

Physical Principles of Ocean Color Remote Sensing

Howard R. Gordon
Distinguished Professor, Emeritus
Department of Physics
University of Miami
Coral Gables, FL, USA

(hgordon@miami.edu)

Copyright © 2019 by Howard R. Gordon

Suggested Citation:

Gordon, H. R. (2019). Physical Principles of Ocean Color Remote Sensing.
University of Miami. <https://doi.org/10.33596/ppocrs-19>

Preface

This work in its present form is meant to be viewed using a PDF viewer such as Adobe Acrobat. The equations and figures are numbered and reference to them is displayed in the color red. Bibliographic references are displayed in the color blue. If your cursor is placed over an equation, figure reference, bibliographic reference, or item entry page in the index, and clicked, the viewer will take you to the particular equation, figure, bibliographic reference, or page on which the index entry appeared. To return to the original position in the text, use the back arrow (on the MacOS operating system; “command”+“left arrow” in Acrobat and “command”+“[” in the Mac application Preview). Note that the actual position of the numbered figures will usually be at the end of the individual chapters

Contents

Preface	iii
0 Preliminaries	1
0.1 Historical Introduction	1
0.2 Scope and Outline of the Present Work	4
0.3 Some Conventions and Notation	12
0.4 Acknowledgments	13
0.5 Dedication	15
I Interaction of Light and Matter	19
1 Review of Electromagnetic Theory	21
1.1 Introduction	21
1.2 Maxwell's Equations	22
1.3 Electromagnetic Waves	26
1.3.1 Harmonic Waves	26
1.3.2 One-Dimensional Plane Waves	27
1.3.3 Three-Dimensional Plane Waves	28
1.3.4 Energy Densities	29
1.3.5 The Poynting Vector	30

1.4	Fields in Matter	32
1.4.1	Four Simple Problems	32
1.4.2	Dielectrics	35
1.4.3	Dense Collection of Atoms in an \vec{E} Field	39
1.4.4	Electromagnetic Waves in Dielectrics	42
1.4.5	Dynamic Properties of Dielectrics	43
1.5	Electromagnetic Wave Propagation in a Dielectric	46
1.6	Reflection of Electromagnetic Waves at a Boundary	51
1.7	Production and Scattering of Electromagnetic Waves	60
1.7.1	Radiation from an Accelerating Charge	60
1.7.2	Radiation from an Oscillating Dipole.	62
1.7.3	The Total Field of a Dipole	64
1.7.4	Scattering of Electromagnetic Radiation by an Atom (Molecule) . .	64
1.7.5	Influence of Molecular Anisotropy	67
1.7.6	Scattering of Electromagnetic Radiation by a Small Dielectric Sphere	72
1.8	The Quantum Nature of Light	73
1.9	Concluding Remarks	74
1.10	Appendix: Analysis of Polarization	75
1.10.1	Plane Waves	75
1.10.2	The Stokes Vector	77
1.10.3	Change in the Reference System	81
1.10.4	The Mueller Matrix	82
1.10.5	Unpolarized (Natural) Light	83
1.10.6	Partially Polarized Light	86
1.10.7	Mueller Matrix for Atomic Scattering and Dielectric Reflection . . .	87

1.10.8 An Alternative Method for Deriving the Mueller Matrix	89
1.11 Bibliographic Notes	92
2 Radiometry and Radiative Transfer	95
2.1 Introduction	95
2.2 Detection of Electromagnetic Radiation	96
2.3 Radiometry	100
2.3.1 Radiant Intensity	100
2.3.2 Radiance	102
2.3.3 Irradiance	110
2.3.4 Radiance of a near parallel beam	111
2.3.5 An example: Is the Moon's surface a lambertian reflector?	113
2.4 Inherent Optical Properties of Media	113
2.5 The Radiative Transfer Equation	120
2.5.1 A Phenomenological Derivation of the Radiative Transfer Equation	120
2.5.2 Uniqueness of RTE Solutions	124
2.5.3 Superposition of Solutions	124
2.5.4 Continuously Varying Refractive Index	125
2.5.5 The Reciprocity Principle	125
2.5.6 Plane-Parallel Media and One-Dimensional Radiative Transfer	127
2.5.7 Irradiances in One-Dimensional Media	128
2.5.8 Measurement of Irradiances	131
2.5.9 Auxiliary Optical Properties	134
2.6 Methods of Solving the RTE	135
2.6.1 Successive Order of Scattering	135
2.6.2 The Monte Carlo Method	139

2.7	Additional Solution Methods	147
2.7.1	Preliminaries	148
2.7.1.1	A Change in Notation	148
2.7.1.2	Phase Function Symmetries and Fourier Analysis	149
2.7.1.3	Gaussian Quadratures	153
2.7.2	The Discrete-Ordinate Method	154
2.7.3	Matrix Operator, Doubling, or Adding Method	159
2.7.4	Invariant Imbedding	161
2.8	Reflecting and Transmitting Boundaries	164
2.8.1	Determination of the Reflection and Transmission Functions	164
2.8.2	Boundaries in Plane-parallel Media	167
2.8.3	A cautionary note	169
2.8.4	The Bi-directional Reflectance Distribution Function — BRDF	170
2.9	An Important Integral — Gershun's Equation	171
2.10	Summary and Concluding Remarks	173
2.11	Appendix 1: First-Order Solutions for 1-d Media	174
2.11.1	First-Order Scattering Solution in Slab Geometry	174
2.11.2	The Single-Scattering Solution	177
2.11.3	The Single-Scattering Solution Including Polarization	180
2.12	Appendix 2: A Multidimensional Integral for Radiance	185
2.12.1	Radiance in a Medium as a Multidimensional Integral	186
2.12.2	Naive Monte Carlo Evaluation of the Integral	190
2.12.3	Forward Monte Carlo Evaluation of the Integral	191
2.12.4	Monte Carlo Solution for Irradiances	197
2.12.5	Backward Monte Carlo Evaluation of the Integral	198

2.12.6 Monte Carlo Solution Including Polarization	200
2.12.7 Connection to the Monte Carlo as Described in Section 2.6.2	203
2.13 Appendix 3: Thermal Radiation	205
2.13.1 Black Body Radiation	206
2.13.2 Radiative Transfer of Thermal Radiation	211
2.14 Bibliographic Notes	215
3 Scattering Theory	217
3.1 Introduction	218
3.2 Rayleigh Scattering	222
3.2.1 Rayleigh Scattering by Atoms and Isotropic Molecules	222
3.2.2 Rayleigh Scattering by Anisotropic Molecules	224
3.2.3 Rayleigh Scattering by Liquids	226
3.3 Scattering by a Homogeneous Sphere	232
3.3.1 The Exact Solution (Mie Theory)	232
3.3.2 Spherical Particles Distributed in Size and Composition	237
3.3.3 Approximations for Small and Large Spheres	238
3.3.3.1 The Rayleigh-Gans Approximation: $ \tilde{m}-1 \ll 1$ and $\kappa R \tilde{m}-1 \ll 1$	238
3.3.3.2 The van de Hulst approximation: $ \tilde{m}-1 \ll 1$ but $\kappa R \gg 1$	244
3.3.3.3 Diffraction	250
3.4 Scattering by Irregularly-Shaped Particles	253
3.4.1 Discrete-Dipole Approximation (DDA)	253
3.4.2 Geometrical-Optics/Diffraction Approximation	257
3.5 Summary	263
3.6 Appendix 1: The Optical Theorem	264

3.7	Appendix 2: Raman Scattering	270
3.8	Appendix 3: Rayleigh and Raman Scattering	273
3.9	Appendix 4. Short Review of Basic Scattering Theory	275
3.10	Bibliographic Notes	278
II	Optical Properties	281
4	Atmospheric Optical Properties	283
4.1	Introduction	283
4.2	Molecular (Rayleigh) Scattering	283
4.2.1	Rayleigh Scattering by the Atmospheric Gas Mixture	284
4.2.2	The Rayleigh Optical Depth	286
4.3	Scattering by Particles (Aerosols)	288
4.3.1	Aerosol Particle Size Distributions	289
4.3.2	Aerosol Refractive Indices	294
4.3.3	Aerosol Models	295
4.3.3.1	Haze C Aerosol Model	296
4.3.3.2	Shettle and Fenn Aerosol Model(s)	299
4.3.3.3	Smirnov et al. Model	305
4.3.3.4	Saharan Dust Models	308
4.3.3.5	Stratospheric Aerosol Models	310
4.3.4	Aerosol Vertical Structure	313
4.4	Absorption by Gases in the Atmosphere	315
4.4.1	General Properties of Gaseous Absorption.	317
4.4.2	Vertical Distribution of Principal Absorbing Gases in the Atmosphere	323

4.4.3	Absorption Spectra of Atmospheric Gases	323
4.4.3.1	Molecular Oxygen	326
4.4.3.2	Water Vapor	327
4.4.3.3	Ozone and NO ₂	328
4.4.3.4	Total Transmittance of the Atmosphere	328
4.5	Simplified Model of the Atmosphere	330
4.6	Concluding Remarks	330
4.7	Appendix: Scattering of Spheres and Spheroids	331
4.8	Bibliographic Notes	333
5	Inherent Optical Properties of Natural Waters	337
5.1	Introduction	337
5.1.1	Definitions	338
5.1.2	Measurement Principles	339
5.2	The IOPs of Pure Water and Sea Water	340
5.2.1	The Absorption Coefficient of Pure Water and Sea Water	340
5.2.2	The Scattering of Pure Water and Sea Water	341
5.2.3	The Raman Scattering of Pure Water	344
5.3	The IOPs of the Constituents of Natural waters	349
5.3.1	Measurement of the Absorption of the Constituents of Case 1 waters	350
5.3.2	Absorption of the Constituents of Case 1 waters	352
5.3.2.1	Absorption of Particles of Case 1 waters	353
5.3.2.2	Absorption by CDOM in Case 1 waters	353
5.3.3	Scattering by Particles in Case 1 Waters	354
5.3.3.1	Total Scattering Coefficient of Particles in Case 1 Waters .	354

5.3.3.2	Volume Scattering Function of Particles in Case 1 Waters .	357
5.3.3.3	The Backscattering Coefficient of Particles in Case 1 Waters	358
5.3.3.4	Polarization of Scattering in Case 1 Waters	360
5.3.3.5	Empirical Model for the Particle Phase Function in Case 1 Waters	363
5.4	Influence of Single Particle Properties on IOPs	364
5.4.1	The Absorption Coefficient	364
5.4.1.1	Applications	367
5.4.2	The Extinction (Attenuation) and Scattering Coefficient	370
5.4.3	The Backscattering Coefficient	372
5.4.4	The Scattering Phase Function	374
5.4.5	Influence of Particle Shape on IOPs	375
5.4.5.1	Absorption	375
5.4.5.2	Extinction and Scattering	376
5.4.5.3	Backscattering	378
5.4.5.4	The Scattering Phase Function	379
5.5	Influence of The Distribution of Particle Size on IOPs	380
5.5.1	Analytical Representations of the Size Distribution	380
5.5.2	Influence of the Distribution of Sizes on IOPs	382
5.5.2.1	Estimation of a_i and \tilde{m} for Size-Distributed Plankton . . .	383
5.5.2.2	IOPs of Power-Law Distributed Particles	384
5.6	Summary	390
5.7	Appendix 1: The Fournier-Forand Phase Function	390
5.8	Appendix 2: Particle Orientational Effects and Averaging	392
5.9	Bibliographic Notes	396

5.10	Figures	401
6	Apparent Optical Properties (AOPs)	435
6.1	Introduction	435
6.2	Apparent Optical Properties	436
6.2.1	Basic Radiometric Quantities	436
6.2.2	Apparent Optical Properties: Definitions	438
6.3	AOP Examples	441
6.3.1	AOPs Derived from Experimental Measurements	442
6.3.2	AOPs Derived from Computed Light Fields	444
6.4	Relationships between AOPs and IOPs	446
6.4.1	The QSSA – A Simple Model to Develop AOP–IOP Relationships	447
6.4.2	The Range of Validity of the QSSA Model for AOPs.	451
6.4.3	Decomposition of Scattering into Water and Particle Components	455
6.4.4	Influence of the Environmental Setting on QSSA Validity	457
6.4.4.1	A Flat Refracting Interface	458
6.4.4.2	A Wind-Roughened Refracting Interface	460
6.4.4.3	An Atmosphere Above a Refracting Interface	462
6.4.4.4	Estimation of E_d^+ and $\langle u_{0w} \rangle_0$ with a Rough Surface, Sun and Sky	463
6.4.4.5	Summary of Environmental Effects on QSSA	467
6.4.5	Depth-Averaged $K_d(z)$	467
6.4.6	Angular Distribution of L_u (BRDF Effect)	470
6.4.7	Degree of Polarization of L_u	472
6.4.8	Summary of QSSA	473

6.4.9	Alternate Formulations of AOP-IOP Relationships	473
6.4.9.1	Hybrid Model for L_u/E_d	474
6.4.9.2	K_d	475
6.4.9.3	$R(0)$	475
6.4.9.4	An Important Observation	476
6.4.10	The Penetration Depth	477
6.4.11	QSSA Estimate of the Influence of Vertical Stratification	480
6.4.12	The Influence of Inelastic Processes on the AOPs	484
6.4.12.1	Radiative Transfer and Inelastic Processes.	485
6.4.12.2	Fluorescence	487
6.4.12.3	Raman Scattering	490
6.4.13	Examples of Modeling with QSSA	500
6.5	Summary	502
6.6	Appendix 1: Estimation of IOPs from AOPs	503
6.6.1	Iterative Inversion Algorithm — Homogeneous Water Bodies	505
6.6.2	Iterative Inversion Algorithm — Stratified Water Bodies	507
6.7	Appendix 2: Bio-Optical Algorithms	509
6.8	Appendix 3: The Zaneveld Solution	511
6.9	Appendix 4: The Modified δ -Isotropic Solution	516
6.10	Bibliographic Notes	521
6.11	Figures	526
7	Optical Effects of the Water Surface	567
7.1	Introduction	567
7.2	Optical Effects of a Flat (Smooth) Water Surface.	568

7.3	Optical Effects of a Wind-ruffled Water Surface.	569
7.3.1	Introduction	569
7.3.2	Reflection of Light by a Wind-Ruffled Surface	570
7.3.2.1	Reflected Direct Solar Radiance: Sun Glitter	574
7.3.2.2	The “Road to Happiness”	574
7.3.2.3	Estimation of $p(\vec{s})$ (Cox and Munk Analysis)	575
7.3.3	Refraction of Light by the Wind-Ruffled Water Surface	580
7.3.4	Requirements for Conservation of Energy Across the Water Surface	582
7.3.5	Reflection and Refraction of Diffuse Radiance by the Water Surface	585
7.3.5.1	General Considerations: Reciprocity	586
7.3.5.2	Reflection of Sky Radiance	588
7.3.5.3	Transmittance of Upwelling Subsurface Radiance (L_u) . . .	589
7.3.6	Incorporation of Wind-Ruffled Surface into Radiative Transfer . . .	590
7.3.6.1	General Methods	590
7.3.6.2	Monte Carlo Methods	592
7.3.7	The Normalized Water-Leaving Radiance	596
7.4	Whitecaps	600
7.5	Numerical Comparisons of Surface Effects	603
7.6	Summary	604
7.7	Appendix 1. Evaluation of an Integral	604
7.8	Appendix 2. Polarization of Sun Glitter	605
7.9	Bibliographic Notes	608
7.10	Figures	611
8	The Sun and Moon	625

8.1	Introduction	625
8.2	The Sun	625
8.2.1	Motion of the Earth relative to the Sun and Fixed Stars	626
8.2.2	Position of the Sun in the Sky	628
8.2.3	Extraterrestrial Solar Irradiance	630
8.3	The Moon	633
8.3.1	Motion of the Moon	633
8.3.2	Radiometry of the Moon	634
8.4	Summary	636
8.5	Appendix: The Equation of Time	637
8.6	Bibliographic Notes	644
8.7	Figures	646
 III Remote Sensing Ocean Color		653
 9 The Remote Sensing Problem		655
9.1	Introduction	655
9.2	The Remote Sensing Problem	656
9.3	Single-Scattering Analysis	656
9.3.1	Contribution of Atmospheric Scattering to $L_{\text{Other}}(0, \hat{\xi})$	657
9.3.2	Identification of Sun Glitter in $L_{\text{Other}}(0, \hat{\xi})$	662
9.3.3	Atmospheric Contribution to Diffuse Transmittance	663
9.3.4	Summary of the Single-Scattering Analysis.	667
9.3.5	Assumption of a Uniform $L_w(\tau_0, u', \phi')$	668
9.4	Accuracy of the Single-Scattering Solution.	671

9.4.1	Accuracy of the Single-Scattering Solution for L_r .	671
9.4.2	Accuracy of the Single-Scattering Solution for $L_r + L_a$.	672
9.4.3	Accuracy of the Single-Scattering Solution for L_a .	673
9.4.4	Accuracy of the Single-Scattering Solution for t_r^* .	673
9.4.5	Accuracy of the Single-Scattering Solution for t^* .	674
9.4.6	Summary	674
9.5	The Water-Leaving Reflectance	675
9.6	Concluding Remarks	678
9.7	L_{Other} Including Polarization	679
9.8	Figures	686
10	Atmospheric Correction	705
10.1	Introduction	705
10.2	Review of the Remote Sensing Problem	707
10.3	A Single-Scattering Solution	710
10.3.1	Application to the CZCS	712
10.3.2	Application to MODIS/SeaWiFS	712
10.3.2.1	Spectral Behavior of ρ_a from Aerosol Models	713
10.3.2.2	Test of SSA Algorithm with Synthetic Data	714
10.4	Multiple Scattering	716
10.4.1	The Multiple-Scattering Retrieval Algorithm	718
10.4.2	Simulated Test of the Multiple-Scattering Algorithm	719
10.4.3	Estimation of Aerosol Optical Depth τ_a	720
10.4.4	The Diffuse Transmittance	722
10.4.5	Ancillary Data Required to Operate the Algorithm	727

10.4.5.1	Ozone Optical Thickness	727
10.4.5.2	Surface Atmospheric Pressure P_0	727
10.4.5.3	Wind Speed W and Wind Vector \vec{W}	728
10.4.5.4	Sea Surface Temperature and Atmospheric Stability	728
10.4.5.5	Relative Humidity RH	728
10.4.6	Annotated Flow Diagram of the Complete Algorithm	728
10.4.7	Non-zero $[\rho_w]_N$ in the NIR	729
10.5	Absorbing Aerosols	729
10.5.1	Atmospheric Correction with a Restricted Set of Candidate Models	731
10.5.2	Aerosol Vertical Structure	731
10.5.3	Strategy for Dealing With Strongly-Absorbing Aerosols	733
10.5.3.1	The Spectral Matching Algorithm (SMA)	734
10.5.3.2	Application of the SMA — A Case Study	736
10.5.3.3	Further Comments Regarding Absorbing Aerosols	740
10.6	Summary and Discussion	741
10.7	Appendix: The SMA Water-Leaving Reflectance Model	743
10.8	Bibliographic Notes	744
10.9	Figures	747
11	Basics of Ocean Color Sensors	773
11.1	Introduction	773
11.2	Radiometric Design Considerations	773
11.2.1	Spectral Bands	774
11.2.2	Acceptable Levels of Sensor Noise	775
11.2.3	Signal Quantization and Saturation Radiance	781

11.3 Imaging with Space-Borne Radiometers	782
11.3.1 Obtaining an Image of the Water-Leaving Radiance	783
11.3.2 Producing a Spectrum of the Water-Leaving Radiance	787
11.3.3 Some Scan Mechanisms	788
11.4 Examples of Ocean Color Scanners	790
11.4.1 The CZCS	790
11.4.2 The MODIS	791
11.5 Summary	793
11.6 Appendix : Comments Regarding Satellite Orbits	793
11.6.1 Polar Orbit	794
11.6.2 Geosynchronous Orbit	795
11.7 Bibliographic Notes	795
11.8 Figures	797
12 Calibration and Characterization	811
12.1 Introduction	811
12.2 Radiometric Calibration	815
12.2.1 Radiance Based Calibration	817
12.2.2 Reflectance Based Calibration	818
12.3 Vicarious Calibration	820
12.3.1 Direct — Using Surface Radiance Measurements	820
12.3.2 Full System Calibration	828
12.4 Characterization	833
12.4.1 Polarization Sensitivity	833
12.4.2 Out-of-Band Response	836
12.4.3 Stray Light	840

12.5 Summary and Discussion	841
12.6 Appendix: Retrieval of ω_a and $P_a(\Theta)$ from $\rho_B(\hat{\xi}_B)$	843
12.7 Bibliographic Notes	847
12.8 Figures	850
13 Vicarious Calibration and Validation	859
13.1 Introduction	859
13.2 How Do We Perform Validation?	860
13.2.1 Surface Measurements of $[\rho_w]_N$	861
13.2.2 How Do We Compare Surface and Sensor Measurements of $[\rho_w]_N$?	864
13.2.3 Validation of Derived Products.	867
13.3 Validation of CZCS and MODIS	868
13.3.1 Verification of CZCS Concept	868
13.3.2 Validation of MODIS Performance	874
13.3.3 Effect and Robustness of Vicarious Calibration	880
13.4 Concluding Remarks	884
13.5 Bibliographic Notes	885
13.6 Figures	887
14 Mathematical Appendix	895
14.1 Complex Numbers and Their Manipulation	895
14.2 Single Variable Calculus	897
14.2.1 Derivatives of a Function	898
14.2.2 Taylor Series Representation of a Function	899
14.2.3 Integration	901
14.2.4 The Dirac Delta Function	902

14.2.5	Differential Equations	904
14.3	Vectors and Multivariable Calculus	909
14.3.1	Vectors — the Elements	909
14.3.2	Functions of Several Variables	911
14.3.3	Differentiation of Multivariable Functions and Vectors	912
14.3.4	Integration of Multivariable Functions and Vectors	915
14.3.5	Functions and Vectors in Spherical Coordinates	918
14.4	Linear Transformations and Matrices	920
14.5	Selected Topics	924
14.5.1	Spherical Triangles: The Laws of Sines and Cosines	924
14.5.2	Least-Squares Analysis	926
14.5.3	Eigenvalues and Eigenfunctions of a Square Matrix	930
14.5.4	Random Variables and the Central Limit Theorem	933
14.5.4.1	Random Variables	934
14.5.4.2	The Central Limit Theorem	936
14.5.5	Monte Carlo Evaluation of Integrals	939
14.5.5.1	Basic Ideas	940
14.5.5.2	Example of an Evaluation	943
14.5.5.3	Reduction of Variance	944
14.5.5.4	Singular Integrands	945
14.5.6	Fourier Series	946
14.5.6.1	Derivation of Fourier Series	946
14.5.6.2	The Fourier Transform	949
14.5.6.3	Generalized Fourier Series	949
14.5.6.4	Relationship Between Fourier series and Least-Squares	950
14.5.7	Orientational Averaging of Direction Cosines	951

14.5.7.1 Direction Cosines	951
14.5.7.2 Orientational Averaging	954
14.6 Bibliographic Notes	957
Bibliography	959
Index	982

Chapter 0

Preliminaries

0.1 Historical Introduction

It has long been known that the sunlight that enters, is backscattered and exits a water body, e.g., the ocean, contains information about the water body itself [Jerlov, 1968, 1975]. The spectral variation of this exiting light is referred to as the “*color*” of the water. Examples of such information are, the near absence of suspended or dissolved matter causing the water to appear a deep blue,¹ the presence of sediment causing the water to be bright throughout the spectrum, the presence of dissolved organic material causing the water to be dark in the blue regions of the spectrum, and the presence of phytoplankton — the first link in the marine food chain — causing the water to appear green. Of course, all of the latter three can be present at the same time. In the late 1960’s and early 1970’s, when Earth-orbiting satellites began to measure sea-surface temperature using the thermal radiation emitted by the water, it became apparent that optical effects such as those mentioned above could be observed from high altitude aircraft, and even from space.

Although there was a vast literature on the optics of natural waters before 1970, arguably the most important work leading to the decision to develop the first satellite-borne ocean color instrument — Coastal Zone Color Scanner (CZCS) — was that of Clarke et al. [1970].

¹Interestingly, Lord Rayleigh, whose theory of molecular scattering explained the blue color of the sky and for whom “*Rayleigh*” scattering is named, believed that the blue color of clear ocean water was due to reflected skylight [Rayleigh, 1910]. It was actually C.V. Raman, for whom “*Raman*” scattering was named, who correctly explained the color of ocean water as being due to the backscattering of the water itself [Raman, 1921], which has the same spectral dependence as the Rayleigh’s molecular scattering (wavelength⁻⁴). Raman won the Nobel Prize in 1930 for his discovery of Raman scattering — a study began because he wanted to understand the color of the ocean.

They combined spectra of the light upwelling above the water over Georges Bank (obtained from an aircraft) with ship-based measurements of the Chlorophyll *a* concentration in the surface water. Although they provided no mathematical relationship between the spectra and the concentration, their results clearly showed that the spectra contained information regarding the Chlorophyll concentration. They also demonstrated that the information was degraded by the backscattering of the atmosphere between the surface and the sensor which they referred to as “air light.”

The first satellite-borne sensor to have several spectral bands in the visible region was Landsat-1 launched July 23, 1972. It was originally named the Earth Resources Technology Satellite (ERTS), but was renamed Landsat 1 in 1975. Its purpose was to assess resources on the land surface. As such its sensitivity was set for imaging the bright land surface rather than the relatively dark (by comparison) ocean. However, even the earliest imagery from Landsat showed that boundaries between ocean currents, e.g., the Gulf Stream and coastal waters [Maul and Gordon, 1975], could be seen in the imagery if the contrast was significantly stretched.² Sediment plumes exiting rivers into the coastal waters were clearly visible as were internal waves propagating near the coasts (seen in and near the Sun’s glitter pattern). However, the oceanographic applications of Landsat imagery were severely limited due to its lack of radiometric sensitivity. This lack of radiometric sensitivity in the visible on Earth-orbiting sensors was remedied by the development and launch of the CZCS on Nimbus 7. The CZCS had approximately an order of magnitude more radiometric sensitivity than Landsat. It was essentially the proof-of-concept mission for ocean color endeavors. The “concept” being that it could produce useful information to the marine science community concerning living marine resources. Interestingly, when the CZCS was selected to be one of the instruments on Nimbus 7, algorithms for processing the data did not exist: there were only data from aircraft and ships showing that there was a relationship, yet to be well developed, between the color of the water and the concentration of Chlorophyll *a* — the photosynthetic pigment in phytoplankton.³

After working on various aspects of ocean optics, light scattering and radiative transfer for about seven years, I became involved in the CZCS project by virtue of being selected as a member of the CZCS Experiment Team through a NASA “Announcement of Opportunity” in the summer of 1975. The full CZCS Experiment Team is provided in Table 1.

²The data from the Multi-Spectral Scanner (MSS) on Landsat-1 were digitized to seven bits, meaning there were 2^7 or 128 radiance values. Imagery delineating the Gulf Stream boundary was stretched so that only digital counts 9 through 13 were displayed. For MSS Band 4 (500-600 nm) the full variation in radiance seen in Case 1 ocean waters was only about 4 digital counts.

³Now, for a satellite sensor to be approved requires that the algorithms be thoroughly tested and that there be no doubt as to the success of the mission — assuming successful launch and instrument operation. However, Earth remote sensing from space at that time was in its infancy and some good ideas were tested on actual satellites, after prototype aircraft instruments showed promise.

Table 1: Members of the CZCS Experiment Team selected in 1975.

Member	Affiliation
W. Hovis (Team Leader)	NASA/GSFC
F. Anderson	NRIO, Capetown, South Africa
R.W. Austin	SIO, Visibility Lab.
E.T. Baker	NOAA/PMEL
D.K. Clark	NOAA/NESS
S.Z. El-Sayed	Texas A&M University
H.R. Gordon	University of Miami
J.L. Mueller*	NASA/GSFC
B. Sturm	JRC Ispra, Italy
R.C. Wrigley	NASA/Ames
C.S. Yentsch	Bigelow Laboratory for Ocean Sciences

* Added after initial formation of the Team.

At the time of the selection of the Team, fabrication of the CZCS was nearly complete.⁴ The basic responsibility of the Team was to develop the necessary scientific algorithms to derive the concentration of Chlorophyll *a* from CZCS data — in other words, make it work. My principal responsibility on the Team was to develop a method for removing the air light from the imagery. This was later termed *atmospheric correction*.⁵ Although funded by NASA and NOAA, there was little support for this effort from the broad oceanographic community, and even hostility from some quarters because the CZCS program was utilizing monies that perhaps could have gone to more “valuable” programs. In addition, computational ability in the late 70’s and early 80’s was crude by standards of even the mid

⁴At that time there was no possibility of modifying the design of the CZCS, although by launch it was known by many in the community that the the instrument could be significantly improved by changing the positions and sensitivities of some of the bands to more effectively carry out atmospheric correction.

⁵It was during my work on atmospheric correction that I came to appreciate the necessity of good radiometric calibration or ocean color scanners. I had developed a preliminary version of the atmospheric correction algorithm and was keen to test the range of its validity using data from the Ocean Color Scanner (OCS) — a NASA aircraft instrument similar to, but preceding, the CZCS. I received data from an OCS flight over the Gulf of Mexico in early 1978 and tried to use them to see if my algorithm made sense. After subtracting an estimate of the water component, I found that the remainder — the air light — had a spectral variation with wavelength (λ) proportional to λ^{-7} . Since molecular scattering is proportional to λ^{-4} , aerosol scattering is approximately proportional to λ^{-1} , and there aren’t any other scatters in the atmosphere, I was understandably concerned. Soon after I learned that the calibration of the OCS could have been in error by as much as 30% in the blue region of the spectrum. I realized then that it would be a waste of time to try to use the OCS data to study the validity of the algorithm. In fact, the CZCS atmospheric correction algorithm was *only* tested prior to launch via modeling — not on actual data.

80's. Computers were slow, a 300 Mb hard drive was considered large (and was expensive), and a few Mb of random access memory (RAM) was rare.⁶ Finally, by 1978 when the first versions of the atmospheric correction and bio-optical algorithms were available, it was apparent that the CZCS spectral band set was not optimum for atmospheric correction — it was not designed with atmospheric correction in mind — and contrary to its name, would probably not work well in the coastal zone.⁷ However, the CZCS program was ultimately highly successful (despite the limitations of the sensor) and its success led to the development of the follow-on U.S. sensors: SeaWiFS, MODIS (Terra), MODIS (Aqua) and VIIRS (Soumi), as well as the European sensors MERIS and OCM, the Indian OCI, the Korean GOCI, and others (22 in total, and counting, as of the fall of 2017).⁸

0.2 Scope and Outline of the Present Work

In this work I present what I believe represents an introduction to the basic physics of ocean color remote sensing. This project was originally conceived as a followup to the work “*Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*” that André Morel and I co-authored in the early 1980's [Gordon and Morel, 1983]. That work followed the first “*Oceanography from Space*” conference in Venice, Italy, at which the early results of the CZCS experiment were first presented to the oceanographic community [Gower, 1981]. We felt that it would be useful to have a document summarizing the algorithms and the preliminary results of what had become known as “*ocean color*.” The work briefly summarized the basic ideas of radiative transfer as applied to optical oceanography, the development of the atmospheric correction algorithm (removal of “air light”) and the in-water or bio-optical algorithms for retrieving the “pigment concentration.”⁹ It also included a more complete discussion of the Case 1 – Case 2

⁶The initial processing of CZCS data was carried out using a mini-computer (PDP 11/55) to manipulate the image, in combination with a mainframe (UNIVAC 1108) that performed the computationally intensive part of the processing. Needless to say, it took hours to process a single 512×512 image.

⁷A few months prior to launch at the “*IUCRM Colloquium: Passive Radiometry of the Ocean*” many researchers from the international ocean optics community were able to exchange ideas concerning the future of ocean color from space. Most knew that the CZCS was not the optimal sensor. The algorithms [Clark et al., 1980; Gordon and Clark, 1980a] had been well enough developed that the “Water Colour Working Group” felt confident in recommending a set of spectral bands for future sensors — three months *before* the launch of CZCS [Morel and Gordon, 1980]. The parallels between these recommendations and the spectral band set on modern sensors — SeaWiFS and MODIS — is striking. An additional recommendation that came out of that working group was to use CZCS to obtain a global data set of phytoplankton pigment concentration — a difficult task because CZCS could only operate two hours per day.

⁸A detailed history of the NASA ocean color missions is available in Aker [2015].

⁹The standard method of measuring Chlorophyll *a* in water samples at that time involved the use of its fluorescence near 685 nm. Phaeophytin *a*, a degradation product of Chlorophyll *a*, interfered with this measurement, so the reported value from such an experiment was the labeled the “pigment concentration:”

optical classification of natural waters. In addition we also presented the first “analytic” bio-optical model that used results of radiative transfer to “explain” the form of the bio-optical algorithms and concluded the report with early validation studies, which showed the pigment concentration could be estimated within a factor of two with CZCS (in 1980). In an appendix we described the (then) recent developments indicating that the factor of two was conservative and the significantly higher accuracy for CZCS-retrieved pigment concentration was possible in the open ocean.

In early 1984, I spent three months at André’s laboratory in Villefranch sur Mer, where we outlined in detail a comprehensive revision of our previous work, even to the point of assigning each other the task of writing on individual topics. However, our continuing research commitments and advocacy for a CZCS follow-on sensor prevented taking any real action on our plan. The present work was to be that revision. I started working on the present manuscript in 2004 and had completed drafts of some of my writing “assignments” at the time of André’s passing in 2012. Feeling inadequate to attempt to cover André’s topics, I decided to reduce the scope of the work to those areas I felt were within my expertise.

In this book I try to present the physics underlying the basic methods of ocean color remote sensing. The goal is not only to provide the reader the basic processing algorithms used to retrieve the desired products, but to show how the algorithms were developed and the reasons for their present form. Also, I hope that the book will provide a basis for the development of new algorithms for processing existing data to yield new products and for developing future projects and sensors.

I have separated the subject into three parts: Part 1, The Interaction of Light and Matter; Part 2, Optical Properties; and Part 3, Remote Sensing of Ocean Color. Parts 1 and 2 provide the background and the theory required for understanding ocean color remote sensing. These are applied in Part 3 to the actual problem of reliably imaging in a quantitative manner the radiance exiting the top of the atmosphere, and retrieving water properties, principally the concentration of Chlorophyll *a*.

Part 1 (The Interaction of Light and Matter) consists of three chapters: Chapter 1, Review of Electromagnetic Theory, Waves, and Scattering; Chapter 2. Radiometry and Radiative Transfer; and Chapter 3. Scattering Theory. Chapter 1 is a review electromagnetic theory describing the nature and production of light as well as its interaction with matter. It is written at a level similar to electrodynamics in the curriculum of an undergraduate physics program (3rd year). Most everything is developed from the Maxwell

the sum Chlorophyll *a* plus Phaeophytin *a*. Now, Chlorophyll *a* is determined by High Pressure Liquid Chromatography (HPLC) leading to the direct use of the Chlorophyll *a* concentration in ocean color remote sensing.

equations, with the exception of the equation for the electromagnetic field of an accelerated charge. This can be developed only after a firm foundation of electrodynamics is in place. Thus, in the case of radiation, we have chosen to simply state the main result and then apply it to situations that will prove useful in the study of environmental optics. Of particular interest for us are (1) the interaction of light with a refracting and absorbing material leading to Snell's law, the Fresnel equations, and the relationship between refraction and absorption of matter (real and imaginary parts of the complex refractive index), (2) the scattering of light from atoms and/or molecules, (3) the Poynting vector, which provides the power per unit area associated with a propagating electromagnetic field, e.g., a light beam, and (4) the polarization of electromagnetic radiation and the manner in which it is described. This chapter requires a higher level of mathematical sophistication (multivariable calculus and vector analysis) than the rest of the material. A reader may skip this chapter and just refer to it when prompted to in the text, if desired.

Chapter 2 deals with the transfer of radiation in a scattering and absorbing medium, e.g., the atmosphere or the ocean. It contains operational definitions¹⁰ of the radiometric quantities important in remote sensing: radiant power, radiant intensity, radiance, and irradiance. Following the radiometric definitions, inherent optical properties (IOPs) — the absorption coefficient, the elastic volume scattering function, and the inelastic volume scattering function — of a homogeneous medium are defined and used to develop a phenomenological derivation of the equation of radiative transfer, which relates the radiance within the medium to its inherent optical properties and to the external (and internal) sources. The radiative transfer equation is then specialized for a medium consisting of homogeneous plan parallel layers (which approximate the ocean and atmosphere very well), essentially rendering the radiance in a given direction dependent only on the spatial coordinate perpendicular the layers. The resulting transfer equation is still an integral-differential equation, and except for the simplest of cases, can only be solved numerically. Numerical methods of solution are presented¹¹ in sufficient detail to be a starting point for readers wishing to write their own radiative transfer code. The presentation ends with a derivation of Gershun's law, which relates the absorption coefficient to the depth variation of the upward and downward propagating irradiances. In an appendix to Chapter 2, the first-order or single-scattering solution to the transfer equation is provided. It is critical to understanding many important topics in the rest of the book, and should be studied and mastered before continuing to the next chapter. Why is the first order solution so important? There are three reasons. First, it provides analytical expressions for the radiance as a function of the inherent optical properties. Next, these expressions applied

¹⁰An operational definition of a quantity is one in which the operations required to measure the quantity are used to define it.

¹¹There are many numerical solutions; however, the ones described are either used to provide examples in the book or to prove certain results, e.g., the existence of the asymptotic light field in a homogeneous medium.

to the atmosphere can yield radiances that are in error by only a few (e.g., 4-5) percent. Finally, a simple modification allows computation of accurate radiances even in media for which single scattering is a poor approximation, e.g., the ocean, and provides simple formulas that can be used to develop more complex retrieval algorithms. As there are many excellent computer codes for solving the radiative transfer equation (including polarization), one might think I am over emphasizing the first-order solution. I am not. Aside from its obvious pedagogical value, historically, it was *actually used* to effect atmospheric correction of CZCS imagery. In addition, it served as a guide for the development of the atmospheric correction algorithm for the follow-on sensors: SeaWiFS, MODIS, and VIIRS. If one wants precise radiances, of course using a radiative transfer code that can produce them is required, but if one wishes to develop an intuitive understanding of the radiative transfer process, the simple solution can go a long way toward providing it. In fact, the theory in much of this book is developed from the first-order solution.

Chapter 3 provides the scattering theory necessary to understand the analysis of the inherent optical properties of the atmosphere and the water and its constituents. First, molecular (Rayleigh) scattering by atoms and molecules (isotropic and anisotropic) is described in detail. A simple model is then used to develop the theory of scattering by liquids. Next the field equations for scattering by a homogeneous sphere (Mie scattering) are provided (not derived) along with some samples of the solution for spheres of a given size and refractive index. Approximate solutions are developed from first principles for both “small” and “large” spheres. Finally, “exact” scattering theory for irregularly-shaped particles is discussed using the concept of the discrete-dipole approximation. In addition, an approximate theory for large (compared to the wavelength of the light) particles based on geometrical optics and diffraction is provided. Selected topics are discussed in appendices. The optical theorem, which is useful in many light scattering calculations is derived in the first appendix. A second appendix describes Raman scattering, and a third develops simple relationships between Raman and Rayleigh scattering.

Part 2 (Optical Properties) describes the optical properties of the various components of the system under consideration. It consists of five chapters: Chapter 4, the Inherent Optical Properties of the Atmosphere; Chapter 5, the Inherent Optical Properties of Natural Waters; Chapter 6 the Apparent Optical Properties of Natural Waters; Chapter 7 the Optics of the Water Surface; and Chapter 8, the Relevant Properties of the Sun and Moon.

Chapter 4, provides the absorption and scattering properties of the components comprising the atmosphere. Topics of particular interest are the scattering by molecules in the atmosphere, the absorption of atmospheric gases such as H_2O , NO_2 , O_3 , and O_2 , and the optics of aerosols. Aerosols are microscopic-sized particles suspended in the air. They both scatter and absorb light. Their scattering is the component of the air light that presents the most difficult problem in atmospheric correction. The principle difficulty is that the

aerosol concentration is highly variable in space and time. This variability is manifest in changes in the size distribution and the refractive index of the aerosol particles, mostly reflecting their source and mode of generation. Here, the aerosol properties are discussed in detail and the scattering theory developed in Chapter 3 is used to provide examples of their scattering and absorption. The aerosol models used in atmospheric correction are described in as well. The typical vertical distribution of the aerosol is illustrated with the help of data from space-borne LIDAR, and a simple (2-layer) model of the atmosphere is proposed for effecting the radiative transfer computations to be utilized in atmospheric correction.

Chapter 5 is the twin of Chapter 4, but focusses on the water (principally the oceans) rather than the atmosphere. However, the basic physics that relates the components of the medium to its inherent optical properties is the same for both the ocean and the atmosphere. Molecular scattering, both Rayleigh and Raman, is developed in detail, but the main emphasis is on the particulate component suspended in the water. The particles of principal interest in the ocean are phytoplankton, the first link in the marine food chain, and their absorbing and scattering detritus. Waters for which these components dominate the optical properties are usually called “Case 1” waters. Optically, the phytoplankton reveal their presence mainly through the absorption of light by the photosynthetic pigment Chlorophyll *a*. Thus, much effort has been focussed on developing empirical relationships between the absorption and scattering coefficients of the planktonic melange to the concentration of Chlorophyll *a*. These relationships are reviewed, and are followed by a theoretical discussion of the optical properties of single particles and their distribution in size and shape, and how these affect the inherent optical properties. A two-component model of particle scattering is developed for use in later chapters.

Chapter 6 concerns the “apparent” optical properties of natural waters (AOPs). These are optical properties that were developed to characterize water masses because they were originally easier to measure than the scattering and absorption properties. Among them are the depth decay coefficients of radiance and irradiance, and the irradiance ratio — the ratio of upwelling irradiance to downwelling irradiance in the water. Of principal interest in remote sensing is the remote sensing ratio, the ratio of the upwelling radiance (toward the zenith) to the downwelling irradiance, both measured just beneath the water surface. The word “apparent” derives from the fact that the AOPs are nearly independent of depth in a homogeneous medium, except near the surface, and thus *appear* to be properties of the medium. However, the actual values of the AOPs depend on the mode of illumination, i.e., the solar zenith angle and the relative amount of cloud cover, as well as the IOPs of the water. Thus, they are *not* solely a property of the medium. The AOPs can be derived by solving the radiative transfer equation given the illumination and the IOPs of the water.

In Chapter 6, first, detailed definitions and examples of the various AOPs based on field

measurements and solutions of the radiative transfer equation, are provided. Then, an approximate, simple model¹² of the AOPs as a function of the IOPs that circumvents solving the transfer equation is developed and applied to many important problems relevant to remote sensing, e.g., the depth to which an external sensor can “see” into the water (the penetration depth), the angular distribution of radiance exiting the water, etc. This model forms the basis for what are termed *semi-analytic algorithms* for retrieving IOPs from remotely sensed radiance. It also illuminates the influence of vertical structure in the IOPs on the remote sensing ratio and the penetration depth. The chapter continues with a detailed study of the effects of inelastic processes — fluorescence and Raman scattering — on the remote sensing ratio.¹³ It ends with examples of using the model and the IOPs as related to the concentration of Chlorophyll *a* from Chapter 5, to delineate the remote sensing ratio as a function of the Chlorophyll *a* concentration. Methods of utilizing measurements of the AOPs to estimate the IOPs are described in an appendix. Three additional appendices describe two alternative approximate solutions to the radiative transfer equation, and provide the bio-optical algorithms actually used in ocean color remote sensing.¹⁴

Chapter 7 deals with the optics of the air-water interface. It begins with the optics of a flat interface (much of which is based on results derived from Chapters 1 and 2), and continues the discussion by examining the reflection and refraction of light from a wind-ruffled water surface, e.g., Sun glitter. Incorporation of a ruffled interface in the radiative transfer equation is described in general and for Monte Carlo simulations — the simulation method used for most of the “exact” radiative transfer calculations presented in Chapter 6. The important concept of *normalized water-leaving radiance* is then defined. It was first developed to approximately relate the actual radiance leaving the water to the radiance that would exit the water in the absence of the atmosphere and with the Sun at the zenith.¹⁵ Finally, the radiative influence of whitecaps on the water surface is discussed.

Chapter 8 describes the relevant facts about radiation from the Sun and the reflective properties of the Moon. The former being important because solar radiation is the ultimate source of the water-leaving radiance, and the latter because it is used to assess the in-orbit radiometric stability of remote sensing systems. We examine the orbital motion of the Earth around the Sun leading to the prediction of the position of the Sun in the sky and hence, the solar zenith and azimuth angles. The extraterrestrial solar irradiance and its spectrum is then described. The radiometry of the Moon, i.e., its reflectance, and the

¹²This model has its roots in the single-scattering solution of the transfer equation. In fact it is a simple modification of the single-scattering solution, and provides additional evidence of its importance.

¹³The single-scattering approximation is shown to be sufficiently accurate to include these processes in the analysis of the of the remote sensing ratio.

¹⁴Algorithms relating water radiance to the concentration of Chlorophyll *a* (or pigment concentration) are usually called “*bio-optical*” algorithms.

¹⁵The phrase “in the absence of the atmosphere” was later modified to “in the absence of atmospheric losses” as this more accurately describes the normalized water-leaving radiance as it is actually computed.

method of its use in stability monitoring is discussed.

Parts 1 and 2 provide the tools necessary to understand the basic principles of ocean color remote sensing, the topic of Part 3. In Part 3 we actually begin the study of remote sensing. We want to use the water-leaving radiance measured from space principally to estimate the concentration of Chlorophyll *a*, but also to estimate other quantities influencing the color of the water, e.g., sediment concentration and detrital concentration. What is required? One might think that the obvious place to start would be the instruments to measure the radiance emanating from the atmosphere at satellite altitudes. However, before we can specify the requirements for such an instrument, such as its radiometric sensitivity and the signal-to-noise ratio, we need to understand in a quantitative manner the entire remote sensing process. For example, we need to know how the measured radiance compares to the water-leaving radiance. This requires understanding in detail the processes that produce the radiance measured at the sensor. We also need to understand how to retrieve the desired water-leaving radiance from the measured radiance, so as to not degrade its information content. This retrieval is usually termed *atmospheric correction* and it will set some of the requirements for the sensor's radiometric accuracy and its noise level. Only after these topics are covered in detail can we provide in a rational manner the attributes that must be required of an instrument.¹⁶ Hence, Part 3 follows this plan and consists of five chapters: Chapter 9, The Remote Sensing Problem; Chapter 10, Atmospheric Correction; Chapter 11, The Basics of Ocean Color Sensors; Chapter 12, Calibration and Characterization; and Chapter 13, Realization and Validation.

Chapter 9 looks in detail at the components of the radiance measured from spacecraft altitudes: the water-leaving radiance; the solar radiation and skylight reflected from the water surface; and the solar radiation backscattered from the atmosphere (the air light). Again, first-order analysis (single scattering) is used to mathematically decompose the effects of these components, and to provide an understanding of each. Then the accuracy of the first-order analysis is demonstrated by comparison with exact computations. Remarkably, the single-scattering development of the atmospheric components of the sensor-measured radiance are only in error by a few percent at most. This is what allowed single scattering theory to serve as the method of atmospheric correction for CZCS and to provide a guide for the development of atmospheric correction of the more-sensitive follow-on sensors. Finally, the effects of the polarization of the radiance in the atmosphere is considered along with the accuracy of the first-order solution in explaining the polarization effects.

Chapter 10 is devoted entirely to atmospheric correction, from its beginnings to more recent developments and methods. The chapter begins with a review of the remote sens-

¹⁶Obviously this program was *not* followed in the design of the CZCS, as the algorithms in question did not exist. However, what did exist were aircraft data on the effect of air light, LANDSAT imagery of coastal phenomena and some measurements that provided estimates of water-leaving radiance.

ing problem. It is followed by an analysis of the accuracy requirements for atmospheric correction based on the accuracy of the bio-optical algorithm relating the concentration of Chlorophyll *a* to the water-leaving radiance. Then single scattering is employed to develop the simple algorithm used for CZCS and to adapt it to newer sensors that possess spectral bands in the near infrared. Next, the single-scattering algorithm is modified to include the effects of multiple scattering through the use of aerosol models and large look up tables (LUTs) derived from exact solutions to the radiative transfer equation.¹⁷ Finally, the difficult problem of an atmosphere containing aerosols that absorb as well as scatter light is addressed using the example of wind-blown dust as a case study.

Chapter 11 concerns ocean color satellite-borne instruments. It begins with a determination of the radiometric design requirements for such a sensor: the position of the spectral bands; the acceptable amount of sensor noise (based on the bio-optical and atmospheric correction algorithms); and the signal quantization and saturation radiances. This is followed by conceptual methods of producing spectral images of the water surface, and their constraints and limitations. The frequent need to address trade-off issues, requiring collaboration between the scientist and the optical engineer is stressed. Finally, examples of ocean color scanners are provided and compared (CZCS and MODIS). An appendix describes the two types of satellite orbits that have been used for ocean color programs: polar (or sun-synchronous) and geosynchronous.

The problem of on-orbit radiometric calibration of ocean color scanners is discussed in detail in Chapter 12. This is vitally important as the processing algorithms cannot perform properly with an inadequate radiometric calibration. We begin by defining just what we mean by *radiometric calibration* and discuss two approaches: radiance based and reflectance based. There are two principal difficulties encountered in radiometric calibration: (1) the calibration accuracy achievable even under laboratory conditions was (at the time of CZCS) less than required; and (2) even if sufficiently accurate calibration could be achieved on the ground, there is no guarantee that this calibration would still be valid after the violence of launch. Thus, on-orbit calibration, and a method of monitoring its stability, is absolutely necessary to ensure a successful mission. Such calibration uses the signal detected by the sensor combined with measurements from the surface. The procedure is called *vicarious calibration*. Two methods of vicarious calibration are discussed in detail. The term “characterization” in the title of Chapter 12 refers to pre-launch assessment of instrument imperfections and peculiarities that can affect the performance of the system.

¹⁷When the author and M. Wang developed the first multiple-scattering atmospheric correction algorithm for SeaWiFS in the early 1990's, computers had maximum sized random-access memories (RAM) of a few megabytes (MB). The full set of LUTs contained about 100 MB of data and therefore could not fit in RAM on most computers. This caused us considerable concern; however, it was not actually a problem as by the time of launch in 1997 even small computers had the requisite amount of RAM. Several times in the ocean color program advances in computation arrived just in time to “save the day.”

Among these are sensitivity to the state of polarization of the radiance, out-of-spectral-band response, and stray light within the instrument that is scattered into the detectors.

Chapter 13 provides a demonstration that all of the above actually works — measurement of spectral radiance at satellite altitudes, along with the various algorithms *does* enable the accurate estimation of water-leaving radiance and through this, the estimation of the concentration of Chlorophyll *a* (or pigment concentration).

0.3 Some Conventions and Notation

With the exception of the present chapter, nearly all references are absent from the actual text, being relegated to a section called “Bibliographic Notes” at the end of each chapter. I preferred this method because it allows some annotation of the papers cited and also allows the addition of references that are extensions or additional works of relevance to the subject, but not described in the text.

As far as notation is concerned, it is standard for physics texts. I use arrows over a quantity to denote a vector quantity, e.g., \vec{Q} , and a hat to denote a unit vector, $\hat{Q} \equiv \vec{Q}/Q$, where Q is the magnitude of \vec{Q} (often written as $|\vec{Q}|$ to explicitly remind the reader that it is the *magnitude* of the vector \vec{Q}). Bold italic symbols represent matrices, e.g., \mathbf{M} . The symbols “ Δ ” for a small increment in a quantity, e.g., Δx , and “ d ” for the differential of the same quantity, dx , are often used interchangeably; however there should be little confusion taking into consideration the context in which they are used. Often I used $\Delta^{(n)}$ to indicate a quantity of n^{th} -order in smallness, if I felt it was necessary for clarity; however, when it is clear as in ΔA , where A is an area, I do not write this as $\Delta^{(2)}A$. Usually if one thinks of the Δ as a differential, then this notation follows immediately. In addition, sometimes I have used ordinary derivatives where strictly partial derivatives are required. Again, there should be little confusion taking into consideration the context in which they are used. I have also used “ \triangleq ” to indicate a *definition*, e.g., an angle θ subtended at the center of a circle of radius r by the arc s of the circle is defined to be s/r or $\theta \triangleq s/r$. Standard functional notation is used throughout the text, i.e., $f(x)$ or $f(\vec{r}, t)$ stands for the value of the function f evaluated at x or at the point \vec{r} and the time t , respectively.

For the space-time variation of electromagnetic waves, I take $\exp[i(\kappa r - \omega t)]$ as opposed to $\exp[-i(\kappa r - \omega t)]$, which requires that the complex refractive index of the medium to be written $m = m_r + im_i$ as opposed to $m = m_r - im_i$. The literature is about equally divided on this with [Mishchenko et al. \[2002a\]](#), [Born and Wolf \[1975\]](#) and [Griffith \[1989\]](#) adopting the former, and [van de Hulst \[1957\]](#), [Hansen and Travis \[1974\]](#) and [Jonasz and Fournier \[2007\]](#) adopting the latter.

I have used MKS (meter-kilogram-second) units throughout the book, except in situations in which hybrid units are more typical or convenient, e.g. for radiance the MKS unit would be “Watt/meter³ Steradian” ($\text{W}/\text{m}^3 \text{ Sr}$) while the hybrid “milliWatts/centimeter² micrometer Steradian” ($\text{mW}/\text{cm}^2 \mu\text{m Sr}$) was used in early ocean color sensing and is more convenient in size. The radiance exiting clear ocean water near noon in the blue part of the spectrum is $\sim 2 \text{ mW}/\text{cm}^2 \mu\text{m Sr}$. Readers should beware that in books on electromagnetic theory that use some variant of the the CGS (centimeter-gram-second) system, such as “gaussian,” “electrostatic,” or “electromagnetic” units, e.g., [Jackson \[1975\]](#),¹⁸ the various formulas will be different from those developed here.

0.4 Acknowledgments

During my 40 plus years working on various ocean color problems, I have had the opportunity to collaborate with many of the leaders in ocean color science. André Morel, obviously, and also Charlie Yentsch and Ros Austin as members of the CZCS Experiment Team, and George Kattawar, Ray Smith, Curt Mobley, and Ron Zaneveld through the ONR ocean optics program. Jerzy Dera introduced me to ocean optics at the very beginning. At the University of Miami I was fortunate to have had many colleagues and collaborators: Ken Voss, Otis Brown, Bob Evans, Barney Balch, Viva Banzon, Kay Kilpatrick, Art Gleason, Paty Matrai, and Chris Boynton. Ken, Otis, Bob, and I were on the MODIS Experiment Team together, having worked on CZCS earlier.¹⁹ Many graduate students and postdocs working with Ken and me made significant contributions to this work both through their Ph.D dissertation research and their subsequent post-doctoral researches. These include M. Wang, K. Ding, T. Du, T. Zhang, R. Chomko, K. Moore, H. Yang, W.R. McCluney, M. Jacobs, C. Hu, Y. Ge, C. Kuchinke, and G. Maul, many of whom suffered though early versions of my course on environment optics, from which the development of many of the topics presented here originated. I recall with great pleasure our Friday afternoon seminars in which ideas were exchanged about planned and ongoing research, often punctuated by cries from a toddler. Also, I was fortunate to be able to host D. Antoine and C. Moulin as visiting scientists in Miami — David produced a CZCS/SeaWiFS 20-year time series and Cyril a spectral matching algorithm for atmospheric correction during Saharan dust outbreaks. I’m grateful to my University of Miami Physics Department colleagues, and

¹⁸The third edition of this famous text uses MKS units, except for parts dealing with the special theory of relativity. The earlier editions used “gaussian” units.

¹⁹Otis Brown and Bob Evans developed the prototype CZCS processing system that was ported to NASA/GSFC and used in processing the first global images for CZCS data by Wayne Esaias, Chuck McClain, and Gene Feldman. In addition, Otis and Bob used an earlier version of their imaging processing system to produce Figure 13.5 (left panel and without the track line) arguably the most famous and widely circulated CZCS image.

former chairs George Alexandrakis and Manuel Huerta in particular, for providing me a supportive and comfortable home in which to pursue my research interests.

The early success of ocean color remote sensing and the CZCS was the result of support (amid much opposition) from Jack Sherman and Harold Yates at NOAA/NESS and Stan Wilson at NASA Headquarters, who had the foresight to hire a Program Manager for ocean biology (on a rotating basis), thus in a sense legitimizing ocean color. This position was filled (in order) by Ken Carder, Wayne Esaias, Curt Davis, Jim Yoder, Frank Muller-Karger, Marlon Lewis, Gregg Mitchell, Robert Frouin, Janet Campbell, John Marra, and Chuck Trees. I, and the whole community as well, owe a debt of gratitude to these scientists who basically put their own research “on hold” for the sake of guiding ocean color through CZCS and the development of SeaWiFS and MODIS. The position was made permanent and is now filled by Paula Bontempi. At Goddard Space Flight Center, the group first known as the SeaWiFS Project (now the “Ocean Biology Processing Group” — “OBPG”), led by Chuck McClain and Gene Feldman, developed the processing system that is second to none. In addition, they effected the application of MOBY data to the calibration of SeaWiFS and MODIS, and developed protocols for “ground truth” validation data (under the direction of Jim Mueller) that are recognized worldwide. Earlier, under the guidance of Wayne Esaias, they had produced the first global CZCS images. These images were instrumental in the acceptance of ocean color by the broader community leading to SeaWiFS as a CZCS follow-on.

I am grateful to NASA and the Program Managers above for their support of my own research in ocean remote sensing, and to ONR under Program Officers Rick Spinrad, Curt Mobley, and Steve Ackleson, for its generous support of my research in other aspects of ocean optics.

Special thanks go to Jim Nearing for providing much-needed advice on the \TeX language (as well as many of his \TeX macros) used in the original draft manuscripts, to Marco Monte whose computer support was indispensable to my research, and to Curt Mobley for reading the entire manuscript and suggesting several modification/additions, which significantly improved the work.

Finally, thanks to my wife and true soul mate Toni, who has made it all possible. Toni has been with me and supported me on my ocean color journey from the start. Our first date was on the day news came that I had been selected as a member of the CZCS Experiment Team, summer 1975.



Dennis Clark aboard the R.V. Athena II, 1978.

0.5 Dedication

I dedicate this work to my friend and colleague Dennis K. Clark, with whom I collaborated throughout my entire career in ocean color remote sensing.

Most researchers in ocean color are aware of Dennis Clark's research focus over the last 20 years of his life: development and deployment of the MOBY system. This was an all-consuming project for Dennis. His goal was to make the best possible measurements of upwelling radiance and water-leaving radiance on a continuing basis for vicarious calibration of ocean color sensors.

What many may not be aware of is the significant impact he and his work had on satellite ocean color remote sensing in its formative years. Dennis was a member of the experiment team (Table 1) developing the first ocean color sensor, the Coastal Zone Color Scanner (CZCS). In the mid 70's CZCS was not a popular program, especially with physical oceanographers, who had their own satellite program — SeaSAT — which was focused on deep-water oceanography, and was a competitor for funding, especially for algorithm development, product validation, ship time, etc. The CZCS concept was largely viewed as the brainchild of a bunch of lunatics — and certainly poor science.

The CZCS focused on U.S. and other coastal regions, which we usually refer to now as Case 2 waters. At that time there was meager evidence that the concept was even viable in such waters — just a few spectra showing that one might be able to measure chlorophyll from aircraft, but there were no algorithms and very little data relating upwelling radiance to chlorophyll. In addition, the atmosphere was known to severely degrade the contrast between low- and high-chlorophyll waters.

The program was continually under attack from those competing for funding, and could easily have been canceled — as many influential oceanographers advocated — or gone on with insufficient funding. However, Dennis knew what needed to be done. With the help of Ros Austin at the Scripps Visibility Laboratory, he built the required instruments, developed the experimental procedures, and set out to collect data for developing a relationship between upwelled radiance and chlorophyll in the coastal waters around the U.S. This led to the first quantitative relationship between the blue-green color ratio and the pigment concentration — the blue-green chlorophyll algorithm. It gave the program needed credence, and he was able to convince NOAA/NESDIS leadership that the program was worth continuing to support, even though NOAA's Environmental Research Laboratories leadership remained firmly opposed to it (ERL was another, more research oriented, part of NOAA).

Dennis' algorithm development program was very ship-time intensive — only one station per day — yet he was able to secure sufficient funds to carry out the program. Although he abhorred the politics of government science, he was a master at it. In addition to the pre-launch effort, he convinced NOAA/NESDIS to provide the funding for ship-time for much of the CZCS validation. In the post-launch validation effort, he was able to make fifty-five stations underneath satellite overpasses using converted U.S. Naval Gunboats from the Vietnam War era (Athena and Athena II) that were capable of 40 knots for short periods. To understand the difficulty presented by validation, it should be noted that, for a variety of reasons (clouds, electronic over-shoot of the sensor, etc.), only nine of these stations provided actual validation points, although the additional fifty-five stations did lead to significant improvement in the pigment algorithm. His validation effort proved that the CZCS worked well in some coastal areas, but more importantly, the CZCS produced

spectacular chlorophyll imagery in the open ocean (Case 1 waters). This open-ocean imagery was validated with Dennis' data collected along-track during the validation cruises. It eventually led to a whole series of ocean color instruments.

Following the success of CZCS and the green light for SeaWiFS as a “new start” at NASA, Dennis started thinking about developing a system for vicarious sensor calibration, a procedure that the CZCS experience showed was an absolute necessity. That led directly to the present-day MOBY. It has been used to provide calibration for most ocean color instruments launched since the mid 90's. The rest is history.

We who care about ocean color owe a depth of gratitude to Dennis.

Part I

Interaction of Light and Matter

Chapter 1

Review of Electromagnetic Theory, Waves, and Scattering

1.1 Introduction

On a fundamental level, this monograph deals with the interaction of electromagnetic waves with matter — living and nonliving. To understand this interaction requires some knowledge of electromagnetic theory. In this chapter, we provide a review of the electromagnetic theory that we believe is required for a through understanding of this interaction, leaving out of the discussion subject matter that by its nature is not required in either oceanic or atmospheric optics, e.g., the properties of magnetic materials and their influence on the fields. Simple classical or semi-classical models of the atom will be used to provide a qualitative description of phenomena of interest. The fact that such a description is possible, when it is known that quantum dynamics is required for a correct treatment of many of the phenomena, is the reason such models held a measure of credence prior to the quantum age. In our review of the subject, we will take Maxwell's field equations and the Lorentz force law as given and proceed with the development using very simple solutions to the field equations. The ultimate goal is to understand electromagnetic waves and their interaction with matter. In addition, knowledge of the sources of radiation, e.g., radiation for an oscillating dipole, is essential for understanding scattering by atoms, molecules, density fluctuations, and small particles.

We begin by presenting Maxwell's equations, for the electric (\vec{E}) and magnetic (\vec{B}) fields, their source densities, and the field energy densities. Then we develop Poynting's theorem, show that the field equations allow the existence of electromagnetic waves in

vacuum, and examine their properties. Next we examine the presence of non-conducting matter (dielectrics) on the fields to ultimately determine their influence on electromagnetic waves. This development is facilitated by first solving four simple problems and applying the solutions to a collection of atoms using a simplified atomic model. This leads to the concept of a dielectric becoming a collection of electric dipoles upon exposure to an electric field. The polarization of the dielectric is then defined as the dipole moment density and the field inside a dielectric medium (as defined through Maxwell's equations) is related to the field seen by an individual atom or molecule. The presence of dielectrics leads to simple modifications of Maxwell's equations, which still support wave motion within dielectrics, but with propagation at a speed less than in a vacuum. The influence of these time varying-fields, within a dielectric medium, on the individual atoms is then developed with a classical atomic model and yields the phenomena of complex refractive index, absorption and dispersion. The propagation of electromagnetic waves across a plane interface between two media (e.g., air and water), is then examined and reflection, refraction, total internal reflection, etc., described. Finally, the field vectors \vec{E} and \vec{B} for a charge in motion with an arbitrary velocity and acceleration are presented (not derived from Maxwell's equations) and it is shown that accelerated charges lead to electromagnetic radiation. This is used to examine radiation from a dipole, scattering of radiation from an atom (or molecule) and scattering of radiation from a small dielectric particle. In addition, a description of the polarization of electromagnetic waves is provided in an appendix.¹

1.2 Maxwell's Equations

The *electric field* (\vec{E}) and the *magnetic field* (\vec{B}) in space influence the motion of a charge q moving with a velocity \vec{v} through the application of a force \vec{F} given by

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}.$$

This is usually referred to as the “Lorentz force.” The sources of the fields are charges and currents. The fields are related to the sources, and to each other, through the Maxwell equations:

$$\begin{aligned} \text{(I)} \quad & \nabla \bullet \vec{E}(\vec{r}, t) = \rho(\vec{r}, t)/\epsilon_0 \\ \text{(II)} \quad & \nabla \bullet \vec{B}(\vec{r}, t) = 0 \\ \text{(III)} \quad & \nabla \times \vec{E}(\vec{r}, t) = -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t} \\ \text{(IV)} \quad & \nabla \times \vec{B}(\vec{r}, t) = \mu_0 \vec{j}(\vec{r}, t) + \mu_0 \epsilon_0 \frac{\partial \vec{E}(\vec{r}, t)}{\partial t}. \end{aligned} \tag{1.1}$$

¹The mathematics required for this chapter, namely vector manipulations, vector calculus, complex numbers, differential equations, and matrices, is reviewed in the Mathematical Appendix.

(Note, when referring to the individual Maxwell equations we will use the roman numerals displayed on the left of Eqs. (1.1).) In these equations, $\rho(\vec{r}, t)$ is the charge density at the position specified by the vector \vec{r} (from the origin of coordinates to the point (x, y, z)) defined through

$$\rho(\vec{r}, t) \equiv \frac{dq(\vec{r}, t)}{dV}$$

where $dq(\vec{r}, t)$ is the total amount of charge in an infinitesimal volume element $dV = dx dy dz$ centered on \vec{r} . The quantity \vec{j} is the density of the electric current associated with the motion of charge, and defined as follows. Consider an infinitesimal area dA_\perp with its normal oriented in the direction of the flow of charge. Then the magnitude of \vec{j} is defined by

$$j(\vec{r}, t) \equiv \frac{dq^{(3)}(\vec{r}, t)}{dA_\perp dt}$$

where $dq^{(3)}(\vec{r}, t)$ is the charge flowing across the dA_\perp in time dt . The electric current $dI = dq/dt$ flowing through a surface of area dA is then $dI = \vec{j} \bullet \hat{n} dA$, where \hat{n} is the normal to dA . In this work, we will employ the word “flux” as commonly used in geophysics, i.e., the flux of any quantity Q is the amount ΔQ that crosses an area ΔA_\perp (orientated normal to the flow of Q , if there is any) in a time Δt divided by $\Delta A_\perp \Delta t$. Thus, the current is the flux of charge. The direction of \vec{j} is in the direction of motion of the charge.

Charge is conserved, that is the rate at which charge enters (leaves) a fixed volume must equal the rate at which the amount of charge inside increases (decreases). The rate at which charge enters a closed volume V through its bounding surface S is

$$\oint_S \vec{j}(\vec{r}, t) \bullet \hat{n} dA = \int_V \nabla \bullet \vec{j}(\vec{r}, t) dV,$$

where \hat{n} is the *outward* normal to the surface, and the right-hand-side of this equation results from application of the divergence theorem (see the Mathematical Appendix). The rate at which charge increases in V is

$$\frac{d}{dt} \int_V \rho(\vec{r}, t) dV = \int_V \frac{\partial}{\partial t} \rho(\vec{r}, t) dV.$$

Combining these, and noting that the volume V is arbitrary, we have

$$\nabla \bullet \vec{j}(\vec{r}, t) + \frac{\partial \rho(\vec{r}, t)}{\partial t} = 0.$$

This is called the equation of continuity.

If we have a single, e.g., positive, charge carrier moving with a velocity $\vec{v}(\vec{r}, t)$ with charge density $\rho(\vec{r}, t)$, it is a simple matter to show that

$$\vec{j}(\vec{r}, t) = \rho(\vec{r}, t) \vec{v}(\vec{r}, t).$$

Table 1.1: SI units of quantities in Maxwell's equations.

Quantity	Symbol	Unit
Length	\vec{r}, x , etc.	Meter [m]
Mass	m	Kilogram [kg]
Time	t	Second [s]
Force	\vec{F}	Newton [N]
Charge	q	Coulomb [C]
Charge Density	ρ	[C/m ³]
Current	I	Ampere [A = C/s]
Current Density	\vec{j}	[A/m ²]
Electric Field	\vec{E}	[N/C]
Magnetic Field	\vec{B}	Tesla [T = N/A·m]
Energy	U	Joules [J = N·m]
Energy Density	u	[J/m ³]
Work	W	[J]
Power	dW/dt	Watt [J/s]

The constants ϵ_0 and μ_0 in the Maxwell equations are, respectively, the permittivity and permeability of free space. Their values are $\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2$ and $\mu_0 = 4\pi \times 10^{-7} \text{ N/A}^2$ (exactly). The SI units of all of the quantities in these equations are presented in Table 1.1.

It is important to note that Maxwell's equations are linear. This means that if \vec{E}_1 and \vec{B}_1 are solutions resulting from the sources ρ_1 and \vec{j}_1 alone and \vec{E}_2 and \vec{B}_2 are solutions resulting from the sources ρ_2 and \vec{j}_2 alone, then $\vec{E}_1 + \vec{E}_2$ and $\vec{B}_1 + \vec{B}_2$ are solutions resulting from the sources $\rho_1 + \rho_2$ and $\vec{j}_1 + \vec{j}_2$, i.e., when both sources are simultaneously present. This is called the principle of superposition and will be used frequently in this chapter.

Consider a distribution of charges and currents that produce the fields \vec{E} and \vec{B} . The charges and currents interact with these fields through the Lorentz force, and will therefore be modified by the very fields that they produce. (Note, a charge $dq = \rho dV$ will not interact with the fields that it produces, but it will interact with the fields produced by all of the other charges and currents.) The work done by the fields on dq in time dt , during which dq moves a distance $d\vec{r}$, will be

$$dW = \vec{F}_{\text{on } dq} \bullet d\vec{r} = dq(\vec{E} + \vec{v} \times \vec{B}) \bullet d\vec{r} = dq(\vec{E} + \vec{v} \times \vec{B}) \bullet \frac{d\vec{r}}{dt} dt = dq(\vec{E} + \vec{v} \times \vec{B}) \bullet \vec{v} dt = dq\vec{E} \bullet \vec{v} dt.$$

Now, multiply and divide the right-hand-side of this by dV , introduce the charge density

ρ , identify $\rho\vec{v}$ with \vec{j} , and integrate over the entire current distribution. The result is

$$\frac{dW}{dt} = \int_V \vec{E} \bullet \vec{j} dV.$$

This is the rate at which the fields do work on the charge and current distributions. Notice that the value of the integral is unchanged if the volume V is enlarged without limit as long as it contains all of the current, i.e., as long as there is no current outside of V . Thus, we take V to be any volume having the current distribution completely inside.

We want to write the right-hand-side of this in terms of the fields alone. This can be done using Maxwell's equations as follows. First, use Eq. (IV) to eliminate \vec{j} :

$$\vec{E} \bullet \vec{j} = \frac{\vec{E} \bullet \nabla \times \vec{B}}{\mu_0} - \epsilon_0 \vec{E} \bullet \frac{\partial \vec{E}}{\partial t},$$

where, for compactness of notation, we have suppressed the explicit dependence of the fields on \vec{r} and t . Next, use the vector identity $\nabla \bullet (\vec{E} \times \vec{B}) = \vec{B} \bullet (\nabla \times \vec{E}) - \vec{E} \bullet (\nabla \times \vec{B})$ and Eq. (III) to replace $\vec{E} \bullet \nabla \times \vec{B}$ with

$$-\nabla \bullet (\vec{E} \times \vec{B}) - \vec{B} \bullet \frac{\partial \vec{B}}{\partial t}.$$

Finally, note that

$$\vec{B} \bullet \frac{\partial \vec{B}}{\partial t} = \frac{1}{2} \frac{\partial (\vec{B} \bullet \vec{B})}{\partial t} = \frac{1}{2} \frac{\partial B^2}{\partial t},$$

etc., and combine all of these into

$$\frac{dW}{dt} = - \int_V \left[\frac{1}{2\mu_0} \frac{\partial B^2}{\partial t} + \frac{\epsilon_0}{2} \frac{\partial E^2}{\partial t} \right] dV - \frac{1}{\mu_0} \int_V \nabla \bullet (\vec{E} \times \vec{B}) dV.$$

Now, if the volume V is fixed (i.e., not moving) the time derivatives can be taken out of the first integral (and become ordinary derivatives since the integrals depend only on t). Then applying the divergence theorem to the second term on the right-hand side, we have

$$\frac{dW}{dt} = - \frac{d}{dt} \int_V \left[\frac{1}{2\mu_0} B^2 + \frac{\epsilon_0}{2} E^2 \right] dV - \frac{1}{\mu_0} \oint_S (\vec{E} \times \vec{B}) \bullet \hat{n} dA, \quad (1.2)$$

where in the last integral \hat{n} is the outward normal (points away from V) and the integration is taken over the surface S bounding the volume V . In the last term on the right-hand side, the quantity

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}$$

is called the Poynting vector. As it has dimensions of energy divided by area times time, i.e., a flux, it is a flux of energy. Therefore, the last term in Eq. (1.2) is the energy

flowing through the surface bounding the volume V per unit time. Equation (1.2) is called Poynting's theorem. This suggests that we should identify

$$u_e = \frac{\epsilon_0}{2} E^2 \quad \text{and} \quad u_m = \frac{1}{2\mu_0} B^2.$$

as the energy densities (energy per unit volume) associated with the electric and magnetic fields within V , respectively. Then Eq. (1.2) expresses the *conservation of energy*: the rate at which work is done by the fields on the charges and currents within V is equal to the rate of decrease of energy stored in the fields within V , minus the rate at which energy is leaving the through surface S bounding V .

1.3 Electromagnetic Waves

The electromagnetic field supports wave motion and in fact can transport energy in the form of electromagnetic waves. To understand how waves are implicit in Maxwell's equations, we consider a region of space free of charges and currents, e.g., a region where $\rho = 0$ and $\vec{j} = 0$. Then if we take $\nabla \times$ Eq. (III) and use Eq. (IV), we obtain

$$\nabla \times \nabla \times \vec{E} = -\mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2},$$

which, using the last entry in Eq. (14.12) from the Mathematical Appendix and noting that $\nabla \bullet \vec{E} = 0$, becomes

$$\nabla^2 \vec{E} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0. \quad (1.3)$$

By taking $\nabla \times$ Eq. (IV) and performing similar operations, we find \vec{B} satisfies the same equation, i.e.,

$$\nabla^2 \vec{B} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2} = 0. \quad (1.4)$$

Thus, we see that both \vec{E} and \vec{B} satisfy the three-dimensional wave equation. We now study some of the properties of electromagnetic waves.

1.3.1 Harmonic Waves

We shall consider only harmonic waves, i.e., waves that have a sinusoidal time dependence. The rational for this is that one can synthesize any time dependence from such sinusoidal functions through Fourier analysis. Thus, we look for solutions of the form

$\vec{E}(\vec{r}, t) = \vec{E}(\vec{r}) \exp[-i\omega t]$. (If the reader is unfamiliar with complex notation, consult the Mathematical Appendix.) Substituting this into the wave equation yields

$$\nabla^2 \vec{E} + \mu_0 \epsilon_0 \omega^2 \vec{E} = 0.$$

1.3.2 One-Dimensional Plane Waves

To simplify the previous equation, let's assume that $\vec{E}(\vec{r}) = \vec{E}(x)$, i.e., \vec{E} depends only on x . This reduces the equation to an ordinary differential equation:

$$\frac{d^2 \vec{E}}{dx^2} + \mu_0 \epsilon_0 \omega^2 \vec{E} = 0. \quad (1.5)$$

A solution of this is

$$\vec{E}(x) = \vec{E}_0 \exp[+i\sqrt{\mu_0 \epsilon_0} \omega x], \quad (1.6)$$

where \vec{E}_0 is a constant vector. The corresponding field is

$$\vec{E}(x, t) = \vec{E}_0 \exp[+i(\sqrt{\mu_0 \epsilon_0} \omega x - \omega t)].$$

This can be rewritten

$$\vec{E}(x, t) = \vec{E}_0 \exp[+i\sqrt{\mu_0 \epsilon_0} \omega (x - \frac{1}{\sqrt{\mu_0 \epsilon_0}} t)],$$

showing that a given value of $\vec{E}(x, t)$ moves (propagates) in the $+x$ direction with a speed $1/\sqrt{\mu_0 \epsilon_0}$. The reader can readily verify that this speed is in fact that of light in a vacuum ($c = 2.99 \times 10^8$ m/s). Henceforth, we replace $\mu_0 \epsilon_0$ by $1/c^2$. Then the field is

$$\vec{E}(x, t) = \vec{E}_0 \exp[+i(\frac{\omega}{c} x - \omega t)] = \vec{E}_0 \exp[+i(\kappa x - \omega t)], \quad (1.7)$$

where $\kappa \triangleq \omega/c$ is called the wave number (also written “wavenumber”). Writing

$$\vec{E}_0 = \hat{i}E_{0x} + \hat{j}E_{0y} + \hat{k}E_{0z},$$

where \hat{i} , \hat{j} , and \hat{k} are unit vectors in the x , y , and z directions, respectively, and applying Eq. (I) (with $\rho = 0$) to $\vec{E}(x, t)$, yields

$$i\kappa E_{x0} = 0,$$

which shows that \vec{E}_0 has no x component. Thus, this wave solution has no \vec{E} component in the direction of propagation. It is a *transverse* wave.

For a given value of time, say $t = t_0$, $\vec{E}(x, t_0)$ is periodic in x repeating for values of x given by $x \pm 2\pi/\kappa$, $x \pm 4\pi/\kappa$, etc. The quantity $2\pi/\kappa$ is defined to be the wavelength λ of the wave. Similarly, fixing x at x_0 , $\vec{E}(x_0, t)$ will repeat when t is increased by an integer times $2\pi/\omega$, a quantity that defines the period T of the wave.

Note that there are other solutions of Eq. (1.5). For example, clearly

$$\vec{E}(x) = \vec{E}_0 \exp[-i\sqrt{\mu_0\epsilon_0}\omega x]$$

also satisfies Eq. (1.5). When combined with the time factor $\exp(-i\omega t)$ this solution represents a wave propagating in the $-x$ direction. In fact, the sum of these two waves, one leading to propagation in the $+x$ direction and the other to propagation in the $-x$ direction, with arbitrary \vec{E}_0 's for each is the *general* solution of Eq. (1.5).

The magnetic field associated with these waves can be found from Eq. (III). Since \vec{E} has no x component,

$$\frac{\partial \vec{B}}{\partial t} = \hat{j} \frac{\partial E_z}{\partial x} - \hat{k} \frac{\partial E_y}{\partial x}.$$

Inserting Eq. (1.7) into this, integrating over time, and setting the constants of integration (time independent \vec{B} and therefore, *not* a wave) to zero, gives

$$B_x = 0, \quad B_y = -\frac{\kappa}{\omega} E_z \quad \text{and} \quad B_z = +\frac{\kappa}{\omega} E_y.$$

This can be written compactly as

$$\vec{B}(x, t) = \frac{\kappa}{\omega} \hat{i} \times \vec{E}(x, t) = \frac{\hat{i} \times \vec{E}(x, t)}{c},$$

and the magnetic field that accompanies \vec{E} in Eq. (1.7) is

$$\vec{B}(x, t) = \vec{B}_0 \exp[+i(\kappa x - \omega t)], \quad \text{with} \quad \vec{B}_0 = \frac{\hat{i} \times \vec{E}_0}{c}. \quad (1.8)$$

This shows that (1) \vec{E} and \vec{B} are perpendicular to each other, (2) the plane formed by \vec{E} and \vec{B} is perpendicular to the direction of propagation of the wave, and (3) there is no phase difference between \vec{E} and \vec{B} , i.e., they are *in phase*. Because the surfaces of constant \vec{E} and \vec{B} are planes (parallel to the y - z plane), waves of this form are called *plane waves*.

1.3.3 Three-Dimensional Plane Waves

It is easy to generalize plane waves to three dimensions. Letting

$$\vec{\kappa} = \hat{i}\kappa_x + \hat{j}\kappa_y + \hat{k}\kappa_z$$

and

$$\vec{r} = \hat{i}x + \hat{j}y + \hat{k}z,$$

we see that

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)] \quad (1.9)$$

satisfies Eq.s (??) and (1.4), if

$$\kappa^2 = \kappa_x^2 + \kappa_y^2 + \kappa_z^2 = \frac{\omega^2}{c^2}.$$

Proceeding as in the above case of waves traveling in the x direction, Eq. (I) shows that

$$\vec{\kappa} \bullet \vec{E}(\vec{r}, t) = 0,$$

and Eq. (III) shows that

$$\vec{B}(\vec{r}, t) = \frac{\vec{\kappa}}{\omega} \times \vec{E}(\vec{r}, t) = \frac{\hat{\kappa} \times \vec{E}(\vec{r}, t)}{c}. \quad (\text{Note: } \hat{\kappa} \text{ not } \vec{\kappa} \text{ in the second equality.})$$

Thus,

$$\vec{B}(\vec{r}, t) = \vec{B}_0 \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)] \quad (1.10)$$

with

$$\vec{B}_0 = \frac{\vec{\kappa}}{\omega} \times \vec{E}_0 = \frac{\hat{\kappa} \times \vec{E}_0}{c},$$

and Eq. (1.9) constitute a solution to the three-dimensional wave equation. This solution corresponds to waves propagating in the $\vec{\kappa}$ direction, i.e., rectilinear propagation. The surfaces of constant \vec{E} and \vec{B} are perpendicular to $\vec{\kappa}$ and \vec{E} and \vec{B} are perpendicular to each other. The wave length is $2\pi/\kappa$ and the period is $2\pi/\omega$.

1.3.4 Energy Densities

The energy densities are proportional to the square of the fields, and as such we must revert to the real representation of the fields in order to compute them, i.e., replace $\exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)]$ by $\cos(\vec{\kappa} \bullet \vec{r} - \omega t)$ (See Mathematical Appendix). The resulting densities associated with a plane electromagnetic wave are

$$u_e = \frac{\epsilon_0}{2} E_0^2 \cos^2(\vec{\kappa} \bullet \vec{r} - \omega t)$$

and

$$u_m = \frac{1}{2\mu_0} B_0^2 \cos^2(\vec{\kappa} \bullet \vec{r} - \omega t),$$

but since $B_0 = E_0/c$ and $\mu_0\epsilon_0 = 1/c^2$, we see that $u_m = u_e$, i.e., the energy densities associated with the electric and the magnetic fields are equal. Thus, in a plane electromagnetic wave, the field energy is equally divided between the electric and magnetic fields. The total field energy density is then $2u_e$.

1.3.5 The Poynting Vector

As with the energy densities, to compute the Poynting vector for a plane wave we need to use the real representation of the fields:

$$\vec{S}(\vec{r}, t) = \frac{\vec{E}_0 \times \vec{B}_0}{\mu_0} \cos^2(\vec{\kappa} \bullet \vec{r} - \omega t),$$

but

$$\vec{E}_0 \times \vec{B}_0 = \frac{\vec{E}_0 \times (\hat{\kappa} \times \vec{E}_0)}{c} = \frac{E_0^2}{c} \hat{\kappa}.$$

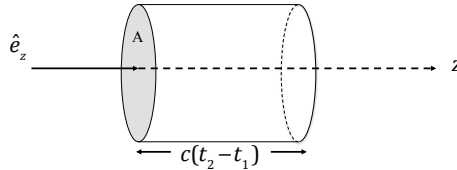
Therefore,

$$\vec{S}(\vec{r}, t) = \frac{E_0^2}{c\mu_0} \hat{\kappa} \cos^2(\vec{\kappa} \bullet \vec{r} - \omega t) = c\epsilon_0 E_0^2 \hat{\kappa} \cos^2(\vec{\kappa} \bullet \vec{r} - \omega t) = \hat{\kappa} c u_{em}(\vec{r}, t),$$

where $u_{em} = u_e + u_m = 2u_e$, the electromagnetic energy density associated with the wave. How do we interpret $\vec{S}(\vec{r}, t) = \vec{\kappa} c u_{em}(\vec{r}, t)$? For simplicity, consider a wave propagating in the z direction; specifically, $\vec{\kappa} = \hat{e}_z \kappa$, so

$$\vec{E}(z, t) = \hat{e}_x E_0 \cos(\kappa z - \omega t).$$

Let this wave be incident on a disk of area A as shown below. Consider the integral of



the Poynting vector over the area A , i.e., $\int_A \vec{S}(z, t) \cdot \hat{e}_z dA$. We have

$$\int_A \vec{S}(z, t) \cdot \hat{e}_z dA = c \int_A u_{em}(z, t) dA.$$

Now average this over a time interval $\mathcal{T} = t_2 - t_1$ (long compared to the period of the wave), i.e.,

$$\frac{1}{\mathcal{T}} \int_{t_1}^{t_2} \int_A \vec{S}(z, t) \cdot \hat{e}_z dA dt = c \frac{1}{\mathcal{T}} \int_{t_1}^{t_2} \int_A u_{em}(z, t) dA dt.$$

Note that $\frac{1}{\mathcal{T}} \int_{t_1}^{t_2} \vec{S}(z, t) \cdot \hat{e}_z dt = \langle \vec{S}(z) \rangle_{\text{Avg}} \cdot \hat{e}_z$, where $\langle \vec{S}(z) \rangle_{\text{Avg}}$ is the time-averaged Poynting vector on the surface A . Also, note that $c(t_2 - t_1)$ is the length of the cylindrical volume shown in the figure. Thus, the integral on the right is just the total electromagnetic energy, U_{em} , contained within the volume $Ac(t_2 - t_1)$, so

$$\int_A \langle \vec{S}(z) \rangle_{\text{Ave}} \cdot \hat{e}_z dA = U_{em}/\mathcal{T}.$$

Since the energy associated with the wave within the volume $\mathcal{T}Ac$ had to pass through A in time \mathcal{T} , U_{em}/\mathcal{T} must be the *flux* of energy through A . In differential form, the above equation reads

$$\langle \vec{S}(z) \rangle_{\text{Ave}} \cdot \hat{e}_z = dU_{em}/dt dA,$$

so we can conclude that for an electromagnetic wave, $\langle \vec{S}(z) \rangle_{\text{Avg}} \cdot \hat{e}_z$ is the time-averaged flux of energy in the direction \hat{e}_z . But, as dt is a differential, $\langle \vec{S}(z) \rangle_{\text{Avg}} \cdot \hat{e}_z = \vec{S}(z) \cdot \hat{e}_z$, and we finally conclude that we can associate \vec{S} with the instantaneous flux of energy (for an electromagnetic wave).²

At optical frequencies, most detectors of electromagnetic waves respond to the power falling on their surface, which is proportional to the Poynting vector; however, those employed in remote sensing cannot respond fast enough to follow the actual oscillations of the fields (or of the Poynting vector). Rather, they respond to the time-averaged Poynting vector over many oscillations of the fields, i.e., they respond to

$$\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} \triangleq \frac{1}{\mathcal{T}} \int_{t_0}^{t_0+\mathcal{T}} \vec{S}(\vec{r}, t) dt,$$

where T is the period, and $\mathcal{T} \gg T$.

²Later in this chapter, we will encounter (spherical) waves having the form

$$\vec{E}(\vec{r}, t) = \frac{\exp[i(\kappa r - \omega t)]}{r} \left(f(\theta, \phi) \hat{\theta} + g(\theta, \phi) \hat{\phi} \right),$$

in spherical coordinates. (The unit vectors are defined in the Mathematical Appendix). A similar computation with this field yields a similar result.

The time-averaged Poynting vector can also be computed in a simple manner using the complex representation of the fields through

$$\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} = \frac{1}{2\mu_0} \Re(\vec{E} \times \vec{B}^*) = \frac{\hat{\kappa}}{2\mu_0 c} \Re(\vec{E} \bullet \vec{E}^*) = \frac{\hat{\kappa}}{2\mu_0 c} E_0 E_0^*, \quad (1.11)$$

where \vec{B}^* and \vec{E}^* are the complex conjugates of \vec{B} and \vec{E} , respectively.

1.4 Fields in Matter

We have seen how the electromagnetic field can propagate energy in the form of waves in vacuum; however, we need to understand how the presence of matter influences these waves, as well as understand how electromagnetic waves influence matter. We shall introduce matter in the form of dielectrics, i.e., the form of matter that in general does not conduct electricity and has no significant magnetic properties. Seawater does conduct electricity somewhat at low frequencies; however, at the optical frequencies of interest in environmental optics, it does not. The electric field in a dielectric will be developed by considering the effect of an external field on a simple model of an atom — a nucleus surrounded by a spherically symmetric cloud of electrons.

1.4.1 Four Simple Problems

Before we begin a discussion of dielectric materials, it is useful to develop some background ideas by solving four simple problems in electrostatics (i.e., all quantities in Maxwell's equations are independent of time). The solutions will result from an application of Gauss' law. If we integrate Eq. (I) over a volume V , we can use the divergence theorem to relate this to a surface integral, i.e.,

$$\int_V \nabla \bullet \vec{E} dV = \oint_S \vec{E} \bullet \hat{n} dA,$$

where S is the bounding surface of V and \hat{n} is the *outward* normal. However, from Eq. (I), the first integral is just

$$\int_V \nabla \bullet \vec{E} dV = \int_V \frac{\rho}{\epsilon_0} dV = \frac{Q}{\epsilon_0},$$

where Q is the total charge within the volume bounded by the surface. Thus, we have the statement of Gauss' law:

$$\oint_S \vec{E} \bullet \hat{n} dA = \frac{Q_{\text{inside } S}}{\epsilon_0}.$$

This equation is useful for computing \vec{E} from charge distributions that possess enough symmetry that \vec{E} can be taken out of the integral on the left-hand-side.

Problem 1: \vec{E} of a point charge.

Consider a point charge of magnitude q at the origin. Symmetry demands that the electric field be radial, i.e., $\vec{E}(\vec{r})$ is in the \hat{r} direction. Imagine a sphere of radius R around the origin. Clearly, symmetry also requires that the electric field has the same magnitude everywhere on the surface of the sphere. Thus, symmetry requires that

$$\vec{E}(\vec{r}) = E(r)\hat{r}.$$

Now, on this sphere, $\hat{n} = \hat{r}$, so the integral in Gauss' law is easily evaluated:

$$\oint_S \vec{E} \bullet \hat{n} dA = E(R) \oint_S dA = E(R)4\pi R^2.$$

Thus, $E(R) = q/(4\pi\epsilon_0 R^2)$, showing that the field at \vec{r} due to a point charge at the origin is

$$E(\vec{r}) = \frac{q}{4\pi\epsilon_0 r^2} \hat{r}.$$

Problem 2: \vec{E} of a spherical ball of uniform charge density.

Consider a spherical ball of charge with uniform density ρ centered at the origin and having a radius R . Again, symmetry demands that $\vec{E}(\vec{r}) = E(r)\hat{r}$. Imagine a sphere of radius $r < R$, also centered on the origin, and apply Gauss' law to this sphere:

$$4\pi r^2 E(r) = \frac{1}{\epsilon_0} \int_V \rho dV = \frac{1}{\epsilon_0} \rho \int_V dV = \frac{1}{\epsilon_0} \rho \frac{4\pi}{3} r^3$$

or

$$\vec{E}(\vec{r}) = \frac{\rho \vec{r}}{3\epsilon_0}.$$

Thus, \vec{E} is radial and *increases* with r inside the ball. To find the field outside the ball, consider an imaginary sphere with $r > R$ and do the same thing. Regardless of how large we make r , the charge inside is the total charge $Q = \rho 4\pi R^3/3$, so the field is that of a point charge with $q = Q$.

Problem 3: \vec{E} in a spherical cavity inside spherical ball of uniform charge density.

Consider a ball of charge with uniform charge density everywhere except for a completely enclosed spherical cavity that is free of charge. The ball and the cavity are not necessarily concentric. To find the field in the cavity we will use the results of Problem 2 in conjunction

with the principle of superposition. Clearly, the situation we are examining is identical to the superposition of a ball of uniform charge density ρ with a smaller ball of charge density $-\rho$ embedded within the larger sphere, i.e., the charge density in the region common to both spheres is $\rho - \rho = 0$. Now the field within each sphere due to the charge carried by the individual spheres is

$$\vec{E}_+(\vec{r}) = \frac{\rho\vec{r}_+}{3\epsilon_0} \quad \text{and} \quad \vec{E}_-(\vec{r}) = -\frac{\rho\vec{r}_-}{3\epsilon_0},$$

where \vec{r}_+ is a vector from the center of the $+\rho$ sphere to any point *within* it, and \vec{r}_- is a vector from the center of the $-\rho$ sphere to any point within it. Note, for the formulas to be valid, the points to which \vec{r}_\pm refer must be *inside* the respective spheres. Consider a point that is inside the $-\rho$ sphere, and therefore inside the cavity. The superimposed fields there are

$$\vec{E}(\vec{r}) = \frac{\rho\vec{r}_+}{3\epsilon_0} - \frac{\rho\vec{r}_-}{3\epsilon_0} = \frac{\rho}{3\epsilon_0}(\vec{r}_+ - \vec{r}_-)$$

The difference vector

$$\vec{r}_+ - \vec{r}_- \triangleq \vec{s}$$

is simply a vector drawn from the center of the original ball to the center of the spherical cavity. Thus, we arrive at the remarkable result that the field within the cavity is *uniform* and its direction is everywhere parallel to a vector drawn from the *center* of the sphere to the *center* of the cavity, i.e.,

$$\vec{E}_{\text{Cavity}} = \frac{\rho\vec{s}}{3\epsilon_0}.$$

Note that if the sphere and the cavity *are* concentric, $\vec{s} = 0$, and there is no field inside the cavity.

Problem 4: \vec{E} of a dipole.

A dipole consists of positive and negative point charges (same magnitude) separated by a small distance s . For specificity, let a charge q be placed at a position $(0, 0, +s/2)$ and a charge $-q$ at $(0, 0, -s/2)$ then, superimposing the fields of these charges (Problem 1) and defining the dipole moment to be $\vec{p} \triangleq qs\hat{k}$ (note, $s\hat{k}$ is a vector from the *negative* to the *positive* charge), \vec{E} at a point specified by \vec{r} , with $r \gg s$, is given by

$$\vec{E}(\vec{r}) \approx \frac{1}{4\pi\epsilon_0} \left[-\frac{\vec{p}}{r^3} + \frac{3\vec{r}(\vec{r} \bullet \vec{p})}{r^5} \right]. \quad (1.12)$$

The approximation becomes exact in the limit that $s \rightarrow 0$ and $q \rightarrow \infty$ such that $sq \rightarrow p$ (this is a “point” dipole). Note that while the field of a point charge falls off as the square of the distance from the charge, the field of a dipole falls off as the cube of the distance. More generally, the dipole moment of a charge distribution specified by a charge density $\rho(\vec{r})$ is $\vec{p} = \int \rho(\vec{r})\vec{r}dV$, and far from the charge distribution the field is given by Eq. (1.12).

Armed with these four simple results, we are now able to proceed to the study of dielectrics.

1.4.2 Dielectrics

We know what happens when individual charges are subjected to an electromagnetic field. They feel a force given by the Lorentz force law. However, atoms and molecules in normal matter are uncharged. What happens when ordinary matter is subjected to electric and magnetic fields? To try to understand this, we need a model of an atom. Consider atomic hydrogen. The structure of this atom is determined by quantum theory to consist of a proton of charge $+e$ surrounded by an electron cloud of charge density

$$\rho(\vec{r}) = \frac{-e}{\pi a_0^3} \exp[-2r/a_0] \quad (1.13)$$

in the ground state. In this equation, a_0 is the “Bohr” radius (0.529×10^{-10} m) and e is the charge on the electron (1.6×10^{-19} C). In excited states, the charge is generally distributed farther from the proton (and is not always spherically symmetric). For the purpose of understanding the influence of fields on a hydrogen atom, we will use a slightly simpler model for ease of computation: we assume the electron cloud is a sphere of radius R_0 and uniform charge density

$$\rho = -\frac{3e}{4\pi R_0^3},$$

and that excited states are similar, but with larger R_0 . Now, place this model atom in a uniform electric field \vec{E}_0 and assume that the electron cloud maintains its shape and uniform charge distribution. Clearly, the result of the field is to force the proton in the direction of the field and the electron cloud in the opposite direction. Equilibrium will be reached when the force on the proton from the external field is balanced by the force from the electron cloud. If the proton has moved a distance r_e relative to the cloud, the cloud’s \vec{E} at the proton’s position is given by Problem 2 above: $\vec{E}(\vec{r}) = (\rho\vec{r}_e/3\epsilon_0)$. The force on the proton by the electron cloud is $e\vec{E}(\vec{r})$ and must be balanced by the force resulting from \vec{E}_0 , i.e.,

$$e\vec{E}_0 - e\frac{3e}{4\pi R_0^3}\frac{\vec{r}_e}{3\epsilon_0} = 0,$$

or

$$e\vec{r}_e = 4\pi\epsilon_0 R_0^3 \vec{E}_0.$$

Thus, at equilibrium the proton is positioned a distance \vec{r}_e from the center of the electron cloud. Now, the total electric field in the medium is the sum of \vec{E}_0 and the field from the atom. Far from the atom, the electron cloud appears as a point charge (Problem 2), so

outside the atom the field is that of two point charges ($+e$ and $-e$) separated by \vec{r}_e , i.e., a dipole of moment $\vec{p} = e\vec{r}_e$. Placing the neutral atom in a uniform electric field caused it to acquire a dipole moment proportional to the field:

$$\vec{p} = 4\pi\epsilon_0 R_0^3 \vec{E}_0 \triangleq \alpha \vec{E}_0,$$

where α is called the *polarizability*. The field of this dipole is given in Problem 4, where it is seen to be proportional to the inverse cube of the distance from the atom. The total field

$$\vec{E}(\vec{r}) = \vec{E}_0 + \frac{1}{4\pi\epsilon_0} \left[-\frac{\vec{p}}{r^3} + \frac{3\vec{r}(\vec{r} \bullet \vec{p})}{r^5} \right]. \quad (1.14)$$

rapidly reduces to \vec{E}_0 with increasing \vec{r} . The polarizability,

$$\alpha = 4\pi\epsilon_0 R_0^3,$$

given by this model is approximately 4.5 times smaller than the actual measured value for hydrogen if R_0 is taken to be a_0 . Were the more correct Eq. (1.13) used for the charge density the result would be 3/4 of that above, i.e., farther from the measured value. Evidently, there is significant distortion of the electron cloud when the atom is placed in a field; however, this simple model does provide the essential physics as well as the correct order of magnitude for α .

It is important to note that for typical values of \vec{E}_0 , \vec{r}_e is very small. For our model,

$$\frac{r_e}{R_0} = \frac{4\pi\epsilon_0 R_0^2}{e} E_0 = \frac{E_0}{E(R_0)},$$

where $E(R_0)$ is the magnitude of the proton's electric field a distance R_0 away, e.g., at the edge of the electron cloud. Taking $R_0 = a_0$, gives $E(R_0) = 5.25 \times 10^{11}$ N/C, so

$$\frac{r_e}{R_0} \approx 1.8 \times 10^{-12} E_0.$$

How large is this when E_0 is the field due to an electromagnetic wave in the visible spectrum? Consider a 5 mW He-Ne laser typically used as a pointer. If it has a beam cross section of 4 mm^2 , the average Poynting vector is 1250 Watts/m^2 . This results in an electric field of 970 N/C (Eq. (??)), so $r_e/R_0 \approx 1.8 \times 10^{-9}$! Furthermore even at the fields high enough to cause dielectric breakdown of air ($\sim 10^6 \text{ N/C}$), i.e., a spark during which air ceases to be a dielectric and becomes a conductor, $r_e/R_0 \approx 1.8 \times 10^{-6}$. Thus, the proton (or nucleus in a more complex atom) never moves a sizable fraction of the atomic radius, and approximation of the atom as a point dipole is very realistic.

If a sufficiently dilute gas of such atoms is subjected to an external field \vec{E}_0 , except for rare near collisions, the field experienced by each atom is \vec{E}_0 . This causes each atom to

develop a dipole moment \vec{p} . If n is the number density (atoms/m³), then in a volume dV there will be ndV dipoles, and the dipole moment $d\vec{p}$ of dV will be $\vec{p}n dV$. It is useful to define the dipole moment density as

$$\vec{P} \triangleq \frac{d\vec{p}}{dV} = n\vec{p}.$$

\vec{P} is called the *polarization* of the medium, and for the dilute gas,

$$\vec{P} = n\alpha\vec{E}_0.$$

Note that when the gas is not dilute, or when we are considering a liquid or solid dielectric, then the second term in Eq. (1.14) will contribute, i.e., the one including $\vec{r} \bullet \vec{p}$, so the field seen by each dipole will not be the externally applied field \vec{E}_0 .

If we have a region of space containing a large number of dipoles, what electric field do these dipoles produce? It is easy to compute the \vec{E} field produced by a sphere of uniform polarization \vec{P} . Such a polarization would develop if a ball of uniform positive charge density and an identical ball of uniform negative charge density were almost, but not quite perfectly superimposed. If their centers are displaced from one another by an amount \vec{s} , where \vec{s} is pointed from the negative to the positive charge, then each element of volume dV will develop a dipole moment³

$$d\vec{p} = \vec{s}dq = \vec{s}\rho dV.$$

The corresponding polarization in is $\vec{P} = \vec{s}\rho$ and it is uniform within the sphere. The electric field in the region common to the two spheres is (Problem 3) $\vec{E} = -\rho\vec{s}/3\epsilon_0$, so

$$\vec{E} = -\frac{\vec{P}}{3\epsilon_0},$$

and the electric field in a sphere of uniform polarization is also uniform and is directed opposite to \vec{P} . Note that the minus sign above arises because in Problem 3, \vec{s} was drawn from the center of the positive sphere to the center of the negative sphere, while \vec{s} here is in the opposite direction.

As there is no net charge anywhere inside the sphere, the reader may be wondering where this field originates. The answer is that there are charges at the polar caps of the sphere (Figure 1.1). If we take the north pole of the sphere to be that toward which \vec{P} is directed, then when the centers of the positive and negative spheres are offset by \vec{s} , there

³Note that in Problem 3, \vec{s} was a vector from the center of the positively charged sphere to the center of the negatively charged sphere. Here, the dipole moment vector $\vec{p} = q\vec{s}$ is from the negative charge to the positive charge, and hence here \vec{s} is the negative of \vec{s} in Problem 3.

will be a cap of positive charge on the northern polar surface and a cap of negative charge on the southern polar surface. Recalling that in a real atom \vec{s} is typically a small fraction of the size of the atom, these charges appear to be residing on the surface of the sphere. We now examine their properties.

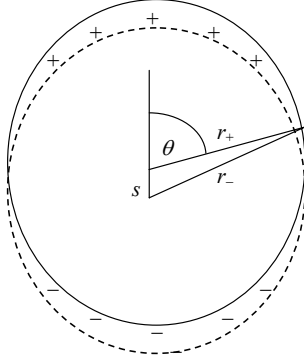


Figure 1.1: Spheres of uniform positive and negative charge density are superimposed and the positive sphere is shifted vertically through a distance s relative to the negative sphere. This results in thin positively and negatively charged polar caps.

Consider the geometry of this situation. To compute the charge on the upper cap in Figure 1.1, we need to compute its thickness. Since the cap is thin, this is the difference between r_+ and r_- . But,

$$r_-^2 = r_+^2 + s^2 + 2r_+s \cos \theta,$$

so, noting that $s \ll r_+$ or r_- ,

$$r_- - r_+ \approx s \cos \theta.$$

If we look at a small area dA on the sphere, the charge dq in the volume $(r_- - r_+)dA$ is $dq = \rho s dA \cos \theta$. As this charge is on the surface of the sphere, it is useful to define the surface charge density σ according to $\sigma \triangleq dq/dA$. Then $\sigma(\theta) = \rho s \cos \theta$. But $\rho = Q/V$, where Q is the total charge on the sphere, so $\rho s = Qs/V = P$, since Qs is the total dipole moment of the system of spheres. Therefore, $\sigma = P \cos \theta = \vec{P} \cdot \hat{n}$ and the effect of the polarization is to cause charge to appear on the surface with a charge density σ . Note that this formula automatically provides correct sign for the surface charge density on the southern

cap.

The above result can be shown to be quite general, i.e., a polarized dielectric will have a charge density

$$\sigma_b = \vec{P} \cdot \hat{n}$$

on its surface. The subscript “ b ” on σ_b is remind us that, unlike charges on conductors that can be manipulated at will using tools such as batteries (and usually referred to as “free” charges), the charges on the surface of a dielectric are bound to the individual atoms. This charge is usually referred to a “bound” charge. Noting that the dielectric is electrically neutral, i.e., no charges have left their individual (neutral) atoms, any volume density ρ_b of bound charge within the dielectric must satisfy

$$\begin{aligned} 0 &= \int_V \rho_b dV + \oint_S \sigma_b dA \\ &= \int_V \rho_b dV + \oint_S \vec{P} \cdot \hat{n} dA \end{aligned}$$

$$= \int_V \rho_b dV + \int_V \nabla \bullet \vec{P} dV.$$

Since V is arbitrary, this requires that

$$\rho_b = -\nabla \bullet \vec{P}.$$

ρ_b and σ_b provide the source of the field, given the polarization of the dielectric.

Typically, we will be interested in electromagnetic waves within dielectrics. In this case \vec{E} will vary in time causing \vec{P} to vary in time, which results in a current. To see how this happens, consider a simple case of a slab of material oriented normal to an electric field \vec{E} . This field causes a polarization \vec{P} in the material in the same direction as \vec{E} , and results in a charge density $\sigma_b = \vec{P} \bullet \hat{n}$ on the surfaces. As the field changes, \vec{P} changes and so does σ_b . If we examine an area dA of the surface, there will be a current $I = dq/dt$ in the column of length ℓ and base dA . The current density in the column is $j = I/dA$, so

$$j = \frac{dq}{dt dA} = \frac{d\sigma_b}{dt} = \frac{dP}{dt}.$$

But \vec{j} is in the same direction as increasing \vec{P} , so we have

$$\vec{j}_b = \frac{d\vec{P}}{dt}$$

as the current density in a dielectric medium with time-varying polarization. (Note as this is also “bound”, we use the subscript b to differentiate it from currents associated with free charge, i.e., “free” currents.)

As an aside, we expect all currents and their associated charge densities to satisfy the continuity equation:

$$\nabla \bullet \vec{j} + \frac{\partial \rho}{\partial t} = 0.$$

Is this true of \vec{j}_b ? Substituting directly for \vec{j}_b and ρ_b shows that this is indeed the case.

1.4.3 Dense Collection of Atoms in an \vec{E} Field

Maxwell’s equations are generally applicable to charges in a vacuum. One is led to wonder then, what fields actually satisfy Maxwell’s equations in matter? After all, in a medium, the field at the microscopic level can be very small, i.e., negligible, outside an atom, but a small distance away near a nucleus \vec{E} is enormous. Thus, in any real medium, \vec{E} fluctuates rapidly and strongly with position. The answer is that the fields in Maxwell’s equations are spatial

averages of the microscopic fields over volumes that include many atoms, e.g., a sphere 100 Angstroms in diameter (for a discussion, see Jackson, *Classical Electrodynamics*). One should not find this mysterious, as one would also average over a similar scale to determine the mass density, which at smaller scales would also fluctuate wildly with position. In such determinations, in essence, matter is being treated as a continuous distribution of mass (or charge). A region of this size will have well-defined charge and current densities (based on similar averaging). When we use the term electric field it will refer to this spatially-averaged field; however, at times when emphasis is needed, we refer to it as the “macroscopic” field. When a dilute medium is subjected to an external electric field the spatially-averaged field and the microscopic field (the actual field at a given point in space) will be virtually identical. This means that when a dilute gas is placed in a field \vec{E}_0 , the field seen by each atom will be \vec{E}_0 , and the development of \vec{P} easy to understand. In contrast, in a dense medium there is no reason to believe that the field seen by an individual atom $\vec{E}_{\text{Microscopic}}$ is equal to the field \vec{E} provided by Maxwell’s equations in the medium. That is, at the molecular scale, the field $\vec{E}_{\text{Microscopic}}$ seen by a single atom (or molecule) will depend on the graininess of matter in the *vicinity* of the atom — the continuum approximation must break down at the atomic scale.

It is customary to relate the polarization of the medium \vec{P} to the macroscopic field in the medium \vec{E} , noting of course that \vec{E} is *not* the field that is externally impressed on the dielectric (as it *is* for a dilute gas). That is, \vec{E} includes any fields created by \vec{P} as well. It is the actual (averaged) electric field in the medium: the macroscopic field. For many dielectrics (called linear-homogeneous-isotropic dielectrics) the relationship,

$$\vec{P} = \epsilon_0 \chi_e \vec{E}, \quad (1.15)$$

is satisfied to a good approximation. The constant χ_e is called the electric susceptibility of the dielectric. This states that the polarization is proportional to the macroscopic field in the dielectric. The question is how do we relate χ_e , which is a bulk property of the dielectric, to α , which is a parameter associated with the individual atoms or molecules?

Consider a slab of dielectric such as shown on the left-hand-side of the equal sign in Figure 1.2, and consider the field at a particular atom in the medium, $\vec{E}_{\text{Microscopic}}$. We expect that the contribution of the near-by atoms to this microscopic field will have to be computed separately from the atoms far away. Thus, we imagine a spherical cavity of radius R containing the atom of interest at its center and compute the field at the center due to the atoms inside the cavity separately and add to it the field at the center due to the rest of the medium. The latter field is treated in the continuum approximation.

Using the principle of superposition (Figure 1.2) the field in the cavity due to \vec{P} outside is easy to compute. On the left-hand-side of the equal sign on the figure, the field at the center of the cavity is just \vec{E} , the macroscopic field. In the first term on the right-hand-side

the field at the center of the cavity is \vec{E}_{Cavity} , while in the second it's the field inside a uniformly polarized sphere which we have found to be $-\vec{P}/3\epsilon_0$. Thus

$$\vec{E} = \vec{E}_{\text{Cavity}} - \vec{P}/3\epsilon_0$$

or

$$\vec{E}_{\text{Cavity}} = \vec{E} + \vec{P}/3\epsilon_0.$$

i.e., the field in the cavity is actually *larger* than the field in the medium. Now, to \vec{E}_{Cavity} we add the sum of the fields due to the individual dipoles within the cavity. These individual dipole fields are given by

$$\vec{E}_{\text{Dipole}}(\vec{r}) = \frac{1}{4\pi\epsilon_0} \left[-\frac{\vec{p}}{r^3} + \frac{3\vec{r}(\vec{r} \bullet \vec{p})}{r^5} \right],$$

where \vec{p} is in the same direction as \vec{P} , and $-\vec{r}$ is the position vector of an individual dipole referenced to the center of the sphere, i.e., relative to the atom at which we wish to compute the field. Clearly, the superposition of the fields due to the individual dipoles is dependent on how they are distributed within the cavity. However, there are two cases in which the dipole field is easy to compute: a cubic lattice of identical atoms (as in a crystal) and a random distribution of atoms (as in a liquid). For these cases the total dipole field actually vanishes, and the field seen by the molecule $\vec{E}_{\text{Microscopic}}$ is just \vec{E}_{Cavity} , i.e.,

$$\vec{E}_{\text{Microscopic}} = \vec{E} + \vec{P}/3\epsilon_0.$$

However, it must be noted that this is not always the case. For example, in a calcite crystal the atoms are not arranged in a cubic lattice and the dipole field does not vanish. In fact \vec{E} and \vec{P} are not even in the same direction in calcite (Eq.(1.15) does not hold), which leads to the phenomenon of double refraction.

The dipole moment of individual atoms is related to the microscopic field through

$$\vec{p} = \alpha \vec{E}_{\text{Microscopic}} = \alpha (\vec{E} + \vec{P}/3\epsilon_0),$$

or using the relationship between \vec{E} and \vec{P} (Eq.(1.15)),

$$\vec{p} = \alpha \vec{E} \left(1 + \frac{\chi_e}{3} \right).$$

But, $\vec{P} = n\vec{p}$, where n is the number density of atoms, so

$$\vec{P} = n\alpha \vec{E} \left(1 + \frac{\chi_e}{3} \right)$$

Using Eq.(1.15) to eliminate \vec{P} , we have the final result:

$$\alpha = \frac{\epsilon_0 \chi_e / n}{1 + \chi_e / 3}$$

or, defining the dielectric constant $K_e \triangleq (1 + \chi_e)$,

$$\alpha = \frac{3\epsilon_0}{n} \left[\frac{K_e - 1}{K_e + 2} \right]. \quad (1.16)$$

This relationship between the dielectric's microscopic (α) and macroscopic (K_e) properties is referred to as the Clausius-Mossotti or the Lorentz-Lorenz law. In a dilute gas, $\chi_e \ll 1$ and $K_e \approx 1$, so

$$\chi_e = \frac{n\alpha}{\epsilon_0} \quad \text{and} \quad K_e = 1 + \frac{n\alpha}{\epsilon_0},$$

which corresponds to

$$\vec{E}_{\text{Microscopic}} = \vec{E}.$$

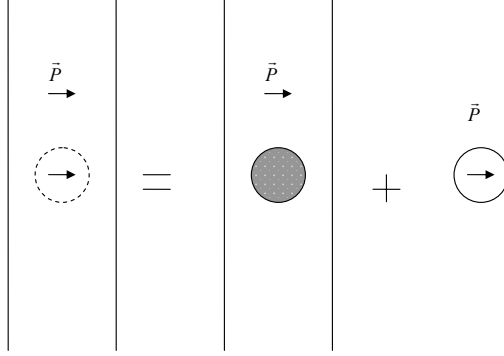


Figure 1.2: The field in the slab of a dielectric with uniform polarization \vec{P} on the left can be formed from a superposition of the field from the slab having a spherical cavity and a sphere of the same polarization sized to just fit into the cavity. This is used in the text to find the field in the cavity.

Some molecules, e.g. water, have permanent dipole moments because of their internal distribution of charge. Consider what happens when a dipole with moment \vec{p} is placed in a field \vec{E} such that there is an angle between \vec{E} and \vec{p} . We model the dipole as a positive charge q separated from a negative charge $-q$ by a small distance. In the field, $+q$ is forced in the direction \vec{E} and $-q$ is forced in the opposite direction. This tends to rotate the dipole so that \vec{p} aligns with \vec{E} , i.e., the dipole is subjected to a torque (given by $\vec{p} \times \vec{E}$). Thus, when an external field is applied to such a medium the dipoles tend to align with the field and the medium develops a polarization \vec{P} . Such media are similar in their properties to the dielectrics we have been discussing above.

1.4.4 Electromagnetic Waves in Dielectrics

In a dielectric medium, the current and charge densities are bound and given by, respectively,

$$\vec{j}(\vec{r}, t) = \frac{\partial \vec{P}(\vec{r}, t)}{\partial t} \quad \text{and} \quad \rho = -\nabla \cdot \vec{P}(\vec{r}, t),$$

so Maxwell's equations become

$$\begin{aligned}\nabla \bullet [\epsilon_0 \vec{E}(\vec{r}, t) + \vec{P}(\vec{r}, t)] &= 0 \\ \nabla \bullet \vec{B}(\vec{r}, t) &= 0 \\ \nabla \times \vec{E}(\vec{r}, t) &= -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t} \\ \nabla \times \vec{B}(\vec{r}, t) &= \mu_0 \frac{\partial}{\partial t} [\epsilon_0 \vec{E}(\vec{r}, t) + \vec{P}(\vec{r}, t)]\end{aligned}$$

The quantity $\epsilon_0 \vec{E} + \vec{P}$ is usually called the electric *displacement* \vec{D} . Note that in a linear-homogeneous-isotropic dielectric

$$\epsilon_0 \vec{E} + \vec{P} = \epsilon_0(1 + \chi_e)\vec{E} = \epsilon_0 K_e \vec{E}.$$

Then, performing the same operations on Maxwell's equations as before in the case of electromagnetic waves in free space, we find

$$\nabla^2 \vec{E} - \mu_0 \epsilon_0 K_e \frac{\partial^2 \vec{E}}{\partial t^2} = 0. \quad (1.17)$$

and

$$\nabla^2 \vec{B} - \mu_0 \epsilon_0 K_e \frac{\partial^2 \vec{B}}{\partial t^2} = 0, \quad (1.18)$$

i.e., the same equations as found for free space, but with ϵ_0 replaced by $\epsilon_0 K_e$. The propagation is rectilinear, and the only change in moving from free space to a dielectric is that the speed of the waves is now given by

$$v^2 = \frac{1}{\mu_0 \epsilon_0 K_e} = \frac{c^2}{K_e}.$$

The index of refraction m of the medium is defined to be $m \triangleq c/v$, so for a dielectric, $m = \sqrt{K_e}$.

1.4.5 Dynamic Properties of Dielectrics

When an electromagnetic wave propagates in a dielectric, the \vec{E} field is varying rapidly in time. We have developed the theory of dielectrics assuming that the field was static. Are the dielectric properties modified by the time variation of \vec{E} ? To answer this question, we need to employ a model of the atom. The correct description requires quantum mechanics; however, it is possible to develop a *qualitative* understanding of most of the effects of the time variation of the fields from the very simple classical model. We assume that the model

of the atom described at the beginning of our discussion of dielectrics is applicable. For atomic hydrogen this is a proton at the center of a spherical electron cloud of radius R_0 and uniform charge density. We assumed that the application of an external electric field did not distort the electron cloud, but caused its center to move relative to the positive charge (that was originally in equilibrium at the center of the sphere). If the positive charge is moved a distance r from the center of the electron sphere, it will be subjected to a restoring force of magnitude

$$F = \frac{e^2 r}{4\pi\epsilon_0 R_0^3}.$$

The same force will act on the electron cloud, but because of its small mass compared to the proton, the motion of the cloud will dominate that of the proton. Letting the vector \vec{r} describe the position of the center of the electron cloud from the proton, the classical equation of motion (Newton's law) of the cloud when the atom is subjected to an oscillating electric field of amplitude \vec{E}_0 is

$$m_e \frac{d^2 \vec{r}}{dt^2} = -\frac{e^2 \vec{r}}{4\pi\epsilon_0 R_0^3} - m_e \gamma \frac{d\vec{r}}{dt} - e \vec{E}_0 \cos(\omega t), \quad (1.19)$$

where m_e is the mass of the electron. This equation will lead to oscillation of the atom, in fact the atom will become an oscillating dipole. As we shall see later, such a dipole emits electromagnetic radiation, and therefore must continually lose energy, so we have introduced such an energy loss as a damping (friction) term,⁴ the frictional force being $-m_e \gamma \vec{v}$. In addition, the material may absorb radiation which also provides a loss mechanism. Rewriting,

$$\frac{d^2 \vec{r}}{dt^2} + \gamma \frac{d\vec{r}}{dt} + \omega_0^2 \vec{r} = \frac{-e}{m_e} \vec{E}_0 \cos(\omega t),$$

where $\omega_0^2 = e^2/(4\pi\epsilon_0 m_e R_0^3)$. To solve this equation, we assume that the electric field is in the x direction, i.e., $\vec{E}_0 = \hat{i}E_0$. Then the y and z motions will be unforced, and if initially excited (by some other process) will damp out, so only the x component of \vec{r} is of interest.⁵ Also, the “transient” solution (solution for $E_0 = 0$) for the x component will damp out as well, so we only need consider the “steady state” solution of

$$\frac{d^2 x}{dt^2} + \gamma \frac{dx}{dt} + \omega_0^2 x = \frac{-e}{m_e} E_0 \cos(\omega t).$$

⁴In a phenomenological treatment of radiation damping, the constant γ can be shown to be $e^2 \omega_0^2 / (6\pi\epsilon_0 m_e c^3)$.

⁵The actual force on a charge is $\vec{F} = q[\vec{E} + \vec{v} \times \vec{B}]$, and the magnetic term can yield forces in directions in addition to x ; however, for an electromagnetic wave $\vec{B} = \hat{k} \times \vec{E}/c$, so the magnetic force is of order v/c times the electric force. Thus, ignoring the magnetic force is justified when $v \ll c$, i.e., when the motion is non-relativistic. Using the example of a 5 mW He-Ne laser discussed earlier, we find $|v/c| \sim 0.01$, and in that case the magnetic force can be safely ignored compared to the electric force.

Replacing $\cos(\omega t)$ by $\exp(-i\omega t)$ and treating x as a complex variable, the solution is

$$x = \frac{(-e/m_e)E_0}{(\omega_0^2 - \omega^2) - i(\gamma\omega)} \exp(-i\omega t).$$

The dipole moment of the atom is then $\vec{p} = -e\vec{r}$, so $\vec{p} = -ex\hat{i}$, and writing the dipole moment as $\vec{p} = \vec{p}_0 \exp(-i\omega t)$, we have

$$\vec{p}_0 = \frac{(e^2/m_e)}{(\omega_0^2 - \omega^2) - i(\gamma\omega)} \vec{E}_0.$$

This implies that \vec{E} and \vec{p} are out of phase, i.e.,

$$\vec{p}_0 = \frac{(e^2/m_e)}{\sqrt{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}} \vec{E}_0 \exp(i\phi),$$

where the phase angle is

$$\phi = \tan^{-1} \left[\frac{\gamma\omega}{\omega_0^2 - \omega^2} \right].$$

Finally, since $\vec{p} = \alpha \vec{E}$, we have

$$\alpha = \frac{(e^2/m_e)}{(\omega_0^2 - \omega^2) - i(\gamma\omega)} = \frac{(e^2/m_e)}{\sqrt{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}} \exp(i\phi), \quad (1.20)$$

Clearly, for low frequencies fields $\omega \ll \omega_0$, $\phi = 0$ and the electron cloud oscillates in phase with \vec{E} , while for high frequencies $\omega \gg \omega_0$, $\phi = \pi$. When $\omega = \omega_0$, i.e., when the frequency of the field is the “natural” frequency of the atom, $\phi = \pi/2$.

The fact that the polarizability is complex (and therefore not in phase with the applied field), means that the dielectric constant will be complex as well. Consider a dilute gas, for which the individual atoms see the externally applied field. Then

$$K_e = 1 + \frac{n\alpha}{\epsilon_0}.$$

Using α from above, and writing the *complex* dielectric constant as $K_e = K_e^r + iK_e^i$, we have

$$K_e^r = 1 + n \frac{e^2(\omega_0^2 - \omega^2)/m_e\epsilon_0}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \quad (1.21)$$

and

$$K_e^i = n \frac{e^2\gamma\omega/m_e\epsilon_0}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}. \quad (1.22)$$

In addition, as the refractive index $m = \sqrt{K_e}$ and $K_e = 1 + n\alpha/\epsilon_0$, noting that $n\alpha/\epsilon_0 \ll 1$, we have $m \approx 1 + n\alpha/2\epsilon_0$, and writing the complex refractive index $m = m_r + im_i$, we have from Eqs. (1.21) and (1.22)

$$m_r = 1 + n \frac{e^2(\omega_0^2 - \omega^2)/2m_e\epsilon_0}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \quad (1.23)$$

and

$$m_i = n \frac{e^2\gamma\omega/2m_e\epsilon_0}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}. \quad (1.24)$$

The variation of the real and imaginary parts of the refractive index with frequency is provided in Figure 1.3. In a dense medium, (1.16) can be rewritten

$$K_e = 1 + \frac{n\alpha/\epsilon_0}{1 - n\alpha/3\epsilon_0},$$

and on inserting α we find that Eq. (1.21) through (5.18) still apply if ω_0^2 is replaced by $\omega_0^2 - ne^2/(3m_e\epsilon_0)$.

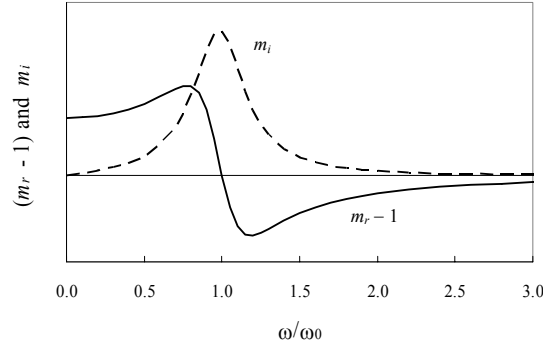


Figure 1.3: Variation of the real part m_r (solid curve) and the imaginary part m_i (dashed curve) of the complex refractive index with frequency ω near the natural frequency ω_0 of an atom of a dilute gas.

1.5 Electromagnetic Wave Propagation in a Dielectric

We have seen that propagation in a dielectric is equivalent to that in a vacuum if the speed of the vacuum waves c is replaced by c/m , where $m^2 = K_e$ is the refractive index of the

medium. Thus, where in a vacuum, for plane waves propagating in the $\hat{\kappa}_v$ direction, we had

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \exp[i(\vec{\kappa}_v \bullet \vec{r} - \omega t)],$$

with $\kappa_v = \omega/c$, we now have

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \exp[i(\vec{\kappa}_m \bullet \vec{r} - \omega t)],$$

with $\kappa_m = m\omega/c = m\kappa_v$. The subscript m on κ stands for “medium,” and the subscript v stands for vacuum. Using the complex refractive index in the above gives

$$\begin{aligned} \vec{E}(\vec{r}, t) &= \vec{E}_0 \exp[i(m_r \vec{\kappa}_v \bullet \vec{r} - \omega t) - m_i \vec{\kappa}_v \bullet \vec{r}] \\ &= \vec{E}_0 \exp(-m_i \vec{\kappa}_v \bullet \vec{r}) \exp[i(m_r \vec{\kappa}_v \bullet \vec{r} - \omega t)]. \end{aligned}$$

This is clearly a wave with an amplitude that decays exponentially with increasing $\vec{\kappa}_v \bullet \vec{r}$. As before,

$$\vec{B}_0 = \frac{m \vec{\kappa}_v \times \vec{E}_0}{\omega},$$

but since m is complex, i.e.,

$$m = \sqrt{m_r^2 + m_i^2} \exp(i\phi_m) = |m| \exp(i\phi_m), \quad \text{where } \phi_m = \tan^{-1}(m_i/m_r),$$

\vec{B} and \vec{E} are out of phase by ϕ_m . The time-averaged Poynting vector is easily seen to be

$$\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} = \frac{|m| \vec{\kappa}_v}{2\omega\mu_0} E_0^2 \cos(\phi_m) \exp(-2m_i \vec{\kappa}_v \bullet \vec{r}), \quad (1.25)$$

and the time-average energy flux decreases as the wave propagates through the medium. Note that $\zeta \triangleq \hat{\kappa} \bullet \vec{r}$ is the distance coordinate measured along the direction of propagation. Then, we see that

$$2m_i \vec{\kappa}_v \bullet \vec{r} = 2m_i \kappa_v \zeta = \frac{4\pi m_i}{\lambda_v} \zeta \triangleq a\zeta.$$

The absorption coefficient a is

$$a = \frac{4\pi m_i}{\lambda_v}, \quad (1.26)$$

where λ_v is the wavelength on the wave in vacuum. (Note that in the medium, the waveform as a function of position is not sinusoidal, it is the product of a sinusoid and a decreasing exponential. In fact, it is not even periodic, i.e., it never repeats. Because of this the notion of a wavelength *in the medium* is somewhat ambiguous unless m_i is small, so the wave form is almost periodic. However, the wavelength would be $\lambda_m \triangleq \lambda_v/m_r$ if m_i were exactly zero, and this is traditionally defined to be the wavelength in the medium.) Thus,

as the wave propagates along $\hat{\kappa}$, it is attenuated according to $\exp(-a\zeta)$. An alternate way of writing the absorption coefficient is clearly

$$a = -\frac{1}{\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}} \frac{d\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}}{d\zeta}, \quad (1.27)$$

i.e., the fractional loss in the energy flux per unit distance in the propagation direction. Also, the fractional loss in $\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}$ traversing a distance of $\lambda_v/4\pi$ is m_i , i.e.,

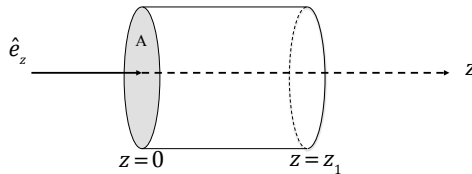
$$m_i = \frac{\lambda_v a}{4\pi}.$$

Figure 1.3 shows how m_i varies in the vicinity of ω_0 for a dilute gas. Clearly the electromagnetic wave will be strongly attenuated (large m_i and a) if the frequency of the wave is close to ω_0 . In contrast, there will be small absorption if ω differs significantly from ω_0 . The variation of m_r in the vicinity of $\omega = \omega_0$ is interesting in that the speed of the wave varies considerably with the frequency. This variation is called *dispersion*. Starting from $\omega \ll \omega_0$, m_r increases with increasing frequency, which is typical for most materials. A similar increase with frequency is seen for $\omega \gg \omega_0$. This is usually referred to as normal dispersion. However, when $\omega \approx \omega_0$, m_r *decreases* with increasing frequency. This behavior is usually termed *anomalous* dispersion. The width of the m_i maximum is determined by γ , and this constant is very small for most atomic absorption features. When γ is small, the absorption feature is very sharp, so the region of anomalous dispersion is narrow and usually difficult to observe in gases. However, in media such as liquids and solids, γ can be quite large and the anomalous dispersion is then easy to observe.

It is instructive to apply Poynting's theorem to this case where the electromagnetic waves undergo absorption. For simplicity, we again consider a wave propagating in the z direction; specifically, $\vec{\kappa} = \hat{e}_z \kappa$, so

$$\vec{E}(z, t) = \hat{e}_x E_0 \exp(-az/2) \cos(\kappa z - \omega t).$$

Let this wave be incident on a disk of area A as shown below, and calculate the total



electromagnetic energy within the cylinder at any time t :

$$U_{em}(t) = \int_V u_{em}(z, t) dV = \epsilon_0 \int_V \vec{E} \bullet \vec{E} dV = \epsilon_0 E_0^2 A \int_0^{z_1} \exp(-az) \cos^2(\kappa z - \omega t) dz.$$

Carrying out the integral yields

$$U_{em}(t) = \frac{\epsilon_0 E_0^2 A}{2(a^3 + 4a^2\kappa)} \left[\left(a^2 + 4\kappa^2 + a^2 \cos(2\omega t) + 2a\kappa \sin(2\omega t) \right) - \exp(-az_1) \left(a^2 + 4\kappa^2 + a^2 \cos(2\kappa z_1 - 2\omega t) + 2a\kappa \sin(2\kappa z_1 - 2\omega t) \right) \right].$$

The time-average of this energy is

$$\langle U_{em} \rangle = \frac{\epsilon_0 E_0^2 A}{2a} [1 - \exp(-az_1)],$$

and by a similar calculation,

$$\left\langle \frac{\partial U_{em}}{\partial t} \right\rangle = 0,$$

i.e., the time-average of the temporal variation in electromagnetic energy content of V is unchanging. Now, Poynting's theorem reads

$$\frac{dW}{dt} = -\frac{\partial U_{em}}{\partial t} - \oint_S S(\vec{r}, t) \bullet \hat{n} dA,$$

and taking the time average,

$$\left\langle \frac{dW}{dt} \right\rangle = -\left\langle \frac{\partial U_{em}}{\partial t} \right\rangle - \oint_S \langle S(\vec{r}, t) \rangle_{\text{Ave}} \bullet \hat{n} dA,$$

so the integral on the right is the rate at which electromagnetic energy is being converted to other forms. This further reinforces the interpretation that for an electromagnetic wave the Poynting vector is the flux of electromagnetic energy. Finally, we can compute this integral directly:

$$\begin{aligned} \oint_S \langle S(\vec{r}, t) \rangle_{\text{Ave}} \bullet \hat{n} dA &= \frac{c\epsilon_0 A E_0^2}{2} \exp(-az_1) - \frac{c\epsilon_0 A E_0^2}{2} \\ &= -\frac{c\epsilon_0 A E_0^2}{2} (1 - \exp(-az_1)) \\ &= -a c \langle U_{em} \rangle. \end{aligned} \tag{1.28}$$

Since $\langle \partial U_{em} / \partial t \rangle = 0$, we see that in this case,

$$\left\langle \frac{dW}{dt} \right\rangle = a c \langle U_{em} \rangle. \tag{1.29}$$

Thus, the time-average rate at which electromagnetic energy within V is being converted to other forms of energy, e.g., heat, is proportional to the time-average electromagnetic energy within V itself.

Although the simple classical model of the atom that we have been using to look at the behavior of α as a function of ω is unrealistic — quantum theory is required for a detailed quantitative description — it does provide a qualitative description of the variation of m_r and m_i in the vicinity of an atomic (or molecular) absorption feature. In addition, the coupling predicted between m_r and m_i is fundamental (it is actually *required* by causality), not just an artifact of the model. Quantum dynamics shows that a hydrogen atom can exist in an infinite number of discrete energy states e_i with each specified by a *wave function* ψ_i . Only radiation of a frequency ω_{ij} can be absorbed (or emitted), with

$$\omega_{ij} = \frac{2\pi(e_j - e_i)}{h},$$

where h is Planck's constant (6.62×10^{-34} J·s), and the state i with energy e_i has lower energy than the state j (with e_j). In addition, the correct quantum analog to Eq.(1.20) can be obtained the replacements

$$\gamma \rightarrow \gamma_{ij}, \quad \omega_0 \rightarrow \omega_{ij} \quad \text{and} \quad \frac{e^2}{m_e} \rightarrow \frac{4\pi\omega_{ij}|e\vec{r}_{ij}|^2}{3h},$$

where

$$\vec{r}_{ij} \triangleq \int \psi_i^*(\vec{r})\vec{r}\psi_j(\vec{r}) dV,$$

ψ_i and ψ_j are the wave functions of the two states. Note that each transition is assumed to have its own damping constant: γ_{ij} . Thus, we have

$$\alpha = \sum_{i < j} \frac{4\pi\omega_{ij}e^2|\vec{r}_{ij}|^2/(3h)}{(\omega_{ij}^2 - \omega^2) - i(\gamma_{ij}\omega)}.$$

It is customary to write this equation in terms of the so-called oscillator strengths f_{ij} defined as follows:

$$\alpha = \sum_{i < j} \frac{f_{ij}}{(\omega_{ij}^2 - \omega^2) - i(\gamma_{ij}\omega)},$$

in which case the refractive indices for a dilute gas (Eqs. (1.23) and (5.18)) become

$$m_r = 1 + \frac{1}{2\epsilon_0} \sum_{i < j} \frac{f_{ij}(\omega_{ij}^2 - \omega^2)}{(\omega_{ij}^2 - \omega^2)^2 + (\gamma_{ij}\omega)^2} (n_i - n_j), \quad (1.30)$$

and

$$m_i = \frac{1}{2\epsilon_0} \sum_{i < j} \frac{f_{ij}\gamma_{ij}\omega}{(\omega_{ij}^2 - \omega^2)^2 + (\gamma_{ij}\omega)^2} (n_i - n_j), \quad (1.31)$$

where n_i and n_j are the number densities of atoms in the lower and upper states, respectively. These number densities depend on the temperature of the medium. The oscillator strengths can be computed, in principle, for any atom (or molecule). Later, in our application, we will fit the observed m_i spectrum (determined by measuring a) to Eq. (1.31) to estimate f_{ij} 's, ω_{ij} 's, and γ_{ij} 's for each absorption feature. These estimates will then be used to estimate the m_r spectrum using Eq. (1.30). This is effected by noting that for a single absorption line, a term in the sum is significant only when $\omega \approx \omega_{ij}$, so

$$(\omega_{ij}^2 - \omega^2) = (\omega_{ij} + \omega)(\omega_{ij} - \omega) \approx 2\omega_{ij}(\omega_{ij} - \omega).$$

Using this, the contribution of a single term in the sums for m_i and $m_r - 1$ (Eq.s (1.30) and (1.31)) is

$$C_{ij} \frac{1}{1 + \eta_{ij}^2} \quad \text{and} \quad C_{ij} \frac{\eta_{ij}}{1 + \eta_{ij}^2},$$

respectively, where $\eta_{ij} \triangleq 2(\omega_{ij} - \omega)/\gamma_{ij}$, $C_{ij} \triangleq f_{ij}(n_i - n_j)/(\gamma_{ij}\omega_{ij})$, and we replaced $\gamma_{ij}\omega$ by $\gamma_{ij}\omega_{ij}$. Thus, if in a given region of the spectrum there are N absorption features, we can write

$$m_i = \sum_{i=1}^N \frac{C_i}{1 + \eta_i^2} \quad \text{and} \quad m_r - 1 = C_0 + \sum_{i=1}^N \frac{C_i \eta_i}{1 + \eta_i^2}, \quad (1.32)$$

where C_0 is the contribution to m_r from all absorption features *outside* the spectral region of the measurements and $\eta_i \triangleq 2(\omega_i - \omega)/\gamma_i$ and the subscript i labels each absorption feature.

What are the consequences of dispersion in a dielectric medium? Recall that nearly any wave form can be synthesized by Fourier analysis. Consider the following situation: an electromagnetic wave is incident on a shutter that is opened for a short period of time. By adding sinusoids of all frequencies with proper amplitudes (\vec{E}_0) and phases, this pulse of radiation can be synthesized. If this pulse then propagates in a dielectric medium, the various sinusoidal components of the pulse will all travel at different speeds, and the pulse will not maintain its shape — it will become distorted. If most of the sinusoids have frequencies far from ω_0 , the variation of n_r with ω is small and the pulse will propagate with only a small change in shape — it will broaden. In a vacuum, the pulse would retain its shape as it propagates.

1.6 Reflection of Electromagnetic Waves at a Boundary

Now, we consider what happens when an electromagnetic wave is incident on the boundary between two dielectrics. This requires knowing how electric and magnetic fields change

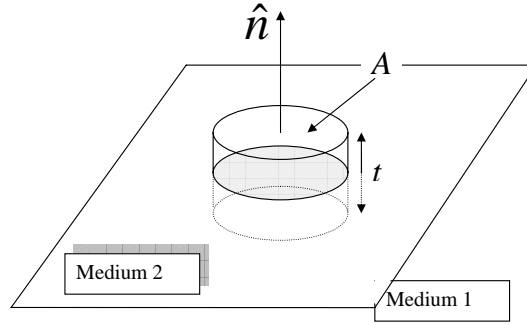
across such a boundary, i.e., the *boundary conditions* on the fields. These boundary conditions can be derived directly from Maxwell's equations. Consider Eq. (I), rewritten to include bound charge density as well as free charge density,

$$\nabla \bullet [\epsilon_0 \vec{E} + \vec{P}] = \rho_{\text{free}},$$

and apply the divergence theorem to deduce

$$\oint_S (\epsilon_0 \vec{E} + \vec{P}) \bullet \hat{n} dA = Q_{\text{free}},$$

where Q_{free} is the total free charge within S . Now consider a cylinder that straddles the



interface between two dielectric media, half above and half below as shown in the figure above. The area of the top is A and the height of the cylinder is t . \hat{n} is the normal to the surface. The shaded area is the portion of the interface inside the cylinder. Now apply Eq. (??) to the cylindrical surface in the following manner: (1) let \vec{E}_i and \vec{P}_i be the electric field and polarization just above ($i = 2$) or below ($i = 1$) the interface between the two media; (2) let the area of the cylinder be so small that these properties do not vary significantly over A ; and let the thickness (t) of the cylinder shrink to zero in such a manner that the interface (the shaded area) is always *inside* the cylinder. If there is free charge within the cylinder after $t \rightarrow 0$, it must be all on the interface, i.e. a surface charge density $\sigma_{\text{free}} = Q_{\text{free}}/A$. Carrying out these operations, we find

$$[(\epsilon_0 \vec{E}_2 + \vec{P}_2) - (\epsilon_0 \vec{E}_1 + \vec{P}_1)] \bullet \hat{n} = \sigma_{\text{free}}.$$

Thus, if there is no free charge on the boundary, the normal component of $(\epsilon_0 \vec{E} + \vec{P})$ is continuous across the boundary. If the dielectric is linear, homogeneous and isotropic (the only kind we will consider), then $(\epsilon_0 \vec{E} + \vec{P}) = \epsilon_0 K_e \vec{E}$, in which case this quantity is continuous:

$$(K_{e2} \vec{E}_2 - K_{e1} \vec{E}_1) \bullet \hat{n} = 0. \quad (1.33)$$

Performing exactly the same steps with Eq. (II) yields

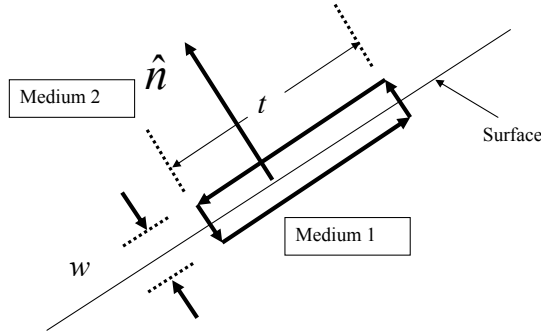
$$(\vec{B}_2 - \vec{B}_1) \cdot \hat{n} = 0, \quad (1.34)$$

i.e., the component of \vec{B} normal to the surface is continuous.

Equations (III) and (IV) can be used to deduce the boundary conditions on the tangential components. From Stokes theorem,

$$\oint_C \vec{E} \cdot d\vec{r} = - \int_S \frac{\partial \vec{B}}{\partial t} \cdot \hat{n} dA,$$

where C is the bounding curve of the area S . Apply this to the curve straddling the interface between the dielectrics as shown below, in which the rectangular loop has length t and width w . The interface is perpendicular to the page (it is viewed on edge) and the plane of the loop is parallel to the page. The width of the loop will shrink to zero in such a manner that the interface is always *inside* the loop. \hat{n} is the normal to the surface. As



before, we integrate \vec{E} around the curve C and take the limit as $w \rightarrow 0$. As $w \rightarrow 0$ the area of the surface S goes to zero so, unless the time derivative of \vec{B} is infinite at the interface, the integral on the right-hand-side is zero. Since $\vec{E} \cdot d\vec{r}$ is proportional to the component of \vec{E} in the direction of $d\vec{r}$, this shows that the tangential component of \vec{E} is continuous across the interface, i.e.,

$$(\vec{E}_2 - \vec{E}_1) \times \hat{n} = 0. \quad (1.35)$$

Similarly from Eq. (IV) we can show that

$$\hat{n} \times (\vec{B}_2 - \vec{B}_1) = \mu_0 \vec{K}_{\text{free}},$$

where \vec{K}_{free} is the free surface current associated with the free surface charge density. When the latter is zero,

$$(\vec{B}_2 - \vec{B}_1) \times \hat{n} = 0, \quad (1.36)$$

and the component of \vec{B} tangent to the surface is continuous. (It should be noted that Eq. (1.36) is modified if magnetic materials are considered.)

We have seen that in an electromagnetic wave $\vec{E} \bullet \vec{\kappa} = 0$ or $\vec{E}_0 \bullet \vec{\kappa} = 0$ so \vec{E} can be in any direction normal to $\vec{\kappa}$. We can always write \vec{E} or \vec{E}_0 (which are in general complex) as the sum of two vectors that are mutually perpendicular and perpendicular to $\vec{\kappa}$, i.e.,

$$\vec{E}(\vec{r}, t) = (\hat{a}E_{0a} + \hat{b}E_{0b}) \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

with $\hat{a} \bullet \hat{b} = 0$, $\hat{a} \bullet \vec{\kappa} = 0$ and $\hat{b} \bullet \vec{\kappa} = 0$. Thus, when the wave interacts with matter, if we understand what happens to each component, e.g., the components along \hat{a} and \hat{b} , we can reconstruct what happens to $\vec{E}(\vec{r}, t)$. The individual waves components above, e.g.,

$$\vec{E}(\vec{r}, t) = \hat{a}E_{0a} \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

and

$$\vec{E}(\vec{r}, t) = \hat{b}E_{0b} \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

have their \vec{E} 's oscillating in a fixed plane: the plane formed by \hat{a} and $\vec{\kappa}$ in the first instance and \hat{b} and $\vec{\kappa}$ in the second. These components are said to be *plane polarized*, with the direction of the \vec{E} being referred to as the *direction* of polarization.

With the boundary conditions on \vec{E} and \vec{B} in hand, we can determine what happens to an electromagnetic wave that is incident on the plane interface between two dielectrics, e.g., when we shine light on the air-water interface. We define the *plane of incidence* to be the plane containing the incident $\vec{\kappa}$ and the surface normal \hat{n} . To understand what happens to the wave, we need to understand what happens, both when \vec{E} is in the plane of incidence or in a plane normal to the plane of incidence. Consider the case when \vec{E} is in the plane of incidence as shown in Figure 1.4. We assume that there will be a transmitted wave and a reflected wave, i.e., when energy is incident from the lower left in the figure we expect some might be reflected from the interface and some might go through; however, we do *not* expect any energy to be propagating toward the interface from the right. Referring to the definitions in Figure 1.4, the three waves are

$$\begin{aligned}\vec{E}_i(\vec{r}, t) &= \vec{E}_{0i} \exp[i(\vec{\kappa}_i \bullet \vec{r} - \omega_i t)], \\ \vec{E}_r(\vec{r}, t) &= \vec{E}_{0r} \exp[i(\vec{\kappa}_r \bullet \vec{r} - \omega_r t)], \\ \vec{E}_t(\vec{r}, t) &= \vec{E}_{0t} \exp[i(\vec{\kappa}_t \bullet \vec{r} - \omega_t t)],\end{aligned}$$

where we have assumed that the wave is propagating with a frequency that is far from any absorption lines in the medium, so the $\vec{\kappa}$'s are all real. The associated magnetic fields are all found from

$$\vec{B}_0 = \frac{\vec{\kappa} \times \vec{E}_0}{\omega}.$$

Now we apply the boundary conditions to the fields.

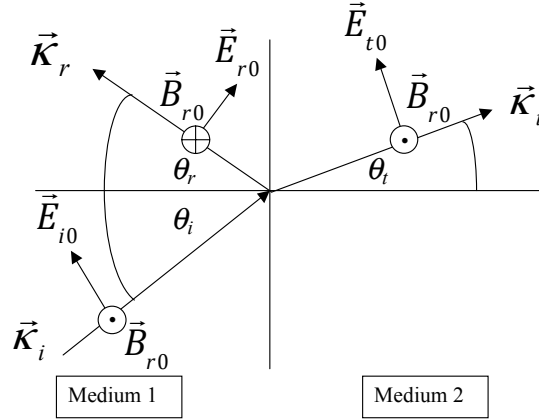


Figure 1.4: Schematic of reflection and refraction of an electromagnetic wave at the interface between two dielectrics. The incident field is from the left propagating toward the interface. The subscripts i , r , and t on all of the quantities refer to the incident, reflected, and transmitted waves, respectively. The amplitudes of the electric vectors are all *in* the plane of the page (the plane of incidence). As shown, the amplitudes of the magnetic fields for the incident and transmitted waves are *out* of the page, while for the reflected wave it is *into* the page. The latter is determined by noting that $\vec{B}_0 = \vec{\kappa} \times \vec{E}_0/\omega$ for each wave.

Take the origin of \vec{r} to the point where all the $\vec{\kappa}$'s meet at the interface in Figure 1.4, and apply the four boundary conditions Eq.s (1.33) through (1.36) to a general point \vec{r} on the interface. From Eq. (1.33), we have

$$\left[K_{e1} \left(\vec{E}_i(\vec{r}, t) + \vec{E}_r(\vec{r}, t) \right) - K_{e2} \vec{E}_t(\vec{r}, t) \right] \cdot \hat{n} = 0.$$

Inserting the explicit forms of the fields,

$$\left[K_{e1} \left(\vec{E}_{0i} \exp[i(\vec{\kappa}_i \cdot \vec{r} - \omega_i t)] + \vec{E}_{0r} \exp[i(\vec{\kappa}_r \cdot \vec{r} - \omega_r t)] \right) - K_{e2} \vec{E}_{0t} \exp[i(\vec{\kappa}_t \cdot \vec{r} - \omega_t t)] \right] \cdot \hat{n} = 0.$$

This equation must hold for all time and every position on the interface (specified by \vec{r}).

This is possible *only* if

$$\omega_t = \omega_r = \omega \triangleq \omega \quad \text{and} \quad \vec{\kappa}_i \cdot \vec{r} = \vec{\kappa}_r \cdot \vec{r} = \vec{\kappa}_t \cdot \vec{r} \quad \text{for any } \vec{r} \text{ on the surface}.$$

Taking \vec{r} to be vertical in Figure 1.4, these give

$$\kappa_i \sin \theta_i = \kappa_r \sin \theta_r = \kappa_t \sin \theta_t.$$

Now, $\kappa_r = \kappa_i = \omega/v_1 = m_1\omega/c$, where m_1 is the refractive index in the incident medium, and $\kappa_t = \omega/v_2 = m_2\omega/c$, so these give

$$\theta_r = \theta_t \quad \text{and} \quad m_1 \sin \theta_i = m_2 \sin \theta_t,$$

the law of reflection and Snell's law of refraction.⁶ With these replacements, all of the exponentials cancel in the application of Eq. (1.33) resulting in

$$K_{e1} [-E_{0i} \sin \theta_i + E_{0r} \sin \theta_r] = K_{e2} E_{0t} \sin \theta_t \quad (1.37)$$

Applying Eq. (1.36), i.e., the component of \vec{B} tangent to the surface is continuous, is easy as all of the \vec{B}_0 's are tangent to the interface. Therefore $\vec{B}_{0i} + \vec{B}_{0r} = \vec{B}_{0t}$, and on using the relationship between \vec{B}_0 and \vec{E}_0 , we get

$$\frac{1}{v_1}(E_{0i} - E_{0r}) = \frac{1}{v_2}E_{0t}$$

or

$$m_1(E_{0i} - E_{0r}) = m_2 E_{0t}. \quad (1.38)$$

Equations (1.37) and (1.38) can be solved for E_{0r} and E_{0t} in terms of E_{0i} , the field of the given incident wave, with the help of the law of reflection and Snell's law. The results are

$$\begin{aligned} E_{0r}^{\parallel} &= \frac{m_1 \cos \theta_t - m_2 \cos \theta_i}{m_1 \cos \theta_t + m_2 \cos \theta_i} E_{0i}^{\parallel}, \\ E_{0t}^{\parallel} &= \frac{2m_1 \cos \theta_i}{m_1 \cos \theta_t + m_2 \cos \theta_i} E_{0i}^{\parallel}, \end{aligned} \quad (1.39)$$

where the superscript \parallel has been added to the fields as an explicit reminder that these apply to the case when the incident electric field is polarized parallel to the plane of incidence. These are called the Fresnel equations. The reader will note that only two of the four boundary conditions were used. The others either give redundant or useless relationships, e.g., $0 = 0$.

Given these fields, it is a simple matter to compute the energy fluxes. For the incident field, the flux *toward* the interface is $\langle \vec{S}_i \rangle_{\text{Ave}} \bullet \hat{n}$, where \hat{n} is directed into medium 2. The energy fluxes of the transmitted and reflected waves *away* from the surface are $\langle \vec{S}_t \rangle_{\text{Ave}} \bullet \hat{n}$ and $\langle \vec{S}_r \rangle_{\text{Ave}} \bullet (-\hat{n})$, respectively. Using (1.25) we see that

$$\begin{aligned} \langle \vec{S}_i \rangle_{\text{Ave}} \bullet \hat{n} &= \frac{m_1}{2\mu_0 c} E_{0i}^2 \cos \theta_i, \\ \langle \vec{S}_t \rangle_{\text{Ave}} \bullet \hat{n} &= \frac{m_2}{2\mu_0 c} E_{0t}^2 \cos \theta_t, \\ \langle \vec{S}_r \rangle_{\text{Ave}} \bullet (-\hat{n}) &= \frac{m_1}{2\mu_0 c} E_{0r}^2 \cos \theta_r. \end{aligned}$$

⁶The laws of reflection and refraction can be placed in vector form. The result for reflection is

$$\hat{k}_r = \hat{k}_i - 2 \cos \theta_i \hat{n}_{1 \rightarrow 2},$$

and for refraction

$$m_2 \hat{k}_t = m_1 \hat{k}_i + (m_2 \cos \theta_t - m_1 \cos \theta_i) \hat{n}_{1 \rightarrow 2},$$

where $\hat{n}_{1 \rightarrow 2}$ is the unit normal to the interface directed from medium 1 toward medium 2.

The Fresnel equations can be used to write these in terms of E_{0i} alone, and defining the Fresnel reflection r_f and transmission t_f coefficients of the interface according to

$$r_f \triangleq -\frac{\langle \vec{S}_r \rangle_{\text{Ave}} \bullet \hat{n}}{\langle \vec{S}_i \rangle_{\text{Ave}} \bullet \hat{n}} \quad \text{and} \quad t_f \triangleq \frac{\langle \vec{S}_t \rangle_{\text{Ave}} \bullet \hat{n}}{\langle \vec{S}_i \rangle_{\text{Ave}} \bullet \hat{n}},$$

respectively, we have that

$$r_f = \left[\frac{E_{0r}}{E_{0i}} \right]^2 \quad \text{and} \quad t_f = \frac{m_2 \cos \theta_t}{m_1 \cos \theta_i} \left[\frac{E_{0t}}{E_{0i}} \right]^2.$$

Inserting the Fresnel equations for \vec{E}_{0i} parallel to the interface in these shows that $r_f^\parallel + t_f^\parallel = 1$, i.e., the energy flux toward the interface equals the energy flux away from the interface.

Examination of the Fresnel equations leads one to ask if it is possible for E_{0r} to vanish, i.e., $m_1 \cos \theta_t = m_2 \cos \theta_i$. Combining this with Snell's law, we find that $E_{0r} = 0$ if $\tan \theta_i = m_2/m_1$. This angle of incidence is called the *Brewster angle* (θ_B). For this angle of incidence $r_f^\parallel = 0$ and $t_f^\parallel = 1$. Figure 1.5 provides the reflection coefficient r_f^\parallel as a function of θ_i for light incident on the air-water interface from the air ($m_1 = 1$ and $m_2 = 1.333$). For this case, $\theta_B = 53.13^\circ$. Figure 1.5 shows $r_f^\parallel \rightarrow 0$ in the vicinity of 50° .

When the incident field is polarized perpendicular to the plane of incidence, a similar analysis yields

$$\begin{aligned} E_{0r}^\perp &= \frac{m_1 \cos \theta_i - m_2 \cos \theta_t}{m_2 \cos \theta_t + m_1 \cos \theta_i} E_{0i}^\perp, \\ E_{0t}^\perp &= \frac{2m_1 \cos \theta_i}{m_2 \cos \theta_t + m_1 \cos \theta_i} E_{0i}^\perp. \end{aligned} \tag{1.40}$$

Is it possible for $E_{0r}^\perp = 0$ for some incident angle? The answer is no, $E_{0r}^\perp = 0$ is incompatible with Snell's law, so r_f^\perp never vanishes. If the refractive indices are eliminated from the Fresnel equations using Snell's law, the reflected and transmitted fields are given by

$$\begin{aligned} E_{0r}^\perp &= -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} E_{0i}^\perp, \\ E_{0t}^\perp &= \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t)} E_{0i}^\perp, \\ E_{0r}^\parallel &= \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} E_{0i}^\parallel, \\ E_{0t}^\parallel &= \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t) \cos(\theta_i - \theta_t)} E_{0i}^\parallel. \end{aligned} \tag{1.41}$$

Then the reflection coefficients are given by

$$r_f^\perp = \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} \quad \text{and} \quad r_f^\parallel = \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)}.$$

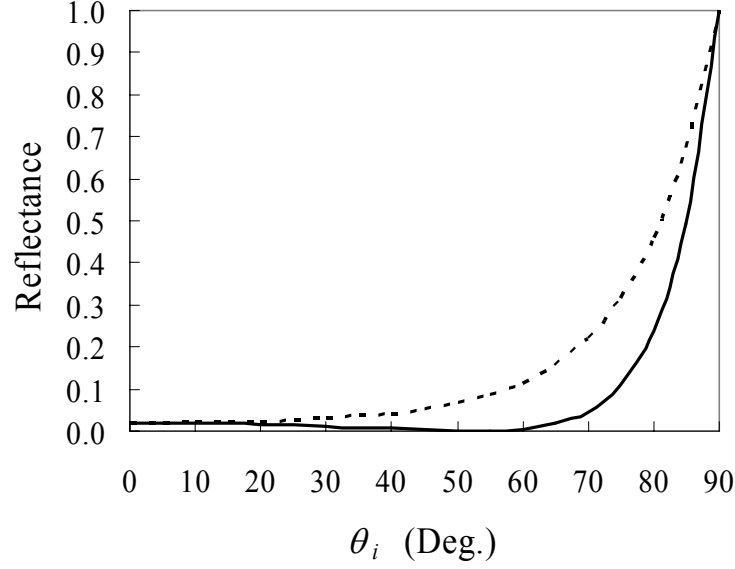


Figure 1.5: The reflectance of the air-water interface as a function of the incidence angle in air. The solid line is for the case with \vec{E} parallel to the plane of incidence and the dotted line is for the case with \vec{E} perpendicular to the plane of incidence

Snell's law tells us that when $m_1 < m_2$, $\theta_t < \theta_i$, and *vice versa* for $m_2 < m_1$. This suggests that there is some incident angle for which $\theta_t = 90^\circ$, i.e., $\sin \theta_i = m_2/m_1$ if $m_2 < m_1$. For this incident angle, called the *critical angle* (θ_c), the transmitted field is propagating parallel to the interface, so $\langle \vec{S}_t \rangle_{\text{Ave}} \bullet \hat{n} = 0$. Thus, at $\theta_i = \theta_c$, $t_f = 0$ and $r_f = 1$. What happens when $\theta_i > \theta_c$? Define a second unit vector $\hat{\ell}$ in a direction parallel to the interface and in the plane of incidence (i.e., upward in the plane of the paper in Figure 1.4). Then

$$\vec{\kappa} = \frac{2\pi}{\lambda_v} (\hat{n} \cos \theta_t + \hat{\ell} \sin \theta_t).$$

Now, for $\theta_i > \theta_c$, $\sin \theta_t$ is real, but $\cos \theta_t$ is pure imaginary:

$$\cos \theta_t = \sqrt{1 - \sin^2 \theta_t} = \sqrt{1 - \frac{m_1^2}{m_2^2} \sin^2 \theta_i} = i \sqrt{\frac{m_1^2}{m_2^2} \sin^2 \theta_i - 1}.$$

Letting

$$\vec{r} = \hat{\ell} \zeta_\ell + \hat{n} \zeta_n + (\hat{\ell} \times \hat{n}) \zeta_\perp,$$

where all of the distances are measured from the intersection of the $\vec{\kappa}$'s in Figure 1.4, we have

$$\vec{\kappa} \bullet \vec{r} = \frac{2\pi}{\lambda_v} (\zeta_n \cos \theta_t + \zeta_\ell \sin \theta_t).$$

Thus, the exponential factor in the transmitted field is

$$\exp \left[-\frac{2\pi}{\lambda_v} \zeta_n \sqrt{\frac{m_1^2}{m_2^2} \sin^2 \theta_i - 1} \right] \exp \left[i \left(\frac{2\pi}{\lambda_v} \zeta_\ell \frac{m_1}{m_2} \sin \theta_i - \omega t \right) \right],$$

and the transmitted field is seen to still propagate in the $\hat{\ell}$ direction (i.e., parallel to the interface) with an amplitude that decays in the \hat{n} direction. How rapid is the decay away from the interface? For the air-water interface, $\theta_c = 48.6^\circ$ for light incident from the water side. If we take $\theta_i = 50^\circ$, i.e., just slightly larger, then the quantity multiplying ζ_ℓ above is $1.3/\lambda_v$, so the field decays to 0.5% of its value at the interface over a distance of $\zeta_\ell = 4\lambda_v$. For larger incidence angles, the decay is even faster.

The transmitted wave provides no average energy flux into the second medium. The easiest way to see this is to examine the reflected wave. The law of reflection, $\theta_r = \theta_i$, is unchanged, so the reflected wave travels as in the case with $\theta_i < \theta_c$. If we substitute the purely-imaginary $\cos \theta_t$ into the Fresnel equations, e.g., Eq. (1.39), it is seen that $E_{0r}^\parallel/E_{0i}^\parallel$ can be written in the form $(-a + ib)/(a + ib) = \exp[i(\pi - 2\phi^\parallel)]$, where $\tan \phi^\parallel = b/a$, or

$$\tan \phi^\parallel = \frac{\sqrt{\sin^2 \theta_i - (m_2/m_1)^2}}{(m_2/m_1) \cos \theta_i}$$

Thus, E_{0r}^\parallel and E_{0i}^\parallel have the same magnitude and differ only in phase, i.e.,

$$\vec{E}_r^\parallel(\vec{r}, t) = \vec{E}_{0r}^\parallel \exp \left[i(\vec{\kappa}_r \bullet \vec{r} - \omega t + \pi - 2\phi^\parallel) \right],$$

and

$$\langle \vec{S}_r^\parallel \rangle_{\text{Ave}} \bullet (-\hat{n}) = \langle \vec{S}_i^\parallel \rangle_{\text{Ave}} \bullet \hat{n},$$

so $r_f^\parallel = 1$, which implies that $t_f^\parallel = 0$. A similar result can be derived for the perpendicular components resulting in $r_f^\perp = 1$, $t_f^\perp = 0$, and

$$\tan \phi^\perp = \frac{\sqrt{\sin^2 \theta_i - (m_2/m_1)^2}}{\cos \theta_i},$$

so $\phi^\perp < \phi^\parallel$ because $m_2 < m_1$. Summarizing, the incident wave is totally reflected from the interface when $\theta_i > \theta_c$: the transmitted wave carries no average energy flux in the \hat{n} direction. In addition, since $\phi^\perp \neq \phi^\parallel$, a linearly polarized wave will become elliptically polarized when it undergoes total internal reflection (see the Appendix, 1.10).

1.7 Production and Scattering of Electromagnetic Waves

In this section we examine how accelerating charges are a source of electromagnetic waves. We then apply the basic results to derive the radiation emitted from an oscillating dipole. Then, as an atom will develop an oscillating dipole moment when placed in an oscillating electric field, we use the result to determine the radiation emitted from such an atom. This emitted radiation is identified as the *scattered* radiation, as it is not traveling in the same direction as the incident electromagnetic wave. The treatment is then extended to the scattering from molecules which, by virtue of their shape, are electrically anisotropic. Finally, we present the scattering of electromagnetic waves by a small dielectric sphere.

1.7.1 Radiation from an Accelerating Charge

The computation of the fields of an arbitrarily moving charge is an extremely difficult problem and far beyond the scope of this review. Therefore, we shall only state the resulting fields and suggest that the reader wishing to understand how they are derived from Maxwell's equations consult advanced texts referenced in the Bibliographic Notes.

Consider a charge of magnitude q located at a point specified by $\vec{r}_0(t)$, where t is time, moving along a curve as shown in Figure 1.6. We want the electric field at the point P specified by the vector \vec{r} . This is given by

$$\vec{E}(\vec{r}, t) = \vec{E}_v(\vec{r}, t) + \vec{E}_a(\vec{r}, t),$$

where

$$\vec{E}_v(\vec{r}, t) = \frac{q}{4\pi\epsilon_0} \frac{[1 - v^2(t_r)/c^2][\hat{R}(t_r) - \vec{v}(t_r)/c]}{R^2(t_r)[1 - \hat{R}(t_r) \bullet \vec{v}(t_r)/c]^3}, \quad (1.42)$$

$$\vec{E}_a(\vec{r}, t) = \frac{q}{4\pi\epsilon_0} \frac{\hat{R}(t_r) \times [(\hat{R}(t_r) - \vec{v}(t_r)/c) \times \vec{a}(t_r)/c]}{cR(t_r)[1 - \hat{R}(t_r) \bullet \vec{v}(t_r)/c]^3},$$

$t_r = t - R(t_r)/c$, and $\hat{R}(t_r) = \vec{R}(t_r)/R(t_r)$. The quantities \vec{v} and \vec{a} are, respectively,

$$\vec{v}(t_r) = \frac{d\vec{R}(t_r)}{dt_r} \quad \text{and} \quad \vec{a}(t_r) = \frac{d^2\vec{R}(t_r)}{dt_r^2},$$

which, since $\vec{R}(t) = \vec{r} - \vec{r}_0(t)$, are related to the actual velocity and acceleration of the charge (relative to the origin of coordinates) by

$$\vec{v}(t_r) = -\frac{d\vec{r}_0(t_r)}{dt_r} \quad \text{and} \quad \vec{a}(t_r) = -\frac{d^2\vec{r}_0(t_r)}{dt_r^2}.$$

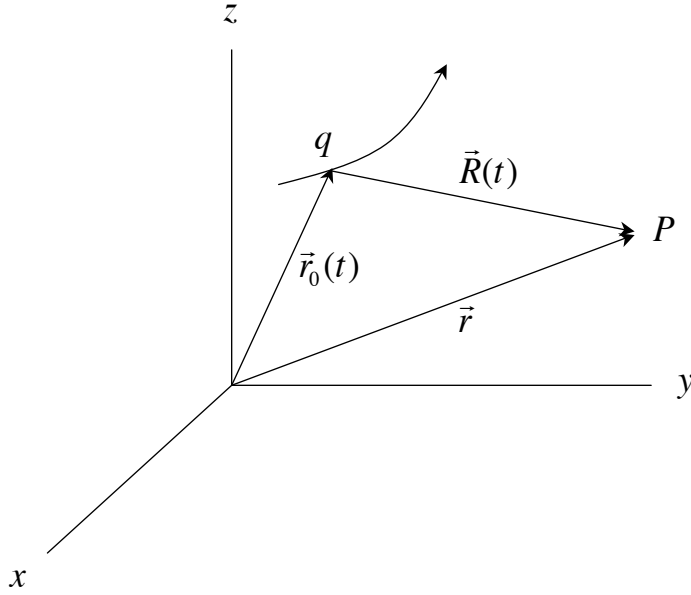


Figure 1.6: A charge of magnitude q moves along the curved path as shown. We want to compute the electric field at the point P .

The time t_r is known as the *retarded* time, i.e., the fields at time t depend on the position, velocity, and acceleration of the charge at an *earlier* time t_r , i.e., *before* the charge arrived at $\vec{r}_0(t)$. The associated magnetic field is given by

$$\vec{B}(\vec{r}, t) = \frac{\hat{R}(t_r) \times \vec{E}(\vec{r}, t)}{c},$$

and the Poynting vector is easily found to be

$$\vec{S} = \frac{\vec{E} \times \vec{B}}{\mu_0} = \frac{1}{\mu_0 c} \vec{E} \times (\hat{R} \times \vec{E}) = \frac{\vec{E} \bullet \vec{E}}{\mu_0 c} \hat{R} + \frac{(\vec{E} \bullet \hat{R}) \vec{E}}{\mu_0 c}. \quad (1.43)$$

To see which of the terms in \vec{E} contributes to radiation we consider a large sphere of radius r around the origin, and the integral

$$\oint_{\text{Sphere}} \vec{S} \bullet \hat{n} dA,$$

which gives the total rate at which energy is *exiting* the sphere. We assume that the charge does not stray too far from the origin, so $\vec{R} \approx \vec{r}$. Noting that $\hat{n} = \hat{r}$, $dA = r^2 \sin \theta d\theta d\phi$, $E_v \sim r^{-2}$, $E_a \sim r^{-1}$, and $\vec{E}_a \bullet \hat{R} = 0$, the only contribution to the integral that remains

as $r \rightarrow \infty$ is that originating from the \vec{E}_a 's in the $\vec{E} \bullet \vec{E}$ term of Eq. (1.43). Thus, only \vec{E}_a produces electromagnetic energy propagating to infinity. This is called *electromagnetic radiation*. Clearly, radiation can be produced by moving charges, but only if they are accelerating.

1.7.2 Radiation from an Oscillating Dipole.

Consider a dipole consisting of a charge $+q$ positioned at $(x, y, z) = (0, 0, z_+)$ and a charge $-q$ positioned at $(0, 0, z_-)$. Let the charges oscillate according to

$$z_+(t) = z_0 \cos(\omega t) \quad \text{and} \quad z_-(t) = -z_0 \cos(\omega t).$$

The dipole moment of this charge distribution $\vec{p} = (qz_+ - qz_-)\hat{e}_z$ so

$$\vec{p}(t) = 2qz_0\hat{e}_z \cos(\omega t) = \vec{p}_0 \cos(\omega t),$$

where $\vec{p}_0 \triangleq 2qz_0\hat{e}_z$ is the *maximum* dipole moment. To compute the radiation from this combination of charges, we consider them individually. We assume that z_0 is sufficiently small that $\vec{R}(t_r) \approx \vec{r}$. For the positive charge,

$$\frac{d^2 z_+}{dt^2} = -\omega^2 z_0 \hat{e}_z \cos(\omega t)$$

so

$$\vec{a}(t_r) = +\omega^2 z_0 \hat{e}_z \cos(\omega t_r) = \omega^2 z_0 \hat{e}_z \cos(\omega t - \omega r/c).$$

Noting that, since \vec{v} and \vec{a} are along the same line (parallel or antiparallel), $\vec{v} \times \vec{a} = 0$, and taking $v \ll c$, we find

$$\vec{E}_a(\vec{r}, t) = \frac{q\omega^2 z_0}{4\pi\epsilon_0 r c^2} \hat{r} \times (\hat{r} \times \hat{k}) \cos(\omega t - \omega r/c).$$

An identical result is obtained for the negative charge because both \vec{a} and q change sign, therefore

$$\vec{E}_a(\vec{r}, t) = \frac{2q\omega^2 z_0}{4\pi\epsilon_0 r c^2} \hat{r} \times (\hat{r} \times \hat{k}) \cos(\omega t - \omega r/c) = \frac{\omega^2}{4\pi\epsilon_0 r c^2} \hat{r} \times (\hat{r} \times \vec{p}_0) \cos(\omega t - \omega r/c).$$

This is an electromagnetic field propagating in the \hat{r} direction with electric field normal to \hat{r} , i.e., $\hat{r} \bullet \vec{E}_a = 0$. It is an example of a spherical wave described in Footnote 2. The wave propagates away from the origin except in directions for which $\vec{k} \times \vec{p}_0 = 0$, i.e., directions for which $\hat{k} = \pm \hat{p}_0$.

The propagation vector is $\vec{\kappa} = \hat{r}\omega/c$, so this can be rewritten

$$\vec{E}_a(\vec{r}, t) = \frac{1}{4\pi\epsilon_0 r} \vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) \cos(\omega t - \kappa r),$$

or in complex form,

$$\vec{E}_a(\vec{r}, t) = \frac{1}{4\pi\epsilon_0 r} \vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) \exp[i(\kappa r - \omega t)].$$

In these equations $\vec{\kappa} \bullet \vec{E}_a = 0$. This is the electric field associated with the radiation propagating in the $\hat{\kappa}$ direction from a dipole of moment $\vec{p}_0 \cos(\omega t)$. The magnetic field is

$$\vec{B}_a = \frac{\hat{r} \times \vec{E}_a}{c} = \frac{\hat{\kappa} \times \vec{E}_a}{c} = \frac{\vec{\kappa} \times \vec{E}_a}{\omega},$$

the usual relationship for an electromagnetic wave. Clearly, $\vec{B}_a \bullet \vec{\kappa} = 0$ as well. Noting that

$$\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) = (\vec{p}_0 \bullet \vec{\kappa}) \vec{\kappa} - \kappa^2 \vec{p}_0,$$

we see that \vec{E}_a is in the plane formed by the direction of propagation $\hat{\kappa}$ and the dipole moment \vec{p}_0 . \vec{B}_a is normal to this plane. In spherical coordinates,

$$\hat{\kappa} = \hat{i} \sin \theta \cos \phi + \hat{j} \sin \theta \sin \phi + \hat{k} \cos \theta$$

so

$$\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) = \left[\hat{i} \sin \theta \cos \theta \cos \phi + \hat{j} \sin \theta \cos \theta \sin \phi + \hat{k} (\cos^2 \theta - 1) \right] \kappa^2 p_0,$$

and the Poynting vector is easily seen to be

$$\vec{S}_a = \frac{1}{\mu_0 c} \frac{\kappa^4 p_0^2 \sin^2 \theta}{(4\pi\epsilon_0)^2 r^2} \hat{\kappa} \cos^2(\omega t - \omega r/c) = \frac{c \kappa^4 p_0^2}{16\pi^2 \epsilon_0} \frac{\sin^2 \theta}{r^2} \hat{\kappa} \cos^2(\omega t - \omega r/c),$$

where we have used $\mu_0 \epsilon_0 = c^{-2}$. As with a plane wave, $\langle \vec{S}_a \rangle_{\text{Ave}} = \vec{S}_a/2$, so the average energy flux through a sphere of radius r is

$$\oint_{\text{Sphere}} \langle \vec{S}_a \rangle_{\text{Ave}} \bullet \hat{n} dA = \frac{1}{4\pi\epsilon_0} \frac{p_0^2 \omega^4}{3c^3}. \quad (1.44)$$

Note that this is proportional to the frequency to the fourth power.

1.7.3 The Total Field of a Dipole

Later, we will need the total field of a dipole, i.e., $\vec{E}_v + \vec{E}_a$. This is straight forward to compute with Eq. (1.42), but the computation is very tedious so we only provide the final result here. Given a dipole located at the origin of coordinates with moment given by $\vec{p}(t) = \vec{p}_0 \exp(-i\omega t)$, the electric field at a point specified by \vec{r} is

$$\vec{E}(r, t) = \frac{1}{4\pi\epsilon_0} \left\{ \frac{\kappa^2 \hat{r} \times (\hat{r} \times \vec{p}_0)}{r} + [\vec{p}_0 - 3\hat{r}(\hat{r} \bullet \vec{p}_0)] \left[\frac{i\kappa}{r^2} - \frac{1}{r^3} \right] \right\} \exp[i(\kappa r - \omega t)]. \quad (1.45)$$

The first term in Eq. (1.45) is from \vec{E}_a and is the radiation field computed above. The second is from \vec{E}_v and is sometimes called the induction field.

1.7.4 Scattering of Electromagnetic Radiation by an Atom (Molecule)

We now examine another of the consequences of illuminating an atom with an electromagnetic wave (propagating in the z direction) given by

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)] = \vec{E}_0 \exp[i(\kappa z - \omega t)].$$

Such an atom will acquire a dipole moment $\vec{p} = \alpha \vec{E}$, and since \vec{E} is oscillating with frequency ω , \vec{p} will as well. Thus, the atom becomes an oscillating dipole (of moment $\alpha \vec{E}_0$) which, as we have seen above, will emit radiation in all directions.⁷ This radiation, which has a frequency identical to that of the incident wave, is the *scattered* field. Define the plane of scattering to be the plane containing the incident propagation vector $\vec{\kappa}_0$ and the scattered propagation vector $\vec{\kappa}$. This is shown in Figure 1.7. In the scattering plane (shaded) ϕ is constant. Traditionally, the scattering angle is denoted by Θ , i.e., $\Theta = \theta$ in the figure. Note that the scattered electric field is parallel (perpendicular) to the scattering plane as the incident field is parallel (perpendicular) to the scattering plane.⁸

The incident field \vec{E}_0 can always be decomposed into components parallel and perpendicular to the scattering plane. Thus, if we understand how these components are scattered, we can understand how a wave with \vec{E}_0 in any orientation is scattered. For the incident radiation we define a set of unit vectors \hat{e}_{0r} and $\hat{e}_{0\ell}$, respectively perpendicular and parallel⁹ to the scattering plane, such that $\vec{\kappa}_0 = \hat{e}_{0r} \times \hat{e}_{0\ell}$. Similarly, for the scattered radiation

⁷In the discussion of dielectrics earlier in this chapter, we found that if q is e (the quantum of electronic charge, i.e., the minimum q), the corresponding z_0 is a just a tiny fraction of the atomic radius.

⁸Note, the scattered field $\vec{E}_s \propto \vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) = \alpha \vec{\kappa} \times (\vec{\kappa} \times \vec{E}_0)$, so \vec{E}_s and \vec{E}_0 are in the same plane.

⁹We employ the common convention of using the last letters of perpendicular and parallel as subscripts to distinguish such unit vectors. Sometimes we will use \hat{r} and $\hat{\ell}$ to represent \hat{e}_r and \hat{e}_ℓ , e.g., Figure 2.17.

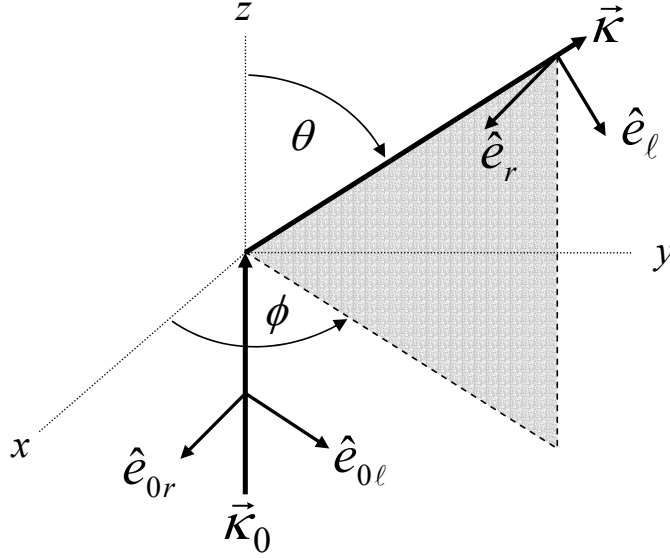


Figure 1.7: An electromagnetic wave with propagation vector \vec{k}_0 is scattered into the direction \vec{k} . The plane formed by \vec{k}_0 and \vec{k} is the *scattering plane*. In this figure the scattering plane is a plane of constant ϕ . Unit vectors \hat{e}_{0r} and $\hat{e}_{0\ell}$ specify directions perpendicular and parallel to the scattering plane for the incident radiation ($\hat{e}_{0r} \times \hat{e}_{0\ell} = \hat{k}_0$). Similarly, \hat{e}_r and \hat{e}_ℓ (with $\hat{e}_r \times \hat{e}_\ell = \hat{k}$) provide the same function for the scattered radiation.

we define unit vectors \hat{e}_r and \hat{e}_ℓ perpendicular and parallel to the scattering plane, with $\vec{k} = \hat{e}_r \times \hat{e}_\ell$. In terms of the cartesian unit vectors,

$$\begin{aligned}\hat{e}_r &= \hat{i} \sin \phi - \hat{j} \cos \phi \\ \hat{e}_\ell &= \hat{i} \cos \theta \cos \phi + \hat{j} \cos \theta \sin \phi - \hat{k} \sin \theta \\ \hat{e}_{0r} &= \hat{e}_r \\ \hat{e}_{0\ell} &= \hat{i} \cos \phi + \hat{j} \sin \phi.\end{aligned}$$

The incident field \vec{E}_0 and the scattered field \vec{E} can be written

$$\begin{aligned}\vec{E}_0 &= \hat{e}_{0r} E_{0r} + \hat{e}_{0\ell} E_{0\ell} \\ \vec{E} &= \hat{e}_r E_r + \hat{e}_\ell E_\ell.\end{aligned}$$

For simplicity (and with no loss in generality), we will take $\phi = 90^\circ$, i.e., the scattering plane is the x - z plane, and note that $\theta = \Theta$, the scattering angle.

We first, consider \vec{E}_0 parallel to the scattering plane, e.g., $\vec{E}_0 = E_{0\ell}\hat{e}_{0\ell} = E_{0\ell}\hat{j}$, so the induced dipole moment is $\vec{p}_0 = p_0\hat{j} = \alpha E_{0\ell}\hat{j}$. Then, noting that $\vec{\kappa} = \kappa(\hat{e}_z \cos \Theta - \hat{e}_y \sin \Theta)$, we have

$$\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) = (-\hat{j} \cos^2 \Theta + \hat{k} \sin \Theta \cos \Theta) \kappa^2 p_0 = -\hat{e}_\ell \kappa^2 p_0 \cos \Theta.$$

The time-averaged Poynting vector for this is

$$\langle \vec{S}_l \rangle_{\text{Ave}} = \frac{c\alpha^2(E_{0\ell})^2\kappa^4}{32\pi^2\epsilon_0} \frac{\cos^2 \Theta}{r^2} \hat{k}, \quad (1.46)$$

For the case where the incident field is normal to the scattering plane, $\vec{E}_0 = E_{0r}\hat{e}_{0r} = E_{0r}\hat{e}_r$, so $\vec{p}_0 = p_0\hat{e}_r = \alpha E_{0r}\hat{e}_r$. This gives

$$\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0) = -\hat{e}_r \kappa^2 p_0,$$

and

$$\langle \vec{S}_r \rangle_{\text{Ave}} = \frac{c\alpha^2(E_{0r})^2\kappa^4}{32\pi^2\epsilon_0} \frac{1}{r^2} \hat{k}. \quad (1.47)$$

Note that $\langle \vec{S}_r \rangle_{\text{Ave}}$ is independent of the scattering angle.¹⁰

Calculating the rate at which energy is scattered (radiated) by the atom out of a large sphere, we need only use Eq. (1.44) from the previous section, replacing p_0 by αE_0 :

$$\oint_{\text{Sphere}} \langle \vec{S} \rangle_{\text{Ave}} \cdot \hat{n} dA = \frac{1}{4\pi\epsilon_0} \frac{\alpha^2 E_0^2 \omega^4}{3c^3}.$$

This shows explicitly that our simple classical model of the atom used to describe α as a function of ω required damping: the atom is continually losing energy through the emission of radiation. However, here the energy is being continually replenished by the incident field. If the incident field were turned off, the atom would continue to radiate, but the dipole moment amplitude would decay exponentially with time, as would the radiated field.

¹⁰Given the incident field $\vec{E}_0 = \hat{e}_{0r}E_{0r} + \hat{e}_{0\ell}E_{0\ell}$, the scattered field $\vec{E}_r = \hat{e}_rE_r + \hat{e}_\ell E_\ell$, is easily seen to be related to the incident fields in matrix form through

$$\begin{pmatrix} E_r \\ E_\ell \end{pmatrix} = \frac{\exp[i(\kappa r - \omega t)]}{-i\kappa r} \begin{pmatrix} A_1(\Theta) & 0 \\ 0 & A_2(\Theta) \end{pmatrix} \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix},$$

where

$$A_1(\Theta) = -\frac{i\kappa^3\alpha}{4\pi\epsilon_0} \quad \text{and} \quad A_2(\Theta) = -\frac{i\kappa^3\alpha}{4\pi\epsilon_0} \cos \Theta.$$

This result will be useful in discussing scattering by particles of arbitrary size.

1.7.5 Influence of Molecular Anisotropy

One might expect that when a diatomic molecule, e.g., N_2 , is placed in an electric field, the dipole moment will depend on the orientation of the molecule with respect to the field, i.e., it would depend on the angle between the inter-nuclear axis and the field, having one value for \vec{E} normal to the axis (\vec{E}^\perp) and another for \vec{E} parallel to the axis (\vec{E}^\parallel). This means there would actually be two polarizabilities:

$$\vec{p} = \alpha^\perp \vec{E}^\perp + \alpha^\parallel \vec{E}^\parallel.$$

If we define a *molecular based* coordinate system with the z axis parallel to the internuclear axis and the x - y plane perpendicular to the internuclear axis, then we can write

$$\vec{p} = \alpha^\perp (E_x \hat{i} + E_y \hat{j}) + \alpha^\parallel E_z \hat{k},$$

or in matrix notation,

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} \alpha^\perp & 0 & 0 \\ 0 & \alpha^\perp & 0 \\ 0 & 0 & \alpha^\parallel \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}.$$

For a more general molecule, the degeneracy seen in the diatomic molecule is removed, however, it is always possible to find a molecular-based coordinate system in which the matrix is diagonal (usually a coordinate system reflecting the symmetry of the molecule). In this coordinate system (called the *principal axis* system), the dipole moment and the electric field are related by (see Section 14.5.3)

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & \alpha_z \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \quad (1.48)$$

Henceforth, we shall use a bold-face type to indicate a 3×3 matrix and a normal vector sign to denote a column vector. In this notation, the above equation reads

$$\vec{p} = \boldsymbol{\alpha} \vec{E}.$$

When such a molecule is subjected to an electromagnetic wave, the induced dipole moment will oscillate with the same frequency as the wave's \vec{E} field, but the direction of the oscillating dipole will not be in the same direction as \vec{E} .

To understand the influence of molecular anisotropy on scattering, we note that the electric field of the wave is provided in the laboratory-fixed system, while the polarizability matrix is diagonal only in the molecular-based system. Therefore, in order to use Eq. (1.48) we need to either transform the polarizability matrix into the laboratory-based system or

\vec{E} into the molecular-based system. Let the laboratory-based coordinate system's cartesian axes be denoted by X , Y , and Z , with unit vectors \hat{e}_X , \hat{e}_Y and \hat{e}_Z , respectively. We shall use upper case indices to denote these unit vectors, e.g., $\hat{e}_I = \hat{e}_X$, for $I = 1$, etc. Likewise let the molecular-based system's cartesian axes be denoted by x , y , and z , with unit vectors \hat{e}_x , \hat{e}_y and \hat{e}_z , respectively. We shall use lower case indices to denote these unit vectors, e.g., $\hat{e}_i = \hat{e}_x$, for $i = 1$, etc. It is shown in the Mathematical Appendix (Section 14.5.7.1) that the \hat{e}_i 's and the \hat{e}_I 's are related by

$$\hat{e}_i = \sum_{I=1}^3 (\hat{e}_i \cdot \hat{e}_I) \hat{e}_I \triangleq \sum_{I=1}^3 a_{iI} \hat{e}_I$$

or in matrix form,

$$\begin{pmatrix} \hat{e}_x \\ \hat{e}_y \\ \hat{e}_z \end{pmatrix} = \begin{pmatrix} a_{xX} & a_{xY} & a_{xZ} \\ a_{yX} & a_{yY} & a_{yZ} \\ a_{zX} & a_{zY} & a_{zZ} \end{pmatrix} = \begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix} \triangleq \mathbf{a} \begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix}.$$

In addition, since $\mathbf{a}^{-1} = \tilde{\mathbf{a}}$,

$$\begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix} = \tilde{\mathbf{a}} \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \\ \hat{e}_z \end{pmatrix}.$$

The 3×3 matrix \mathbf{a} where $a_{iI} = (\hat{e}_i \cdot \hat{e}_I)$, is the matrix of *direction cosines* relating one coordinate system to the other. Using the subscript “Lab” for a vector \vec{V} written in terms of unit vectors in the laboratory-fixed system, i.e., \hat{e}_X , etc., and “Mol” for the *same vector* written in terms of unit vectors in the molecular-fixed system, i.e., \hat{e}_x , etc., we have

$$\vec{V}_{\text{Mol}} = \sum_{i=1}^3 V_i \hat{e}_i \quad \text{and} \quad \vec{V}_{\text{Lab}} = \sum_{I=1}^3 V_I \hat{e}_I.$$

These are related by

$$\vec{V}_{\text{Mol}} = \mathbf{a} \vec{V}_{\text{Lab}} \quad \text{and} \quad \vec{V}_{\text{Lab}} = \tilde{\mathbf{a}} \vec{V}_{\text{Mol}}.$$

In the case of interest here, $\vec{p}_{\text{Mol}} = \boldsymbol{\alpha} \vec{E}_{\text{Mol}}$ and we want \vec{p}_{Lab} in terms of \vec{E}_{Lab} . The required transformations are simple:

$$\vec{E}_{\text{Mol}} = \mathbf{a} \vec{E}_{\text{Lab}} \quad \text{and} \quad \vec{p}_{\text{Lab}} = \tilde{\mathbf{a}} \vec{p}_{\text{Mol}}, \quad \text{so}$$

$$\vec{p}_{\text{Lab}} = \tilde{\mathbf{a}} \boldsymbol{\alpha} \mathbf{a} \vec{E}_{\text{Lab}},$$

where, to reiterate, \mathbf{a} is the matrix of direction cosines above and the subscript “Lab” means the quantities are referenced to the laboratory-based system. Written out in detail,

$$p_I = \sum_{i=1}^3 \sum_{J=1}^3 \alpha_i a_{iI} a_{iJ} E_J, \quad (1.49)$$

Here, and in what follows, we use the notation α_i as an abbreviation for the diagonal elements α_{ii} of the matrix $\boldsymbol{\alpha}$.

What influence does molecular anisotropy have on scattering? Following the derivation in the isotropic case, we need to compute

$$|\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0)|^2,$$

where \vec{p}_0 was $\alpha \vec{E}_0$; however, now \vec{p}_0 depends on the molecular orientation. In the laboratory-based system, we know \vec{E}_0 , and we need \vec{p}_0 in the same system, i.e., $\vec{p}_0 = p_X \hat{e}_X + p_Y \hat{e}_Y + p_Z \hat{e}_Z$. From Figure 1.7 with $\phi = 0$, $\vec{\kappa} = \hat{e}_Z \cos \Theta + \hat{e}_X \sin \Theta$, and

$$\vec{\kappa} \times \vec{p}_0 = \kappa [-\hat{e}_X p_Y \cos \Theta + \hat{e}_Y (p_X \cos \Theta - p_Z \sin \Theta) + \hat{e}_Z p_Y \sin \Theta],$$

so

$$|\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0)|^2 = \kappa^2 |\vec{\kappa} \times \vec{p}_0|^2 = \kappa^4 [p_Y^2 + p_X^2 \cos^2 \Theta + p_Z^2 \sin^2 \Theta - 2p_X p_Z \sin \Theta \cos \Theta].$$

To proceed further we need a definite direction for \vec{E}_0 , so we assume \vec{E} is normal to the scattering plane, i.e., $\vec{E}_0 = -E_{0r} \hat{e}_Y$ (Figure 1.7, $\phi = 0$), and then the only nonzero E_J in Eq. (1.49) is E_Y . Then,

$$\begin{aligned} p_X &= -E_{0r} \sum_{i=1}^3 \alpha_i a_{iX} a_{iY}, & p_X^2 &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iX} a_{iY} a_{jX} a_{jY}; \\ p_Y &= -E_{0r} \sum_{i=1}^3 \alpha_i a_{iY} a_{iY}, & p_Y^2 &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iY} a_{iY} a_{jY} a_{jY}; \\ p_Z &= -E_{0r} \sum_{i=1}^3 \alpha_i a_{iZ} a_{iY}, & p_Z^2 &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iZ} a_{iY} a_{jZ} a_{jY}; \\ & & p_X p_Z &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iX} a_{iY} a_{jZ} a_{jY}. \end{aligned}$$

In a collection of such molecules in a liquid or a gas, the orientation of the individual molecules would be completely random. Therefore any experimental measurement of the Poynting vector would be a measure of a quantity proportional to

$$\langle |\vec{\kappa} \times (\vec{\kappa} \times \vec{p}_0)|^2 \rangle,$$

where here the angle brackets denote averages over all orientations of the molecular-based coordinate system relative to the laboratory-based system. Thus, we need to compute

$\langle p_X^2 \rangle$, $\langle p_Y^2 \rangle$, $\langle p_Z^2 \rangle$, and $\langle p_X p_Z \rangle$. The method of computing these averages is spelled out in the Mathematical Appendix and the results we need are summarized in Table A-1.

Consider $\langle p_Y^2 \rangle$.

$$\begin{aligned} \langle p_Y^2 \rangle &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j \langle a_{iY}^2 a_{jY}^2 \rangle \\ &= (E_{0r})^2 \left[\sum_{i=1}^3 \alpha_i \alpha_i \langle a_{iY}^4 \rangle + 2 \sum_{i=1}^3 \sum_{j>i=1}^3 \alpha_i \alpha_j \langle a_{iY}^2 a_{jY}^2 \rangle \right] \\ &= (E_{0r})^2 \left[\sum_{i=1}^3 \frac{\alpha_i \alpha_i}{5} + 2 \sum_{i=1}^3 \sum_{j>i=1}^3 \frac{\alpha_i \alpha_j}{15} \right]. \end{aligned}$$

Likewise

$$\langle p_X^2 \rangle = \langle p_Z^2 \rangle = (E_{0r})^2 \left[\sum_{i=1}^3 \frac{\alpha_i \alpha_i}{15} - \sum_{i=1}^3 \sum_{j>i=1}^3 \frac{\alpha_i \alpha_j}{15} \right], \quad \text{and} \quad \langle p_X p_Z \rangle = 0.$$

Now, define

$$\begin{aligned} \alpha &\triangleq \frac{1}{3}(\alpha_x + \alpha_y + \alpha_z) \\ \beta^2 &\triangleq \frac{1}{2}[(\alpha_x - \alpha_y)^2 + (\alpha_y - \alpha_z)^2 + (\alpha_z - \alpha_x)^2], \end{aligned}$$

so

$$\begin{aligned} \sum_{i=1}^3 \alpha_i \alpha_i &= \frac{9\alpha^2 + 2\beta^2}{3}, \\ \sum_{i=1}^3 \sum_{j>i=1}^3 \alpha_i \alpha_j &= \frac{9\alpha^2 - \beta^2}{3}. \end{aligned}$$

Collecting all of these results, the time average-orientation average Poynting vector is found to be,

$$\langle \langle \vec{S}_r \rangle_{\text{Ave}} \rangle = \frac{c(E_{0r})^2 \kappa^4}{32\pi^2 \epsilon_0} \left[\frac{45\alpha^2 + 7\beta^2}{45} \right] \frac{1}{r^2} \hat{k}, \quad (1.50)$$

which reduces to Eq. (1.47) in the isotropic case ($\alpha_x = \alpha_y = \alpha_z = \alpha$, $\beta = 0$).¹¹ This is the power per unit area that would be observed by a detector measuring radiant power

¹¹Note that in $\langle \langle \vec{S}_r \rangle_{\text{Ave}} \rangle$ the outside brackets indicate the average over molecular orientation.

without regard for polarization state (i.e., there is no polarization analyzer in front of the detector), when the incident beam is polarized perpendicular to the scattering plane.

Similarly letting \vec{E} for the incident field be in the plane of scattering, i.e., $\vec{E}_0 = E_{0\ell}\hat{e}_X$ we get¹²

$$\langle\langle\vec{S}_\ell\rangle_{\text{Ave}}\rangle = \frac{c(E_{0\ell})^2\kappa^4}{32\pi^2\epsilon_0} \left[\frac{6\beta^2 + (45\alpha^2 + \beta^2)\cos^2\Theta}{45} \right] \frac{1}{r^2}\hat{k}. \quad (1.51)$$

In contrast to the isotropic case, in which the Poynting vector vanishes for $\Theta = 90^\circ$ when the incident field is parallel to the scattering plane, Eq. (1.51) shows that there is no value of Θ for which $\langle\langle\vec{S}_\ell\rangle_{\text{Ave}}\rangle$ vanishes in the anisotropic case. For $\Theta = 90^\circ$, with $E_{0r} = E_{0\ell}$ the ratio

$$\frac{\langle\langle\vec{S}_\ell\rangle_{\text{Ave}}\rangle}{\langle\langle\vec{S}_r\rangle_{\text{Ave}}\rangle} \triangleq \delta = \frac{6\beta^2}{45\alpha^2 + 7\beta^2}.$$

The quantity δ is called the *depolarization* (or depolarization *factor*)¹³ and is zero when $\alpha_x = \alpha_y = \alpha_z$ (isotropic polarizability), and $\frac{1}{2}$ when two of the α_i 's are zero (maximum

¹²For completeness, when $\vec{E}_0 = E_{0\ell}\hat{e}_X$, we have

$$\begin{aligned} p_X &= E_{0\ell} \sum_{i=1}^3 \alpha_i a_{iX} a_{iX}, & p_X^2 &= (E_{0\ell})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iX} a_{iX} a_{jX} a_{jX}; \\ p_Y &= E_{0\ell} \sum_{i=1}^3 \alpha_i a_{iY} a_{iX}, & p_Y^2 &= (E_{0\ell})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iY} a_{iX} a_{jY} a_{jX}; \\ p_Z &= E_{0\ell} \sum_{i=1}^3 \alpha_i a_{iZ} a_{iX}, & p_Z^2 &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iZ} a_{iX} a_{jZ} a_{jX}; \\ & & p_X p_Z &= (E_{0r})^2 \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j a_{iX} a_{iX} a_{jZ} a_{jX}. \end{aligned}$$

$$\begin{aligned} \langle p_X^2 \rangle &= (E_{0\ell})^2 \left[\sum_{i=1}^3 \frac{\alpha_i \alpha_i}{5} + 2 \sum_{i=1}^3 \sum_{j>i=1}^3 \frac{\alpha_i \alpha_j}{15} \right] = \frac{45\alpha^2 + 4\beta^2}{45} \\ \langle p_Y^2 \rangle &= \langle p_Z^2 \rangle = (E_{0\ell})^2 \left[\sum_{i=1}^3 \frac{\alpha_i \alpha_i}{15} - \sum_{i=1}^3 \sum_{j>i=1}^3 \frac{\alpha_i \alpha_j}{15} \right] = \frac{\beta^2}{15} \\ \langle p_X p_Z \rangle &= 0. \end{aligned}$$

¹³There is an alternate definition of the depolarization factor: it is the ratio of light scattered at 90° with its electric field parallel to the scattering plane to that scattered with electric field perpendicular to the scattering plane, when the incident radiation's electric field is perpendicular to the scattering plane. When $\vec{E}_0 = -E_{0r}\hat{e}_Y$ this is proportional to $\langle p_Z^2 \rangle / \langle p_Y^2 \rangle$. Thus, the alternate depolarization factor ρ is given by

$$\rho = \frac{\langle p_Z^2 \rangle}{\langle p_Y^2 \rangle} = \frac{3\beta^2}{45\alpha^2 + 4\beta^2}.$$

anisotropy), i.e.,

$$0 \leq \delta \leq \frac{1}{2}.$$

1.7.6 Scattering of Electromagnetic Radiation by a Small Dielectric Sphere

We now do one last computation, the scattering of an electromagnetic wave from a small dielectric sphere. We shall define “small” to mean that at a given instant of time the electric field does not vary significantly over the volume of the sphere. Therefore, we can assume that the sphere is in a spatially *uniform* electric field and as the wave propagates the field seen by the sphere is simply harmonic in time. This requires that the wavelength of the radiation is much larger than the radius (R) of the sphere. The problem of a dielectric sphere in an otherwise uniform \vec{E} field is straightforward to solve by advanced methods; however, we will provide a simple solution, that is possible *only* because the field inside a uniformly polarized dielectric sphere is spatially uniform.

If we place a dielectric sphere in a uniform field \vec{E}_0 , there will be induced a uniform polarization $\vec{P}_0 = \chi_e \epsilon_0 \vec{E}_0$. This, in turn, produces a field inside the dielectric

$$\vec{E}_1 = -\frac{\vec{P}_0}{3\epsilon_0} = -\frac{\chi_e}{3} \vec{E}_0.$$

But \vec{E}_1 produces additional polarization $\vec{P}_1 = \chi_e \epsilon_0 \vec{E}_1$, yielding

$$\vec{E}_2 = -\frac{\chi_e}{3} \vec{E}_1 = +\frac{\chi_e^2}{9} \vec{E}_0$$

Continuing, and summing the various \vec{E} 's, we have

$$\vec{E} = \sum_{i=0}^{\infty} \vec{E}_i = \left[1 - \frac{\chi_e}{3} + \frac{\chi_e^2}{9} + \cdots \right] \vec{E}_0 = \frac{\vec{E}_0}{1 + (\chi_e/3)}.$$

Thus, the field *inside* the sphere is uniform, and noting that $K_e = 1 + \chi_e$, is

$$\vec{E}_{\text{in}} = \frac{3}{K_e + 2} \vec{E}_0,$$

The relationship between δ and ρ is

$$\delta = \frac{2\rho}{1+\rho} \quad \text{or} \quad \rho = \frac{\delta}{2-\delta}.$$

so the polarization of the sphere is uniform and given by

$$\vec{P}_{\text{in}} = \chi_e \epsilon_0 \vec{E}_{\text{in}} = 3\epsilon_0 \left[\frac{K_e - 1}{K_e + 2} \right] \vec{E}_0$$

But since \vec{P}_{in} is uniform, the sphere will develop an induced dipole moment \vec{p} given by \vec{P} times the volume of the sphere:

$$\vec{p} = 4\pi\epsilon_0 R^3 \left[\frac{K_e - 1}{K_e + 2} \right] \vec{E}_0. \quad (1.52)$$

Since \vec{E}_0 is oscillating with a frequency ω , \vec{p} will as well, and as such will emit radiation. As with the case of an atom, this is the scattered radiation. The results of the computation for scattering by an atom are then directly applicable, and the Poynting vectors of the scattered radiation for the two incident directions are given by, Eq.s (1.46) and (1.47) with $\alpha\vec{E}_0$ there replaced by \vec{p} above.

1.8 The Quantum Nature of the Interaction of Light with Matter

In general, to correctly describe matter, electromagnetic radiation, and their interaction, the quantum nature of both must be recognized. Such a description exists and is called quantum electrodynamics (QED), wherein both matter (charged particles, e.g., electrons) and the associated fields (electric and magnetic) are quantized and governed by the laws of quantum mechanics. In the so-called semi-classical theory of the interaction of light with matter, the electromagnetic field is treated classically (as in this chapter), but matter is described by quantum mechanics (as opposed to classical mechanics). This is the theory that was actually used to relate the oscillator strengths to the refractive index in Eqs. (1.30) and (1.31) in Section 1.5.

In the early development of the quantum theory the classical notion of electromagnetic radiation as a continuous flow of energy was abandoned in favor of a model in which the energy associated with radiation was deposited at its destination in discrete packets called quanta or photons possessing an energy given by $E = h\nu = \hbar\omega$, where ν is the frequency of the radiation, and ω the angular frequency. The constant $h = 6.62 \times 10^{-34}$ J·s is called Planck's constant, and $\hbar = h/2\pi$. In 1905, Albert Einstein explained previous studies of the photoelectric effect — the ejection of electrons from a metal by light incident on its surface, and which we will consider in our discussion of radiation detectors (Chapter 2) — by assuming that a single quantum of energy associated with the light (a photon) simply gave up its energy to the electron in a mechanical-like collision. In fact, this explanation is

not correct, and we now know that the semi-classical theory (as well as QED) can provide an explanation of all of the details of the photoelectric effect without the photon concept. Because a proper understanding of the photoelectric effect requires knowledge of both quantum mechanics and electromagnetic theory, most elementary explanations still rely on Einstein's model of a simple collision between a photon and an electron, subject to energy conservation. We shall follow this tradition in discussing the photoelectric effect as used for radiation detectors (Chapter 2). Indeed, when the simplicity of associating photons with light is desirable, we will associate the energy flux of a light beam, i.e., the time-averaged Poynting vector, with a stream of photons, resulting in a photon flux. If we define this flux to be \mathcal{N} , the number of photons passing a unit area (normal to the beam) per unit time, then

$$\mathcal{N} = \frac{1}{\hbar\omega} |\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|.$$

For the 5 mW He-Ne laser ($\lambda = 632.8$ nm) with a beam cross section of 4 mm^2 , that we discussed earlier, the average Poynting vector was 1250 Watts/m^2 . Each photon in the beam has an energy of about $3.2 \times 10^{-19} \text{ J}$, which results in a photon flux \mathcal{N} of about $4 \times 10^{22} \text{ photons/m}^2 \cdot \text{s}$. Approximately 1.6×10^{16} photons pass the 4 mm^2 cross section of the beam each second. Clearly, the flux \mathcal{N} associated with a typical light beam is enormous. We will find this association useful in discussing the photoelectric effect as well as in our preliminary discussion of Monte Carlo techniques for solving radiative transfer problems; however, for the most part we will avoid the photon concept.

1.9 Concluding Remarks

In this review, we have developed the concepts relating to electromagnetic waves that will be needed throughout the rest of this work. For some parts we have had to simply state important results, e.g., fields of a moving particle, but for most, we have derived the results from first principles — the Maxwell equations and the Lorentz force law. In summary, for the interaction of light with matter, one should understand the basic properties of materials (which in this context means the complex refractive index), the meaning of the Poynting vector, the Fresnel equations for reflectance and transmittance through an interface between two dielectrics, the radiation from an oscillating dipole and scattering by a small dielectric sphere. Finally, the apparatus for dealing with the polarization properties of electromagnetic radiation in the following appendix is crucial for understanding the role of such properties in environmental optics.

1.10 Appendix: Analysis of Polarization of Electromagnetic Radiation

In this appendix we develop in detail the apparatus for dealing with the polarization of electromagnetic radiation. In particular, to describe the polarization properties of radiation, we define the Stokes vector and show how optical processes, such as atomic scattering and Fresnel reflection, relate the Stokes vector of the incident radiation to the Stokes vector of the scattered (or reflected) radiation through objects known as Mueller matrices.

1.10.1 Plane Waves

Consider a plane electromagnetic wave propagating in the $\hat{\kappa}$ direction. Its \vec{E} field will be in a plane perpendicular to $\hat{\kappa}$, and can be resolved into components in two mutually perpendicular directions in this plane

$$\vec{E} = E_r \hat{e}_r + E_\ell \hat{e}_\ell,$$

where $\hat{e}_r \times \hat{e}_\ell = \hat{\kappa}$. One application of the present discussion is the analysis of polarized light reflecting from a dielectric, in which case \hat{e}_ℓ and \hat{e}_r will be, respectively, parallel and perpendicular to the plane of incidence. Likewise, in analysis of scattering they will be parallel and perpendicular to the scattering plane. However, for the moment they are just two orthogonal directions that can be specified by the observer and are normal to $\hat{\kappa}$. The electric field of a plane wave is then

$$\vec{E}(\vec{r}, t) = (E_r \hat{e}_r + E_\ell \hat{e}_\ell) \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)], \quad (1.53)$$

where in general the *complex* components E_r and E_ℓ do not have the same phase, i.e.,

$$E_r = |E_r| \exp(i\phi_r) \quad \text{and} \quad E_\ell = |E_\ell| \exp(i\phi_\ell)$$

with $\phi_r \neq \phi_\ell$.

Before continuing with the formal development, it is instructive to examine some specific cases. For example, assume that $\phi_r = \phi_\ell \triangleq \phi$. Consider an observer located at a fixed position \vec{r}_0 looking toward the source. The field that would be seen is

$$\vec{E}(\vec{r}_0, t) = (|E_r| \hat{e}_r + |E_\ell| \hat{e}_\ell) \exp[i(\vec{\kappa} \bullet \vec{r}_0 - \omega t + \phi)],$$

or, reverting to the real representation of fields,

$$\vec{E}(\vec{r}_0, t) = (|E_r| \hat{e}_r + |E_\ell| \hat{e}_\ell) \cos(\vec{\kappa} \bullet \vec{r}_0 - \omega t + \phi).$$

This field oscillates with amplitude $\sqrt{E_r^2 + E_\ell^2}$ along the direction shown in Figure 1.8, with $\tan \gamma = |E_r|/|E_\ell|$. This is a *linearly polarized plane wave*.

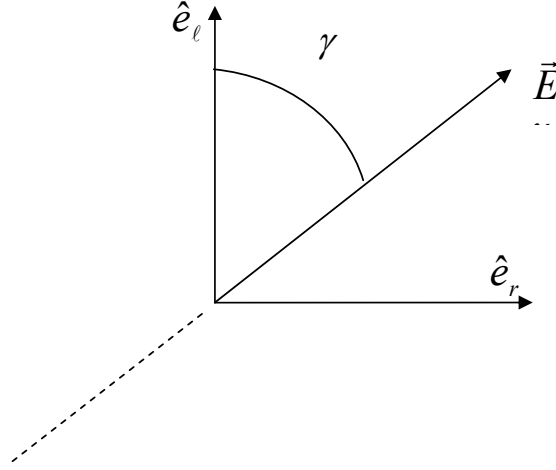


Figure 1.8: The electric field at a given location and a given instant of time. For the linearly polarized wave described in the text, as time increases, the tip of the field vector moves sinusoidally along the dashed line shown that makes an angle γ with the \hat{e}_ℓ axis. Propagation of the wave is *out* of the paper (in the direction of $\hat{e}_r \times \hat{e}_\ell$).

If we relax the requirement that $\phi_r \neq \phi_\ell$, we have

$$\vec{E}(\vec{r}, t) = (|E_r| \hat{e}_r \exp(i\phi_r) + |E_\ell| \hat{e}_\ell \exp(i\phi_\ell)) \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

This is a rather complex wave to visualize — it is *elliptically* polarized. To make the visualization more tractable, we will look at a special case that carries all of the essential subtleties: $\phi_r = 0$, $\phi_\ell = -\pi/2$, $|E_r| = |E_\ell| \triangleq E_0$, and assume the wave propagates in the z direction. Then

$$\vec{E}(z, t) = E_0(\hat{e}_r - i\hat{e}_\ell) \exp[i(\kappa z - \omega t)],$$

and reverting to the real representation of the fields,

$$\vec{E}(z, t) = E_0(\hat{e}_r \cos(\kappa z - \omega t) + \hat{e}_\ell \sin(\kappa z - \omega t)).$$

How does this field vary in time at a given location, e.g., $z = 0$? At $z = 0$, the field reduces to

$$\vec{E}(0, t) = E_0(\hat{e}_r \cos(\omega t) - \hat{e}_\ell \sin(\omega t)).$$

Here, the tip of the electric vector is located on a circle of radius E_0 and traverses the circle *clockwise* (looking toward the source), making a complete revolution in $2\pi/\omega$ seconds. This is shown schematically in Figure 1.9 (left panel). In contrast, how does the field vary in space at a given time, e.g., $t = 0$? At $t = 0$,

$$\vec{E}(z, 0) = E_0(\hat{e}_r \cos(\kappa z) + \hat{e}_\ell \sin(\kappa z)).$$

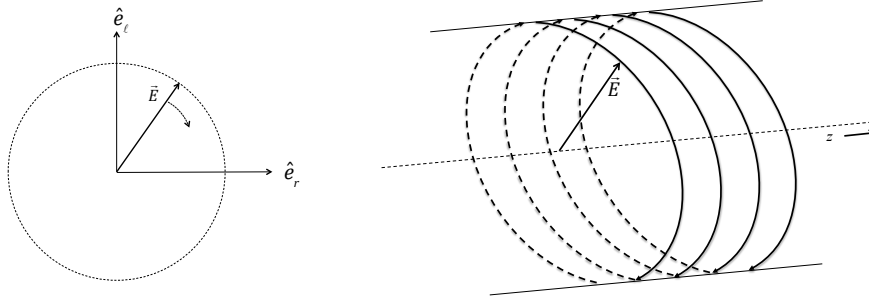


Figure 1.9: Schematic of the electric field in a right-circularly polarized wave at a given location (left panel) and a given instant of time (right panel). At a given location, as time increases, the tip of the field vector moves around the circle in a *clockwise sense* (looking *toward* the source), making one revolution in a time $2\pi/\omega$ (propagation of the wave is *out* of the paper, i.e., in the direction of $\hat{e}_r \times \hat{e}_l$). At a given time, the tip of the field vector at each location (right panel) falls on a curve that is a right-handed helix.

In this case, the tip of the electric vector for a particular z is located on a right-handed cylindrical helix as shown in Figure 1.9 (right panel). This wave is said to be *right circularly polarized* (RCP) — because of the *right*-handed helix. The same result is obtained whenever $\Delta\phi \triangleq \phi_r - \phi_l = +\pi/2$, i.e., \vec{E}_l leads \vec{E}_r at a given location by $\pi/2$. If $\Delta\phi = -\pi/2$, then in the left panel of Figure 1.9 \vec{E} rotates in the *counter clockwise* direction and the helix in the right panel is left handed. This is called *left circularly polarized radiation* (LCP). Finally, in the general case, arbitrary phases and amplitudes, the wave has characteristics similar to those in Figure 1.9, except that the circle in the left panel and the cylinder in the right panel are now both ellipses, and the ellipses are tilted relative to the \hat{e}_r and \hat{e}_l axes (e.g., Figure 1.10).

1.10.2 The Stokes Vector

To completely specify the field, we need to specify two complex quantities E_r and E_l or four real quantities $|E_r|$, $|E_l|$, ϕ_r , and ϕ_l . Nearly all detectors, in which we have interest, measure the time-averaged Poynting vector $\langle \vec{S} \rangle_{\text{Ave}}$. For the general wave in Eq. (1.53) the magnetic field is

$$\vec{B}(\vec{r}, t) = \frac{E_r \hat{e}_l - E_l \hat{e}_r}{c} \exp[i(\vec{k} \bullet \vec{r} - \omega t)],$$

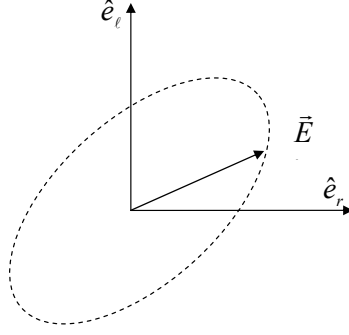


Figure 1.10: The electric field at a given location and a given instant of time for an elliptically polarized wave. As time increases, the tip of the field vector moves clockwise (for right-elliptical polarization) or counter clockwise (for left-elliptical polarization) around the dashed ellipse, making one revolution in a time $2\pi/\omega$. Propagation of the wave is *out* of the paper (in the direction of $\hat{e}_r \times \hat{e}_l$).

so the time-averaged Poynting vector is

$$\langle \vec{S} \rangle_{\text{Ave}} = \frac{\hat{\kappa}}{2\mu_0 c} (E_l E_l^* + E_r E_r^*).$$

Measurement of this only gives the combination $E_l E_l^* + E_r E_r^*$. Are there experiments that can be performed to provide the full two complex or four real quantities that are required to specify the polarization? The answer is yes, given optical devices such as *linear polarizers* that only pass electric fields oscillating in a given direction (the *pass* direction), and *wave plates* that introduce a phase difference between fields oscillating along \hat{e}_r and \hat{e}_l . Such devices are readily available.

Consider a linear polarizer that will totally transmit all electromagnetic waves with fields parallel to \hat{e}_p and completely absorb all fields perpendicular to \hat{e}_p , i.e., complete absorption in the direction of $\hat{\kappa} \times \hat{e}_p$. The direction \hat{e}_p is the *pass direction* of the polarizer. Placing the polarizer with $\hat{e}_p = \hat{e}_r$ in the electromagnetic wave passes the field

$$\hat{e}_p \left[\hat{e}_p \bullet \vec{E}(\vec{r}, t) \right] = E_r \hat{e}_r \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

and the associated Poynting vector is

$$\langle \vec{S}_r \rangle_{\text{Ave}} = \frac{\hat{\kappa}}{2\mu_0 c} E_r E_r^*.$$

Likewise, with $\hat{e}_p = \hat{e}_l$

$$\langle \vec{S}_l \rangle_{\text{Ave}} = \frac{\hat{\kappa}}{2\mu_0 c} E_l E_l^*.$$

Thus, measurement of $\langle \vec{S}_r \rangle_{\text{Ave}}$ and $\langle \vec{S}_\ell \rangle_{\text{Ave}}$ provides $E_r E_r^* = |E_r|^2$ and $E_\ell E_\ell^* = |E_\ell|^2$. But, what about their phases?

Let $\hat{e}_p = (\hat{e}_r + \hat{e}_\ell)/\sqrt{2}$, i.e., let the polarizer be oriented so that its pass direction is at an angle of 45° with both \hat{e}_r and \hat{e}_ℓ . Then the transmitted field will be

$$\vec{E}_p(\vec{r}, t) = \hat{e}_p \frac{E_r + E_\ell}{\sqrt{2}} \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

and

$$2\langle \vec{S}_p \rangle_{\text{Ave}} = \frac{\hat{\kappa}}{2\mu_0 c} [E_\ell E_\ell^* + E_r E_r^* + E_r E_\ell^* + E_r^* E_\ell].$$

Thus, the measurement of $\langle \vec{S}_p \rangle_{\text{Ave}}$ along with $\langle \vec{S}_r \rangle_{\text{Ave}}$ and $\langle \vec{S}_\ell \rangle_{\text{Ave}}$ allows the determination of $E_r E_\ell^* + E_r^* E_\ell$.

Finally, a device called a *quarter-wave plate*, when properly oriented has the ability to shift the phase of the field along \hat{e}_r by $\pm\pi/2$ with respect to the field along \hat{e}_ℓ . Introducing such a device in the original wave, shifting phase of the field E_r by $-\pi/2$ relative to E_ℓ , results in the transmitted field

$$\begin{aligned} \vec{E}(\vec{r}, t) &= (E_\ell \hat{e}_\ell + E_r \hat{e}_r \exp(-i\pi/2)) \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)] \\ &= (E_\ell \hat{e}_\ell - i E_r \hat{e}_r) \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)]. \end{aligned}$$

Now, follow this by a linear polarizer oriented as in the last paragraph and the transmitted field through both is

$$\vec{E}(\vec{r}, t) = \hat{e}_p \frac{(E_\ell - i E_r)}{\sqrt{2}} \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)],$$

with the associated Poynting vector

$$2\langle \vec{S}_q \rangle_{\text{Ave}} = \frac{\hat{\kappa}}{2\mu_0 c} [E_\ell E_\ell^* + E_r E_r^* + i(E_r^* E_\ell - E_r E_\ell^*)],$$

where the subscript “*q*” is used to indicate “quarter-wave plate.” Thus, measurement of $\langle \vec{S}_q \rangle_{\text{Ave}}$ along with $\langle \vec{S}_p \rangle_{\text{Ave}}$, $\langle \vec{S}_r \rangle_{\text{Ave}}$, and $\langle \vec{S}_\ell \rangle_{\text{Ave}}$ allows the determination of $E_r^* E_\ell - E_r E_\ell^*$. These are sufficient to determine the polarization state of the wave, and allow determination of what is known as the *Stokes parameters* I , Q , U , and V , given by

$$\begin{aligned} I &= \frac{1}{2\mu_0 c} (E_\ell E_\ell^* + E_r E_r^*) = [\langle S_\ell \rangle + \langle S_r \rangle] = \langle S \rangle, \\ Q &= \frac{1}{2\mu_0 c} (E_\ell E_\ell^* - E_r E_r^*) = [\langle S_\ell \rangle - \langle S_r \rangle], \\ U &= \frac{1}{2\mu_0 c} (E_\ell E_r^* + E_\ell^* E_r) = [2\langle S_p \rangle - \langle S_\ell \rangle - \langle S_r \rangle], \\ V &= \frac{1}{2\mu_0 c} i(E_\ell E_r^* - E_\ell^* E_r) = [2\langle S_q \rangle - \langle S_\ell \rangle - \langle S_r \rangle], \end{aligned} \tag{1.54}$$

where the notation $\langle S \rangle$ has been used for the magnitude of $\langle \vec{S} \rangle_{\text{Ave}}$. Writing

$$E_r = |E_r| \exp(i\phi_r) \quad \text{and} \quad E_\ell = |E_\ell| \exp(i\phi_\ell)$$

it is easy to show that

$$\begin{aligned} I &= \frac{1}{2\mu_0 c} [|E_\ell|^2 + |E_r|^2], \\ Q &= \frac{1}{2\mu_0 c} [|E_\ell|^2 - |E_r|^2], \\ U &= \frac{1}{2\mu_0 c} 2|E_\ell||E_r| \cos(\phi_r - \phi_\ell), \\ V &= \frac{1}{2\mu_0 c} 2|E_\ell||E_r| \sin(\phi_r - \phi_\ell), \end{aligned} \tag{1.55}$$

or, recalling $\tan \gamma = |E_r|/|E_\ell|$ (Figure 1.8), $\sin \gamma = |E_r|/\sqrt{|E_r|^2 + |E_\ell|^2}$, etc.,

$$\begin{aligned} Q &= I \cos 2\gamma \\ U &= I \sin 2\gamma \cos(\phi_r - \phi_\ell) \\ V &= I \sin 2\gamma \sin(\phi_r - \phi_\ell). \end{aligned} \tag{1.56}$$

Thus, measurement of $\langle \vec{S}_q \rangle_{\text{Ave}}$ along with $\langle \vec{S}_p \rangle_{\text{Ave}}$, $\langle \vec{S}_r \rangle_{\text{Ave}}$, and $\langle \vec{S}_\ell \rangle_{\text{Ave}}$ allows the determination of the amplitude of, and the phase *difference* between, the two components. Note that because the four parameters are written in terms of only three quantities, there must be a relationship between them. It is

$$I^2 = Q^2 + U^2 + V^2. \tag{1.57}$$

It should be noted that the absolute phases of the wave components are impossible to determine, which should not be surprising as they are arbitrary — the absolute phase depends on how we define $t = 0$.

Usually the Stokes parameters are arranged in a column vector, the *Stokes vector*:

$$\mathbf{I} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix}.$$

Stokes vectors are easy to evaluate by virtue of their definition, e.g., in the case of circular polarization,

$$\mathbf{I} = \begin{pmatrix} I \\ 0 \\ 0 \\ \pm I \end{pmatrix},$$

where $+$ and $-$ are for right and left circularly polarized light, respectively.

1.10.3 Change in the Reference System

It is of interest to understand how the Stokes vector is modified if the reference system is changed, i.e., how does the vector \mathbf{I}' defined with respect to

$$\begin{aligned}\hat{e}'_r &= \hat{e}_r \cos \psi + \hat{e}_\ell \sin \psi \\ \hat{e}'_\ell &= -\hat{e}_r \sin \psi + \hat{e}_\ell \cos \psi,\end{aligned}$$

relate to \mathbf{I} defined with respect to \hat{e}_ℓ and \hat{e}_r ? Note that the primed system \hat{e}'_ℓ is simply

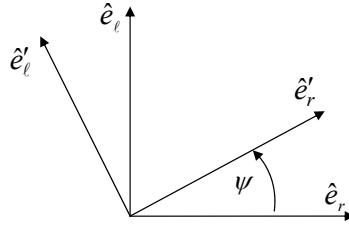


Figure 1.11: Rotation of the primed axes with respect to the unprimed axes through an angle ψ , *counterclockwise* looking *at* the source ($\hat{\kappa} = \hat{e}_r \times \hat{e}_\ell = \hat{e}'_r \times \hat{e}'_\ell$).

rotated from \hat{e}_r toward \hat{e}_ℓ about an axis parallel to the propagation direction, through an angle ψ (Figure 1.11), where, the angle ψ is positive if the rotation is counterclockwise looking at the source. The inverse transformation is

$$\begin{aligned}\hat{e}_r &= \hat{e}'_r \cos \psi - \hat{e}'_\ell \sin \psi \\ \hat{e}_\ell &= \hat{e}'_r \sin \psi + \hat{e}'_\ell \cos \psi,\end{aligned}$$

so

$$\begin{aligned}\vec{E}(\vec{r}, t) &= (E_r \hat{e}_r + E_\ell \hat{e}_\ell) \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t)] \\ &= [(E_r \cos \psi + E_\ell \sin \psi) \hat{e}'_r + (E_\ell \cos \psi - E_r \sin \psi) \hat{e}'_\ell] \exp[i(\vec{\kappa} \bullet \vec{r} - \omega t),\end{aligned}$$

and the components of the field amplitudes in the primed system are

$$\begin{aligned}E'_r &= E_r \cos \psi + E_\ell \sin \psi \\ E'_\ell &= -E_r \sin \psi + E_\ell \cos \psi.\end{aligned}$$

Inserting these into Eq. (1.54), with the primed fields replacing the unprimed fields, yields

$$I' = I,$$

$$\begin{aligned}
Q' &= Q \cos 2\psi - U \sin 2\psi, \\
U' &= Q \sin 2\psi + U \cos 2\psi, \\
V' &= V,
\end{aligned}$$

or in matrix notation, $\mathbf{I}' = \mathbf{R}(\psi)\mathbf{I}$, where

$$\mathbf{R}(\psi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\psi & -\sin 2\psi & 0 \\ 0 & \sin 2\psi & \cos 2\psi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (1.58)$$

It will be important to remember that here the angle ψ is *positive* when the new axes are rotated *counterclockwise* relative to the old axes, looking *toward the source*. This matrix will be important in the study of radiative transfer when polarization is considered.

1.10.4 The Mueller Matrix

When a wave passes through a linear polarizer with pass axis in a general direction \hat{e}_p , the transmitted field amplitude is

$$\begin{aligned}
\hat{e}_p [\hat{e}_p \bullet (E_r \hat{e}_r + E_\ell \hat{e}_\ell)] &= \hat{e}_p [E_r \sin \theta + E_\ell \cos \theta] \\
&= \hat{e}_\ell (E_r \sin \theta \cos \theta + E_\ell \cos^2 \theta) + \hat{e}_r (E_r \sin^2 \theta + E_\ell \cos \theta \sin \theta),
\end{aligned}$$

where θ is the angle between \hat{e}_p and \hat{e}_ℓ measured toward \hat{e}_r . Inserting these field components into Eq. (1.54) yields for the transmitted Stokes vector $\mathbf{I}_t = \mathbf{M}\mathbf{I}$, or

$$\begin{pmatrix} I_t \\ Q_t \\ U_t \\ V_t \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \cos 2\theta \sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta \sin 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}. \quad (1.59)$$

The 4×4 matrix \mathbf{M} is called the Mueller matrix.

As an example of the use of Eq. (1.59), consider a linearly polarized light. From Eq. (1.56) the Stokes vector is

$$\mathbf{I} = \begin{pmatrix} I \\ I \cos 2\gamma \\ I \sin 2\gamma \\ 0 \end{pmatrix},$$

where γ is the angle \vec{E} makes with the \hat{e}_ℓ direction (Figure 1.8). When such a beam passes through a linear polarizer, the resulting Stokes vector is

$$\mathbf{I} = \frac{I}{2} \begin{pmatrix} 1 + \cos 2\gamma \cos 2\theta + \sin 2\gamma \sin 2\theta \\ \cos 2\theta + \cos^2 2\theta \cos 2\gamma + \cos 2\theta \sin 2\theta \sin 2\gamma \\ \sin 2\theta + \cos 2\theta \sin 2\theta \cos 2\gamma + \sin^2 2\theta \sin 2\gamma \\ 0 \end{pmatrix} = \frac{I}{2} \begin{pmatrix} 1 + \cos 2(\gamma - \theta) \\ \cos 2\theta (1 + \cos 2(\gamma - \theta)) \\ \sin 2\theta (1 + \cos 2(\gamma - \theta)) \\ 0 \end{pmatrix}.$$

Clearly, if $\theta = \gamma$, \mathbf{I} is unchanged when passing through the polarizer, but if $\theta = \gamma \pm \pi/2$, then the beam is totally extinguished. Notably, the first component $I(1 + \cos 2(\gamma - \theta))/2$, would vary from I to 0 as $\gamma - \theta$ is varied for 0 to 90° . The *direction* of polarization of the original beam is defined to be the direction for which the transmission of the polarizer is maximum, i.e., $\theta = \gamma$.

The Mueller matrix can be written for any combination of optical elements. Clearly, if a wave passes through two elements with Mueller matrices \mathbf{M}_1 and \mathbf{M}_2 (labeled in the order of passage), the transmitted Stokes vector will be $\mathbf{I}_t = \mathbf{M}_2 \mathbf{M}_1 \mathbf{I}$. The fact that the order of polarizing operation is important (e.g., a wave passing through a polarizer with $\hat{e}_p = (\hat{e}_r + \hat{e}_\ell)/\sqrt{2}$ and then a second with $\hat{e}_p = \hat{e}_\ell$ will emerge with field polarized along \hat{e}_ℓ , while in reverse order the emerging field will be along $(\hat{e}_r + \hat{e}_\ell)/\sqrt{2}$) is manifest in the fact that in general matrices do not commute: $\mathbf{M}_2 \mathbf{M}_1 \neq \mathbf{M}_1 \mathbf{M}_2$.

1.10.5 Unpolarized (Natural) Light

We know from experience that when light from a natural source, e.g., an incandescent lamp or a candle, is passed through a perfect linear polarizer, the transmitted light's brightness is decreased by half, independent of the direction of \hat{e}_p . (Brightness is proportional to the Poynting vector and will be defined with precision in the next chapter.) Furthermore, we know that if a second polarizer with pass direction \hat{e}'_p , that is perpendicular to \hat{e}_p , is introduced after the first polarizer, the light will be completely extinguished. The first observation suggests that the light entering the first polarizer is completely *unpolarized*, i.e., \vec{E} does not appear to be oscillating in any particular direction. However, the second observation suggests that the wave emerging from the first polarizer is linearly polarized with \vec{E} along \hat{e}_p . In fact, if we take $\hat{e}_p = \hat{e}_\ell$, the incident and transmitted Stokes vectors would be found to be

$$\begin{pmatrix} I \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \frac{1}{2} \begin{pmatrix} I \\ I \\ 0 \\ 0 \end{pmatrix},$$

respectively, when the measured Poynting vectors $\langle S_\ell \rangle_{\text{Ave}}$, etc., are inserted into Eq. (1.54). Clearly, the *incident* radiation does not have the characteristics of the electromagnetic

waves that we have been dealing with so far, e.g., $I^2 \neq Q^2 + U^2 + V^2 = 0$; however, it is turned into a *completely polarized beam* by passing it through a polarizer. How can this be?

Thus far we have been considering only monochromatic waves (single ω) which exist over all space for all time. However, when an atom emits light, the emitted radiation has a small spread in frequency $\Delta\omega \ll \omega$ and the associated wave at a given point \vec{r}_0 is something like that drawn in Figure 1.12, where $\Delta T \sim 1/\Delta\omega$. In reality, there are many more oscillations in the region of more-or-less constant amplitude than shown in the figure. For an atom undergoing spontaneous emission, ΔT in the visible is $\sim 3 \times 10^{-6}$ s, while the period of visible light is $\sim 2 \times 10^{-15}$ s, so there are $\sim 1.5 \times 10^9$ oscillations.

A wave such as shown in Figure 1.12 can be written

$$\vec{E}(\vec{r}, t) = [E_r(t)\hat{e}_r + E_\ell(t)\hat{e}_\ell] \exp[i(\vec{k} \bullet \vec{r} - \omega t)],$$

where $E_r(t)$ and $E_\ell(t)$ are slowly varying functions of time, compared to the rapid variations of the exponential factor, and vanish outside the time interval ΔT . The resulting field is often referred to as a *wave packet*. For such a wave, the average Poynting vector will be significant only within ΔT , and in that interval will be very nearly that of a plane wave with fixed amplitudes. There will be a definite polarization state within ΔT , i.e., the \vec{E} vector will oscillate like that of a plane wave. Another atom in the source will emit a similar wave, but the polarization will be different and the radiation will be emitted at a different time so the phase will be different. Considering a collection of such atoms emitting wave packets, each with arbitrary phase and polarization (and arrival time at \vec{r}_0), the field amplitude and its direction at \vec{r}_0 will fluctuate. To find the Poynting vector in this situation it is useful to examine two extreme cases. First, we assume that the time interval ΔT is so small that none of the wave packets overlap. Then,

$$\begin{aligned} \langle \vec{S}(\vec{r}_0, t_0) \rangle_{\text{Ave}} &= \frac{1}{T_0} \int_{t_0 - T_0/2}^{t_0 + T_0/2} \vec{S}(\vec{r}_0, t) dt \\ &= \frac{1}{T_0} \int_{t_0 - T_0/2}^{t_0 + T_0/2} \sum_{i=1}^N \vec{S}_i(\vec{r}_0, t) dt \\ &= \frac{1}{T_0} \sum_{i=1}^N \int_{t_0 - T_0/2}^{t_0 + T_0/2} \vec{S}_i(\vec{r}_0, t) dt \\ &= \frac{\Delta T}{T_0} \sum_{i=1}^N \langle \vec{S}_i(\vec{r}_0, t_{ic}) \rangle_{\text{Ave}} \end{aligned}$$

where

$$\langle \vec{S}_i(\vec{r}_0, t_{ic}) \rangle_{\text{Ave}} = \frac{1}{\Delta T} \int_{t_{ic} - \Delta T/2}^{t_{ic} + \Delta T/2} \vec{S}_i(\vec{r}_0, t) dt,$$

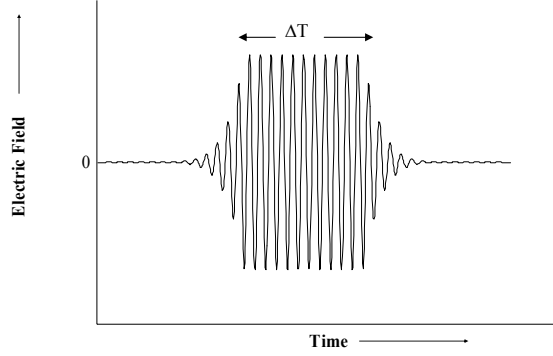


Figure 1.12: Schematic of the electric field of a light wave emitted by an atom. The amplitude of the field is essentially constant over a time interval ΔT . In reality, there are many more oscillations of the electric field during ΔT than shown here.

and t_{ic} is the time that the center of the i^{th} packet reaches \vec{r}_0 . Thus, the time-averaged Poynting vector is *proportional to* the sum of the time-averaged Poynting vectors of the individual packets. Since each term in the sum represents a different state of polarization, if we perform the experiments required to compute the Stokes vector, we will find

$$\langle \vec{S}_\ell \rangle_{\text{Ave}} = \langle \vec{S}_r \rangle_{\text{Ave}} = \langle \vec{S}_p \rangle_{\text{Ave}} = \langle \vec{S}_q \rangle_{\text{Ave}} = \frac{1}{2} \langle \vec{S} \rangle_{\text{Ave}}, \quad (1.60)$$

so $Q = U = V = 0$ and $I(t_0) = \langle \vec{S}(t_0) \rangle_{\text{Ave}}$.

In the other extreme, assume that ΔT is so large that all of the packets overlap, e.g, each packet is very nearly a monochromatic plane wave. In this case,

$$\vec{S}(\vec{r}_0, t) = \frac{1}{2\mu_0 c} \vec{E}_0 \bullet \vec{E}_0^*, \quad \text{with} \quad \vec{E}_0 = \sum_{i=1}^N \vec{E}_{0i},$$

where \vec{E}_{0i} is the complex amplitude of the i^{th} packet. To evaluate this, let $\vec{E}_{0i} = \hat{e}_i |E_{0i}| \exp(i\phi_i)$. Then

$$\begin{aligned} \vec{E}_0 \bullet \vec{E}_0^* &= \sum_{i=1}^N \sum_{j=1}^N \hat{e}_i \bullet \hat{e}_j |E_{0i}| |E_{0j}| \exp[i(\phi_i - \phi_j)] \\ &= \sum_{i=1}^N |E_{0i}|^2 + \sum_{i=1}^N \sum_{j>i}^N 2\hat{e}_i \bullet \hat{e}_j |E_{0i}| |E_{0j}| \cos(\phi_i - \phi_j). \end{aligned}$$

Since the waves are all emitted by different atoms, their phases are random, so the second sum vanishes, and

$$\vec{S}(\vec{r}_0, t) = \frac{1}{2\mu_0 c} \sum_{i=1}^N |E_{0i}|^2.$$

(Note that this is true even if $\hat{e}_i = \hat{e}_j$ and $|E_{0i}| = |E_{0j}|$.) Time averaging this result yields

$$\langle \vec{S}(\vec{r}_0, t_0) \rangle_{\text{Ave}} = \frac{\Delta T}{T_0} \sum_{i=1}^N \langle \vec{S}_i(\vec{r}_0, t_{ic}) \rangle_{\text{Ave}},$$

where we are implicitly assuming that $\Delta T < T_0$, and as before, $Q = U = V = 0$, and $I = \langle \vec{S}(\vec{r}_0, t_0) \rangle_{\text{Ave}}$.

The light (radiation) we have described here, for which the average Poynting vector of a superposition of waves is the sum of the average Poynting vector of the individual waves, is usually called *incoherent* light. Incoherent light from a source such as a candle or incandescent lamp is often called *natural* light and is unpolarized; however, it is important to note that incoherent light can be fully polarized. For example, if unpolarized incoherent light is passed through a linear polarizer, the \vec{E} vector of the individual emerging wave packets will all be along \hat{e}_p , so the stream of packets will be completely polarized, but it is still incoherent. The polarization state of incoherent light is analyzed in the same manner as that of a plane wave.

1.10.6 Partially Polarized Light

One can imagine situations in a source that would cause the wave packets emitted by individual atoms to have a bias for alignment in a certain direction, e.g., the source placed in a magnetic field. Then Eq. (1.60) would no longer be valid and the stream of packets would have some polarization. Formally, it can be shown (e.g., Born and Wolf, *Principles of Optics*) that radiation with these properties can be represented as a linear superposition of two independent streams of unpolarized and completely polarized radiation, and that this representation is unique. Thus, for any state of radiation,

$$\mathbf{I} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} (1-P)I \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} PI \\ Q \\ U \\ V \end{pmatrix} \quad (1.61)$$

For the completely polarized part Eq. (1.57) tells us that $(PI)^2 = Q^2 + U^2 + V^2$, so

$$P = \sqrt{\frac{Q^2 + U^2 + V^2}{I^2}}.$$

The quantity P is called the *degree of polarization*. For completely unpolarized light $P = 0$ and for completely polarized light $P = 1$. For a partially (linearly) polarized beam, the second term can be written

$$\begin{pmatrix} PI \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} PI \\ PI \cos 2\gamma \\ PI \sin 2\gamma \\ 0 \end{pmatrix}$$

so for such a beam

$$\mathbf{I} = \begin{pmatrix} I \\ PI \cos 2\gamma \\ PI \sin 2\gamma \\ 0 \end{pmatrix}.$$

If this beam is passed through a linear polarizer, i.e., that in Eq. (1.59), the resulting Stokes vector is

$$\mathbf{I} = \frac{I}{2}(1 + P \cos 2(\gamma - \theta)) \begin{pmatrix} 1 \\ \cos 2\gamma \\ \sin 2\gamma \\ 0 \end{pmatrix}.$$

As earlier, the first (top) element of \mathbf{I} is maximum (I_{Max}) when $\theta = \gamma$ and minimum (I_{Min}) when $\theta = \gamma + \pi/2$. Clearly

$$P = \frac{I_{\text{Max}} - I_{\text{Min}}}{2},$$

which provides a simple method to measure P for cases where $V = 0$ (linearly polarized light).¹⁴

1.10.7 Mueller Matrix for Atomic Scattering and Dielectric Reflection

We now provide examples of how to find Mueller matrices. First, we show how to derive the Mueller matrix for atomic scattering. The resultant fields were found in the text. Figure 1.7 provides the unit vectors \hat{e}_ℓ and \hat{e}_r for the scattered radiation, and $\hat{e}_{0\ell}$ and \hat{e}_{0r} for the incident radiation. The field amplitudes were

$$E_\ell = \frac{\kappa^2 \alpha}{4\pi\epsilon_0 r} \cos \Theta E_{0\ell} \quad \text{and} \quad E_r = \frac{\kappa^2 \alpha}{4\pi\epsilon_0 r} E_{0r},$$

where $E_{0\ell}$ and E_{0r} are, respectively, the amplitudes of the incident fields parallel and perpendicular to the scattering plane. From these, we can straightforwardly relate the

¹⁴Most light in the environment of interest in remote sensing is unpolarized or linearly partially polarized.

scattered Stokes vector \mathbf{I} to the incident \mathbf{I}_0 , i.e., find the Mueller matrix for scattering: $\mathbf{I} = \mathbf{M}\mathbf{I}_0$. For example, the scattered Stokes parameter Q is

$$\begin{aligned}
 Q &= \frac{1}{2\mu_0 c} [E_\ell E_\ell^* - E_r E_r^*] \\
 &= \frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \frac{1}{2\mu_0 c} [E_{0\ell} E_{0\ell}^* \cos^2 \Theta - E_{0r} E_{0r}^*] \\
 &= \frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \left[\frac{I_0 + Q_0}{2} \cos^2 \Theta - \frac{I_0 - Q_0}{2} \right] \\
 &= \frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \left[-\frac{I_0}{2} \sin^2 \Theta + \frac{Q_0}{2} (1 + \cos^2 \Theta) \right] \\
 &= M_{21} I_0 + M_{22} Q_0 + M_{23} U_0 + M_{24} V_0,
 \end{aligned}$$

so

$$M_{21} = - \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \right] \frac{\sin^2 \Theta}{2}, \quad M_{22} = \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \right] \frac{1 + \cos^2 \Theta}{2}, \quad \text{and} \quad M_{23} = M_{24} = 0.$$

The other elements of the Mueller matrix can be found in the same manner, and the result is

$$\mathbf{M} = \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2} \sin^2 \Theta & 0 & 0 \\ -\frac{1}{2} \sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix}$$

If the incident light is unpolarized, the scattered light is linearly partially polarized, having a Stokes vector

$$\left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) \\ -\frac{1}{2} \sin^2 \Theta \\ 0 \\ 0 \end{pmatrix} I_0$$

and a degree of polarization P

$$P = \sqrt{\frac{Q^2 + U^2 + V^2}{I^2}} = \frac{\sin^2 \Theta}{1 + \cos^2 \Theta}$$

Note that $P = 1$ for $\Theta = 90^\circ$, i.e., right-angle scattered light from atoms is fully polarized. For scattering at other angles the scattered radiation is partially polarized. Thus, scattering can produce polarized light starting from unpolarized light.

Now consider reflection from a plane interface between two dielectrics, where $m_{\text{Incident}} < m_{\text{Transmitted}}$. We have $\mathbf{I}_r = \mathbf{M}\mathbf{I}_i$, and the Mueller matrix \mathbf{M} can easily be derived by using

Eq.s (1.41). The result is

$$\mathbf{M} = \begin{pmatrix} \rho_+ & \rho_- & 0 & 0 \\ \rho_- & \rho_+ & 0 & 0 \\ 0 & 0 & \rho_{33} & 0 \\ 0 & 0 & 0 & \rho_{33} \end{pmatrix}, \quad (1.62)$$

with

$$\rho_{\pm} = \frac{1}{2} \left[\frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} \pm \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} \right] = \frac{1}{2} [r_f^{\parallel} \pm r_f^{\perp}],$$

and

$$\rho_{33} = -\frac{\tan(\theta_i - \theta_t) \sin(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t) \sin(\theta_i + \theta_t)}.$$

We see that reflection from an interface between two dielectrics, e.g., air and water, can also produce polarized light starting from unpolarized light.

1.10.8 An Alternative Method for Deriving the Mueller Matrix

We conclude this appendix with an alternate method for finding \mathbf{M} . Given the 2×2 transformation matrix between the incident and scattered fields,¹⁵

$$\vec{E} = \begin{pmatrix} E_r \\ E_{\ell} \end{pmatrix} = \begin{pmatrix} T_{rr} & T_{r\ell} \\ T_{\ell r} & T_{\ell\ell} \end{pmatrix} \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix} = \mathbf{T} \vec{E}_0$$

there is a simple method of constructing the Mueller matrix. Consider the so-called *coherency matrix*

$$\mathbf{J} = \begin{pmatrix} \langle E_r E_r^* \rangle & \langle E_r E_{\ell}^* \rangle \\ \langle E_{\ell} E_r^* \rangle & \langle E_{\ell} E_{\ell}^* \rangle \end{pmatrix} = \left\langle \begin{pmatrix} E_r E_r^* & E_r E_{\ell}^* \\ E_{\ell} E_r^* & E_{\ell} E_{\ell}^* \end{pmatrix} \right\rangle$$

and the four matrices,¹⁶

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_4 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

¹⁵Note, \mathbf{T} is the *transformation* matrix, *not* the *transmission* matrix, i.e., it can apply equally well to reflection, transmission, and scattering.

¹⁶The first matrix is the unit matrix, $\mathbf{1}$, the others are related to the Pauli spin matrices usually denoted by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is easily verified that any 2×2 matrix can be expanded as a linear combination of Pauli matrices and the unit matrix, i.e.,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{2} [(a+d)\mathbf{1} + (b+c)\sigma_x + i(b-c)\sigma_y + (a-d)\sigma_z].$$

It is easy to verify that

$$\begin{aligned} \text{Tr}(\boldsymbol{\sigma}_i \boldsymbol{\sigma}_j) &= 2\delta_{ij}, \\ \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i &= \mathbf{1}, \end{aligned}$$

where Tr indicates the *trace* (the sum of the diagonal elements of the resulting matrix), $\delta_{ij} = 1$ if $i = j$ and zero otherwise, and $\mathbf{1}$ is the identity matrix (same as $\boldsymbol{\sigma}_1$). Let us write

$$\mathbf{J} = \frac{1}{2} \sum_{i=1}^4 I_i \boldsymbol{\sigma}_i,$$

form the matrix product ($\boldsymbol{\sigma}_1 \mathbf{J}$), and take the trace of the result:

$$\text{Tr}(\boldsymbol{\sigma}_1 \mathbf{J}) = \frac{1}{2} \sum_{i=1}^4 I_i \text{Tr}(\boldsymbol{\sigma}_1 \boldsymbol{\sigma}_i) = I_1.$$

Direct calculation of $\text{Tr}(\boldsymbol{\sigma}_1 \mathbf{J})$ yields $\langle E_r E_r^* \rangle + \langle E_\ell E_\ell^* \rangle$, the Stokes parameter I . In a similar manner, we can show that

$$Q = -\text{Tr}(\boldsymbol{\sigma}_2 \mathbf{J}), \quad U = +\text{Tr}(\boldsymbol{\sigma}_3 \mathbf{J}), \quad V = -\text{Tr}(\boldsymbol{\sigma}_4 \mathbf{J}),$$

so, writing the Stokes vector as

$$\mathbf{S} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix},$$

we have shown that

$$\mathbf{J} = \frac{1}{2} \sum_{i=1}^4 (-1)^{i+1} S_i \boldsymbol{\sigma}_i. \quad (1.63)$$

Now, direct matrix multiplication shows that we can write \mathbf{J} in the form

$$\mathbf{J} = \left\langle \begin{pmatrix} E_r \\ E_\ell \end{pmatrix} \begin{pmatrix} E_r & E_\ell \end{pmatrix} \right\rangle,$$

and the row matrices

$$\begin{pmatrix} E_r^* & E_\ell^* \end{pmatrix} = \begin{pmatrix} E_r & E_\ell \end{pmatrix}^* = \begin{pmatrix} E_{0r} & E_{0\ell} \end{pmatrix}^* \mathbf{T}^\dagger,$$

where \mathbf{T}^\dagger is the Hermitian conjugate of \mathbf{T} :

$$\mathbf{T}^\dagger = \begin{pmatrix} T_{rr}^* & T_{\ell r}^* \\ T_{r\ell}^* & T_{\ell\ell}^* \end{pmatrix}.$$

Defining the incident coherency matrix \mathbf{J}_0 through

$$\mathbf{J}_0 = \left\langle \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix} \begin{pmatrix} E_{0r}^* & E_{0\ell}^* \end{pmatrix} \right\rangle,$$

direct calculation shows that

$$\mathbf{J} = \mathbf{T} \mathbf{J}_0 \mathbf{T}^\dagger.$$

Using Eq. (1.63) with $\mathbf{J}_0 = \frac{1}{2} \sum_{i=1}^4 (-1)^{i+1} S_{0i} \boldsymbol{\sigma}_i$, we have

$$\frac{1}{2} \sum_{i=1}^4 (-1)^{i+1} S_i \boldsymbol{\sigma}_i = \mathbf{T} \frac{1}{2} \sum_{i=1}^4 (-1)^{i+1} S_{0i} \boldsymbol{\sigma}_i \mathbf{T}^\dagger,$$

and multiplying by $\boldsymbol{\sigma}_j$ and taking the trace

$$(-1)^{j+1} S_j = \frac{1}{2} \sum_{i=1}^4 (-1)^{i+1} S_{0i} \text{Tr}(\boldsymbol{\sigma}_j \mathbf{T} \boldsymbol{\sigma}_i \mathbf{T}^\dagger),$$

or

$$S_j = \frac{1}{2} \sum_{i=1}^4 (-1)^{i-j} \text{Tr}(\boldsymbol{\sigma}_j \mathbf{T} \boldsymbol{\sigma}_i \mathbf{T}^\dagger) S_{0i}.$$

This shows that the Mueller matrix element M_{ji} is given by

$$M_{ji} = \frac{1}{2} (-1)^{i-j} \text{Tr}(\boldsymbol{\sigma}_j \mathbf{T} \boldsymbol{\sigma}_i \mathbf{T}^\dagger). \quad (1.64)$$

Although formulated using notation based on scattering, this method is applicable to determining the Mueller matrix for any optical element for which \mathbf{T} is known, e.g., linear polarizers, wave retarders, etc.

As an example of the use of Eq. (1.64), consider atomic scattering where,

$$\mathbf{T} = \frac{\kappa^2 \alpha}{4\pi \epsilon_0 r} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} = \mathbf{T}^\dagger.$$

Then, for example

$$\begin{aligned} M_{21}(\Theta) &= -\frac{1}{2} \text{Tr}(\boldsymbol{\sigma}_2 \mathbf{T} \boldsymbol{\sigma}_1 \mathbf{T}^\dagger) \\ &= -\frac{1}{2} \left(\frac{\kappa^2 \alpha}{4\pi \epsilon_0 r} \right)^2 \text{Tr} \left[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \right] \\ &= -\frac{1}{2} \left(\frac{\kappa^2 \alpha}{4\pi \epsilon_0 r} \right)^2 \sin^2 \Theta, \end{aligned}$$

as before.¹⁷

For a second example we derive the Mueller matrix for total internal reflection, i.e., reflection when the incident angle is greater than the critical angle. In this case, from Section 1.6,¹⁸

$$\mathbf{T} = \begin{pmatrix} \exp[-2i\phi^\perp] & 0 \\ 0 & \exp[i(\pi - 2\phi^\parallel)] \end{pmatrix} = \begin{pmatrix} \exp[-2i\phi^\perp] & 0 \\ 0 & -\exp[-2i\phi^\parallel] \end{pmatrix},$$

and $\mathbf{T}^\dagger = \mathbf{T}^*$. In this case, for example,

$$M_{34} = -\frac{1}{2}Tr(\boldsymbol{\sigma}_3 \mathbf{T} \boldsymbol{\sigma}_4 \mathbf{T}^*) = \sin 2(\phi^\parallel - \phi^\perp)$$

In a similar manner the full Mueller matrix is found to be

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos 2(\phi^\parallel - \phi^\perp) & \sin 2(\phi^\parallel - \phi^\perp) \\ 0 & 0 & -\sin 2(\phi^\parallel - \phi^\perp) & \cos 2(\phi^\parallel - \phi^\perp) \end{pmatrix}$$

Thus, if the incident light is linearly polarized with a nonzero U component and $V = 0$, the reflected V component will be $-\sin 2(\phi^\parallel - \phi^\perp)U_{\text{Incident}}$, and the reflected light will be elliptically polarized as claimed in the text. We shall see later that when unpolarized solar radiation is scattered in an ocean-atmosphere system consisting of only Rayleigh scatterers, e.g., air molecules and water molecules, the only process by which elliptically polarized light can be generated in this environment is through upward-scattered radiation within the water undergoing total internal reflection at the surface.

1.11 Bibliographic Notes

Graduate-level texts that are usually considered “chapter and verse” of classical electrodynamics are Jackson, *Classical Electrodynamics* [Jackson, 1975] and Panofsky and Phillips *Classical Electricity and Magnetism* [Panofsky and Phillips, 1964]. For students specializing in physics, these provide a clear exposition of the subject. A valuable review of the electrodynamics required for optics in general is provided in the first chapter of *Principles of Optics* by Born and Wolf [Born and Wolf, 1975]. On the introductory level, there are

¹⁷It is a simple matter to show that then T_{12} and T_{21} are both equal to zero, i.e., when \mathbf{T} is diagonal, that $M_{ji} = 0$ when $j = 1$ or 2 and $i = 3$ or 4 .

¹⁸In Section 1.6 we used the superscripts \perp and \parallel to denote cases where the electric vector is perpendicular or parallel to the plane of incidence, respectively

several excellent books, the best two being *the Feynman Lectures on Physics, Volume II* by Feynman, Leighton, and Sands [Feynman et al., 1964], and *Introduction to Electrodynamics* by Griffiths [Griffith, 1989]. The latter two have been used extensively in the preparation of this chapter. In particular, the development of the properties of dielectrics through simple charge configurations, the electric fields of which can be computed using Gauss' law, was based on examples in these texts. The derivation of \vec{P} for scattering by small spheres was taken from Griffiths' *Introduction to Electrodynamics* [Griffith, 1989]. Moving from the classical to the quantum domain in electrodynamics is a large step, which fortunately is not really needed for our purposes. For the student with a basic knowledge of quantum mechanics, we recommend the *Quantum Electronics* by Yariv [Yariv, 1988], e.g., for a discussion of the quantum theory of polarizability and quantization of the electromagnetic field. A very readable non-mathematical account of the subject is found in *QED* by Feynman [Feynman, 1985].

In graduate physics courses, radiation from accelerating charges is treated only after a firm foundation of electrodynamics is in place. Thus, we have chosen to simply state the main result and then apply it to situations that will prove useful in the study of environmental optics. Most texts on electrodynamics treat this particular subject well, and we have found the development in *Classical Electricity and Magnetism* by Barger and Olsson [Barger and Olssen, 1986] to be particularly valuable. Given the general formula for the fields from accelerating charges, the discussion of the scattering by atoms and molecules then follows with little difficulty.

Finally, the development of the Stokes representation of polarized light in the Appendix, that was based on a set of four measurements using polarizers and quarter wave plates was inspired by a similar development in *Absorption and Scattering of Light by Small Particles*, by Bohren and Huffman [Bohren and Huffman, 1983], and the relationship between the coherency matrix and the Stokes vector, developed from a similar approach taken by *States Waves and Photons: A Modern Introduction to Light* by Simmons and Guttman [Simmons and Guttman, 1970].

Chapter 2

Radiometry and Radiative Transfer

2.1 Introduction

Now we develop the framework for understanding the propagation of electromagnetic radiation in an absorbing and scattering medium, e.g., the ocean-atmosphere system. This requires characterizing the properties of the radiation and developing the basic law governing its transport through the medium. We begin by describing radiation detectors that can be used to measure radiant power, a quantity that is proportional to the time-averaged Poynting vector. This quantity is used to define various radiometric quantities that characterize the radiation, the most fundamental being the *radiance*. Then the basic optical properties influencing light propagation in a medium, scattering, absorption, etc., are defined and used to develop the *radiative transfer equation*. This equation governs the transport of radiance through an absorbing-scattering-emitting medium. Several solution techniques are described in some detail, and the first-order solution for a homogeneous slab with plane-parallel boundaries is presented in Appendix 1. The first-order solution is the starting point for analysis of radiative transfer in remote sensing and is used extensively throughout the rest of this work.

2.2 Detection of Electromagnetic Radiation

One of the earliest, and simplest to understand, methods of detecting electromagnetic radiation relies on the heating that occurs upon exposure to it. Consider a small wafer, coated with a flat black material (to reduce reflection), exposed to a beam of radiation. The wafer will absorb a significant portion of the radiation and the associated absorbed energy will heat the wafer causing its temperature to increase. The temperature increase can be measured by a thermometer, e.g., a thermocouple. Assume that the radiation falls on the wafer at normal incidence and is totally absorbed. Since the amount of energy absorbed is proportional to the time-averaged Poynting vector, we have that the absorbed energy will be related to the temperature change (ΔT) through

$$mc_P\Delta T = |\langle \vec{S} \rangle_{\text{Ave}}| A \Delta t,$$

where m is the wafer mass, c_P is its specific heat at constant pressure, A is its area, and Δt is the time over which the wafer is exposed to the radiation. Clearly,

$$|\langle \vec{S} \rangle_{\text{Ave}}| = \frac{mc_P\Delta T}{A\Delta t},$$

and such a sensor, and the associated thermometer, can be used to effect a straightforward measurement of the time-averaged Poynting vector, although typically with a slow response time.

The radiation detectors used in remote sensing are more complex. They utilize materials that have the property that when light falls on them and is absorbed, a quantum transition takes place within the detector material that yields an electron freed in some manner to contribute to a flow of electricity. For example, when electromagnetic radiation is incident on the surface of a metal, electrons are ejected from the metal by a process known as the *photoelectric effect*. Albert Einstein's explanation of the experimental observations regarding the photoelectric effect earned him the Nobel Prize. A detector employing the photoelectric effect is called a *phototube*.

Einstein hypothesized that when light falls on the surface of a metal it can give up energy to the electron only in discrete units called quanta or photons, the energy of which is given by $\hbar\omega$, where ω is the angular frequency of the light. In this picture, light was imagined to consist of a stream of photons, and he modeled the photoelectric effect as a mechanical-like collision between a photon and an electron. The collision was assumed to be elastic, i.e., the sum of the energies of the electron and photon is conserved. In the simplest terms (Figure 2.1), the photon collides with the electron in a metal plate (cathode), gives its energy to the electron, and the electron escapes the cathode. The electron is then accelerated to the positively charged metal plate (the anode) constituting

an electric current that will be recorded on the ammeter. The negatively charged metal is usually referred to as the *photocathode*, the ejected electron as the *photoelectron*, and the current as the *photocurrent*. The *photocurrent* is determined by the rate at which electrons pass between the photocathode and anode. This photocurrent is proportional to the time-average Poynting vector of the incident light. In Einstein's picture, the photocurrent was interpreted as being proportional to the rate (N) at which photons are incident on the photocathode surface. This required the Poynting vector be interpreted as a flux of photons as described in Chapter 1:

$$\mathcal{N} = \frac{1}{\hbar\omega} |\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|,$$

where \mathcal{N} is the flux of photons. So the rate at which photons fall on the surface (N) is $N = \mathcal{N}A_D$, where A_D is the area of the surface of the detector.¹

Other types of detectors behave in essentially the same way, but are based on different phenomena. For example, when a photon gives its energy to an electron in an insulator or a semiconductor it can cause the electron to join the *conduction* electrons in the material, increasing the material's ability to conduct electricity. Detectors that operate on this principal are called *photoconductive* detectors. In both photoelectric and photoconductive detectors, the photon gives its energy to an electron, which is then detected in some manner. The resulting effect is proportional to the number of photons absorbed. If the effect is a current (it usually is) then the current is proportional to the *rate* at which photons fall on the surface (N), and so, proportional to the time-averaged Poynting vector of the incident light.

For the rest of our discussion we will take as given that detectors are available that, when illuminated at near-normal incidence by a beam of radiation, respond in a manner that can be calibrated to yield the time-average Poynting vector of the radiation, e.g., the current in the detector i_D is proportional to $|\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|$:

$$i_D \propto |\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|.$$

Radiation detectors such as these are often referred to as *square-law* detectors because they respond to the square of the electric field of the radiation. In addition to $|\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|$, in order to specify more completely the electromagnetic radiation from a source, e.g., a candle or the sky, we need to specify its frequency distribution – its *spectrum*. This can be accomplished in a variety of ways, e.g., passing the radiation through a colored filter that absorbs all radiation incident upon it except for wavelengths in a narrow band $\Delta\lambda$ which

¹Note that $\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} = \hat{\xi} |\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|$, where $\hat{\xi}$ is in the direction of propagation, so we have implicitly assumed here that the normal to the detector surface (\hat{n}_D) is parallel to $\hat{\xi}$, i.e., A_D is oriented perpendicular to the beam. If $|\hat{\xi} \bullet \hat{n}_D| \neq 1$, then $N = \mathcal{N}A_D |\hat{\xi} \bullet \hat{n}_D|$.

it passes completely. Thus, we will also take as given that filters are available that pass a band of wavelengths $\Delta\lambda$ (or frequencies $\Delta\omega$) perfectly, and absorb radiation at all other wavelengths.

In what follows, we consider a light detecting device. It consists of a radiation detector that produces a current proportional to $|\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|$ at the detector surface, and it has a spectral filter that transmits only a narrow band of wavelengths $\Delta\lambda \ll \lambda$. If desired, it can also be equipped with polarizers and quarter-wave plates to determine the polarization state of the light. The detecting surface has an area A_D (ΔA_D or dA_D , if it is considered to be infinitesimal). For simplicity, we will refer to this arrangement as simply *a* detector or *the* detector.

For such a detector illuminated by a plane (or spherical) wave generated by a single source, the current is proportional to the square of the electric field amplitude $|\vec{E}_0|^2$; however, except in special situations, like viewing laser light, in the natural environment the detector will be illuminated by radiation emitted by a collection of independent sources (e.g., radiation emitted by individual atoms in a heated gas). In this case, the current is proportional to the sum of the individual Poynting vectors of the radiation from each source. To justify this assertion, consider the example of a set of sources generating electromagnetic waves propagating in the z direction. If the field of the i^{th} source is $\vec{E}_i = \vec{E}_{i0} \exp[-i(\kappa z - \omega t + \phi_i)]$, the time-averaged Poynting vector will be

$$\begin{aligned}
 \langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} &= \frac{1}{2\mu_0} \left(\sum_{i=1}^N \vec{E}_i \right) \times \left(\sum_{i=1}^N \vec{B}_i^* \right) \\
 &= \frac{\hat{e}_z c \epsilon_0}{2} \left(\sum_{i=1}^N \vec{E}_i \right) \cdot \left(\sum_{i=1}^N \vec{E}_i^* \right) \\
 &= \frac{\hat{e}_z c \epsilon_0}{2} \left(\sum_{i=1}^N (\vec{E}_i \cdot \vec{E}_i^*) + \sum_{i=1}^N \sum_{j \neq i}^N \vec{E}_i \cdot \vec{E}_j^* \right) \\
 &= \frac{\hat{e}_z c \epsilon_0}{2} \left(\sum_{i=1}^N |\vec{E}_{i0}|^2 + \sum_{i=1}^N \sum_{j \neq i}^N \vec{E}_{i0} \cdot \vec{E}_{j0} \exp[i(\phi_i - \phi_j)] \right) \\
 &= \frac{\hat{e}_z c \epsilon_0}{2} \left(\sum_{i=1}^N |\vec{E}_{i0}|^2 + \sum_{i=1}^N \sum_{j < i}^N 2 \vec{E}_{i0} \cdot \vec{E}_{j0} \cos(\phi_i - \phi_j) \right).
 \end{aligned}$$

In the sum in the second term the cosine will randomly have values ranging between 1 and

–1 and this term will sum to zero if there is no fixed phase between the individual sources.² Such a field might be produced by a set of identical sources in random motion or having random positions. The light from such a collection of sources is said to be *incoherent*. When the second term is zero,

$$\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} = \sum_{i=1}^N \langle \vec{S}_i(\vec{r}) \rangle_{\text{Ave}}.$$

All of the sources that we will deal with in this work are incoherent.

We take the current from a detector equipped with a spectral filter at λ_0 to be

$$\Delta i_D(\lambda_0) = \mathcal{K}(\lambda_0) \Delta \left(|\langle \vec{S}(\vec{r}, \lambda_0) \rangle_{\text{Ave}}| \right) A_D,$$

where as above, $\Delta \left(|\langle \vec{S}(\vec{r}, \lambda_0) \rangle_{\text{Ave}}| \right)$ is the sum of the time-averaged Poynting vectors having wavelengths (or frequencies) within $\Delta\lambda_0$ (or $\Delta\omega_0$). Note that $\Delta \left(|\langle \vec{S}(\vec{r}, \lambda_0) \rangle_{\text{Ave}}| \right) \times A_D$ is the radiant power $\Delta\mathcal{P}(\lambda_0)$ within $\Delta\lambda_0$ received by the detector. \mathcal{K} is a proportionality constant, usually called the *calibration constant*. It could be determined by utilizing the simple thermal detector (described at the beginning of this section) to measure $|\langle \vec{S} \rangle_{\text{Ave}}|$ for a set of stable sources, e.g., “standard lamps,” then viewing these sources with the detector in question. Treating $|\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}|$ as a continuous function of λ , we can define the *spectral radiant power* as

$$P(\lambda_0) \equiv \frac{\Delta\mathcal{P}(\lambda_0)}{\Delta\lambda_0} = \frac{1}{\mathcal{K}} \frac{\Delta i_D(\lambda_0)}{\Delta\lambda_0}. \quad (2.1)$$

Thus, measuring $\Delta i_D(\lambda_0)$ gives $P(\lambda_0)$. The total radiant power within a spectral band from λ_1 to λ_2 is just

$$\mathcal{P}(\lambda_1 \rightarrow \lambda_2) = \int_{\lambda_1}^{\lambda_2} P(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{1}{\mathcal{K}} \left(\frac{di_D}{d\lambda} \right) d\lambda.$$

²Perhaps this is easier to visualize if all of the sources have the same amplitude E_0 . In this case,

$$\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}} = \frac{\hat{e}_z c \epsilon_0}{2} \sum_{i=1}^N |\vec{E}_0|^2 \left(1 + 2 \sum_{j < i}^N \cos(\phi_i - \phi_j) \right),$$

and we can expect the cosine to sum to zero, when ϕ_i and ϕ_j are independent.

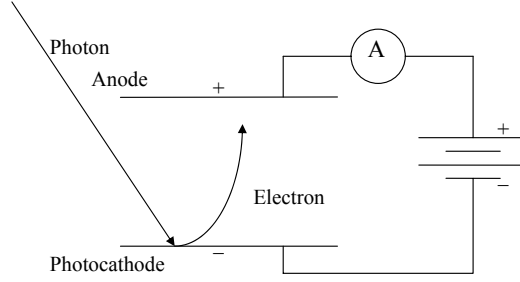


Figure 2.1: Schematic of a phototube. A photon strikes the photocathode and ejects an electron. The electron is accelerated toward the positively charged anode. A stream of such photons will be recorded as a current through the ammeter “A.” The current is proportional to the rate at which photons strike the plate.

2.3 Radiometry

In the last section we defined the spectral radiant power $P(\lambda)$ and related it to the totality of the time-average Poynting vectors (from each independent source) within a narrow band of wavelengths $\Delta\lambda$. All radiometric quantities are defined through the spectral radiant power and they will now be developed. For convenience the symbol λ will be suppressed from all radiometric quantities, unless its deletion would result in confusion.

2.3.1 Radiant Intensity

Sources of radiation that are either very far away, e.g., the stars, or by their nature are very small, appear as *point* sources to an observer. An observer can use a detector to measure the power from the source, but the power received depends on the distance to the source. However, to make a measurement that provides a characteristic of the source, this distance must be taken into consideration. This is accomplished by dividing the measured spectral radiant power by the *solid angle* subtended by the *detector at the source*.³ The *radiant intensity* (J) of the source is the power per unit solid angle emitted by the source. If \hat{n}_D is the unit normal to the detector of area dA_D , as shown in Figure 2.2, then the intensity

³The solid angle is the three-dimensional analog of an ordinary angle in plane geometry. For the latter, consider a circle of radius r and an arc of length s along the circle. Then the angle θ subtended by the arc (at the center of the circle) is given by $\theta = s/r$. In analogy, consider a sphere of radius R and a (closed) curve of area A on its surface. Then the solid angle Ω subtended by A at the center of the sphere is given by $\Omega = A/R^2$. The angle θ , although dimensionless, is usually given the unit *radians* and the solid angle Ω , also dimensionless, is given the unit *steradians* (Sr).

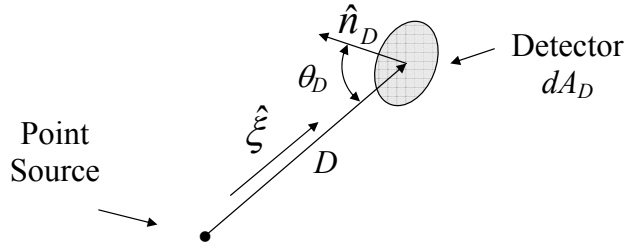


Figure 2.2: A point source illuminates a detector of area dA_D in the direction $\hat{\xi}$, a distance D away. \hat{n}_D is the normal to the detector surface. The angle θ_D is given by $\cos \theta_D = |\hat{n}_D \bullet \hat{\xi}|$.

of the source in the direction specified by the unit vector $\hat{\xi}$ is given by

$$\begin{aligned}
 J(\hat{\xi}) &= \frac{dP(\hat{\xi})}{d\Omega(\hat{\xi})} \\
 &= \frac{dP(\hat{\xi})}{|\hat{n}_D \bullet \hat{\xi}| dA_D / D^2} \\
 &= \frac{dP(\hat{\xi})}{\cos \theta_D dA_D / D^2}
 \end{aligned} \tag{2.2}$$

$dP(\hat{\xi})$ is the power received by the detector in the solid angle $d\Omega(\hat{\xi})$ around the direction $\hat{\xi}$. (Note that $\hat{\xi}$ designates the direction in which energy is propagating.) The total power emitted by the source in all directions is

$$P = \int_{\text{All } \hat{\xi}} J(\hat{\xi}) d\Omega(\hat{\xi}).$$

If there are no losses of radiant power in the space between the source and the observer (in any direction) then the total power is independent of the distance from the source. The only way this is possible is if the power within any fixed solid angle (subtended at the source) is independent of the distance from the source. This implies that J is independent of the distance from the source. Thus, in the absence of losses between a point source and an observer, the radiant intensity is independent of the distance between them, and therefore is characteristic of the *source*. The most common symbol for intensity is I ; however, we will use the symbol J to avoid confusion with the first element of the Stokes vector.

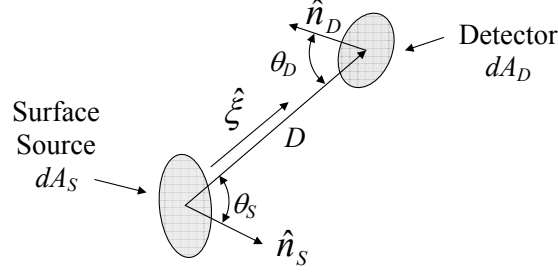


Figure 2.3: A surface source of area dA_S and normal \hat{n}_S illuminates a detector of area dA_D , in the direction $\hat{\xi}$, a distance D away. \hat{n}_D is the normal to the detector surface.

2.3.2 Radiance

When the source is extended, the radiant intensity is not a useful concept, e.g., which point on the source is used as reference for the solid angle? For an extended source another quantity is needed. This quantity is called *radiance*. The *radiance* leaving a surface is the radiant power emitted (reflected or scattered) per unit area (projected normal to the direction of energy flow) per unit solid angle. Consider the surface element dA_S of an extended source shown in Figure 2.3. If the power received by the detector is $d^2P(\hat{\xi})$, then the radiance of the surface (assuming no loss along D) is defined to be

$$L(\hat{\xi}) \equiv \frac{d^2P(\hat{\xi})}{\cos \theta_S dA_S d\Omega_D}, \quad (2.3)$$

where

$$d\Omega_D = \frac{\cos \theta_D dA_D}{D^2}$$

is the solid angle subtended by the detector at any point on the source. Radiance is of primary importance in radiometry because in the absence of losses, it is a conserved quantity in the sense that it is independent of the distance from the surface. To illustrate this, consider a surface S which consists of a set of independent⁴ (uniformly distributed) point sources, each of intensity J_0 . Let n be the number of sources per unit area of the surface so the total number in an area A is $N_A = nA$. Consider two of these sources, symmetrically spaced with respect to the origin, as shown in Figure 2.4. A detector of area dA_D is located at $(0, y_D, z_D)$ and its normal \hat{n}_D lies in the y - z plane. The power falling

⁴The sources are “independent” in the sense of Section 2.2, i.e., in combination they produce incoherent radiation.

on the detector from these two sources is

$$dP = J_0 dA_D \left[\frac{\hat{\xi} \cdot \hat{\xi}_1}{r_1^2} + \frac{\hat{\xi} \cdot \hat{\xi}_2}{r_2^2} \right],$$

where the r_i are the distances from the individual sources to the detector. From Figure 2.4 it is seen that

$$\begin{aligned} \hat{\xi}_i &= \frac{x_i \hat{e}_x + (y_D - y_i) \hat{e}_y + z_D \hat{e}_z}{r_i}, \\ r_i^2 &= x_i^2 + (y_D - y_i)^2 + z_D^2, \\ \hat{\xi} &= \frac{y_D \hat{e}_y + z_D \hat{e}_z}{D}, \end{aligned}$$

and

$$D^2 = [y_D^2 + z_D^2],$$

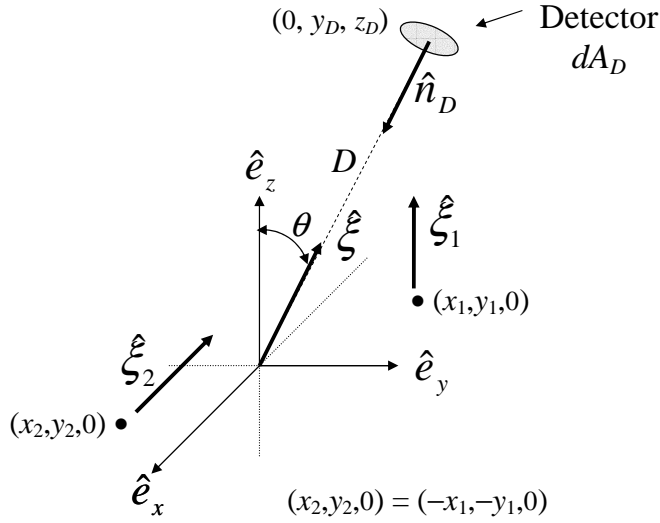


Figure 2.4: Two point sources, symmetrically located with respect to the origin, illuminate a detector of area dA_D in the direction $\hat{\xi}$ a distance D from the origin. \hat{n}_D is the normal to the detector surface, with $\hat{n}_D = -\hat{\xi}$. The unit vectors $\hat{\xi}_1$ and $\hat{\xi}_2$ are directed toward the detector. The detector is located at the point $(0, y_D, z_D)$.

where $i = 1$ or 2 . Thus, dP becomes

$$dP = \frac{J_0 dA_D}{D} \left[\frac{y_D(y_D - y_1) + z_D^2}{r_1^3} + \frac{y_D(y_D - y_2) + z_D^2}{r_2^3} \right].$$

Noting that $x_2 = -x_1$ and $y_2 = -y_1$, performing binomial expansions on r_1^{-3} and r_2^{-3} , we find

$$dP = \frac{J_0 dA_D}{D^2} \left[2 - 3 \left(\frac{x_1}{D} \right)^2 - 3(1 + 2 \sin^2 \theta) \left(\frac{y_1}{D} \right)^2 + \dots \right]. \quad (2.4)$$

If the second order terms in x_1/D and y_1/D are ignored, i.e., $x_1 \ll D$ and $y_1 \ll D$,

$$dP = \frac{2J_0 dA_D}{D^2},$$

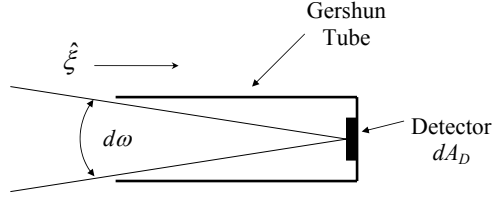


Figure 2.5: A Gershun tube limits the field of view of a detector of area dA_D to a solid angle $d\omega$. The solid angle $d\omega$ is simply the area of the open end divided by the square of the length of the tube. We shall see later that the radiance propagating in the direction $\hat{\xi}$ is the radiant power measured by the detector divided by the product $dA_d \times d\omega$.

and each point source contributes the same power. Now, let's determine the radiance of the surface. Operationally, we want to examine an area dA_S of the surface and measure the power received by dA_D from that area. Therefore, it is necessary to restrict the field of view (FOV) of the detector. This can be accomplished most simply by means of a *Gershun tube* as shown in Figure 2.5. The Gershun tube is simply a cylindrical tube through which the detector views the surface. A detector equipped with such a tube is called a radiometer. Attachment of the Gershun tube to the detector limits the field of view of the detector to an area dA_S of the surface (Figure 2.6), so the power received by the detector is from $n dA_S$ sources. The radiance of the surface is by definition

$$L(\hat{\xi}) = \frac{d^2 P(\hat{\xi})}{\cos \theta dA_S d\Omega_D}. \quad (2.5)$$

Again in the absence of losses, all the power emitted toward dA_D falls on the detector, so the radiance is given by

$$L(\hat{\xi}) = \frac{ndA_S J_0 dA_D / D^2}{\cos \theta dA_S d\Omega_D}.$$

But,

$$d\Omega_D = \frac{dA_D}{D^2},$$

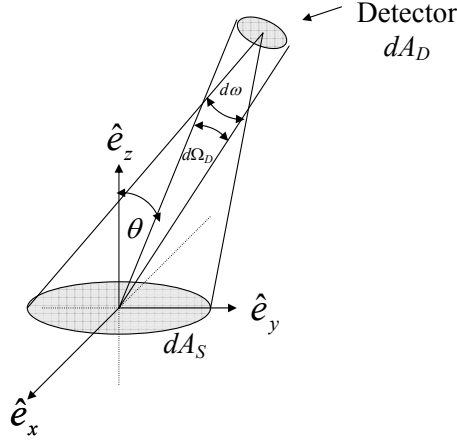


Figure 2.6: The detector in Figure 2.4 is able to view only an area dA_S of the $x - y$ plane because its field of view has been restricted to a solid angle $d\omega$, e.g., by a Gershun tube (not shown). The solid angle $d\Omega_D$ is that subtended by the detector by *any* point in dA_S .

so,

$$L(\hat{\xi}) = \frac{nJ_0}{\cos \theta}, \quad (2.6)$$

and L is independent of the solid angles, as well as the distance between the source and the radiometer. It depends only on the properties of the individual sources.⁵

It is useful to try to eliminate specific references to the surface under examination from these relationships. Consider again the power measured by the detector. This is

$$\begin{aligned} d^2P(\hat{\xi}) &= \frac{ndA_S J_0 dA_D}{D^2} \\ &= \frac{L(\hat{\xi}) \cos \theta dA_S dA_D}{D^2}. \end{aligned}$$

Since

$$\frac{\cos \theta dA_S dA_D}{D^2} = d\omega dA_D,$$

$$d^2P(\hat{\xi}) = L(\hat{\xi}) dA_D d\omega$$

and hence, the power measured by the detector at the end of the tube is proportional to the radiance of the surface. Note that the $d\omega$ in the above equation is the solid angle of view of the detector as shown in Figure 2.5, and should not be confused with the spectral band pass of the filter, i.e., $\Delta\omega$. If we were unaware of the existence of the surface or if,

⁵If J_0 is proportional to $\cos \theta$, then $L(\hat{\xi})$ is independent of $\hat{\xi}$, and the surface is said to be a *lambertian* emitter.

as in the case of the sky, a physical surface does not exist, this equation can be used to provide an alternate definition of radiance based *only* on the properties of the radiometer:

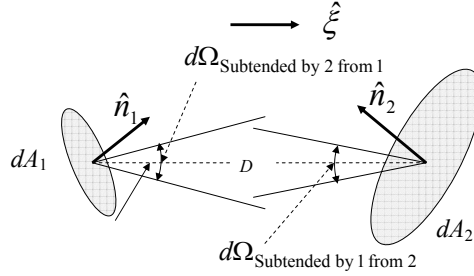
$$L(\hat{\xi}) = \frac{d^2 P(\hat{\xi})}{dA_D d\omega}. \quad (2.7)$$

The radiance defined in Eq. (2.7) is sometimes called the *apparent radiance* of the surface, the *aperture radiance* of the detector, or the *field radiance*.⁶

There are conditions that must be met for the radiometer to provide an accurate radiance. Looking back over our derivation of Eq. (2.7), we find that the validity of Eq. (2.5) required ignoring the terms $(x_1/D)^2$ and $(y_1/D)^2$ in Eq. (2.4). Consider Eq. (2.4) for $\theta = 0$ and let the radius of the Gershun tube be R . Then if the tube is close to the surface, the sources farthest from the axis have $x_1^2 + y_1^2 = R^2$, and

$$dP = \frac{J_0 dA_D}{D^2} \left[2 - 3 \left(\frac{R}{D} \right)^2 - \dots \right],$$

⁶The basis of the definition of the field radiance in Eq. (2.7) is a result concerning the area-solid angle product. Consider two areas and the solid angles that each subtends at the other as shown in the figure below.



The figure shows two small areas and the solid angles that each subtends at the other. The solid angle subtended by dA_2 at the center of dA_1 is $d\Omega_{\text{Subtended by 2 from 1}}$, and the solid angle subtended by dA_1 at the center of dA_2 is $d\Omega_{\text{Subtended by 1 from 2}}$. The individual solid angles are

$$d\Omega_{\text{Subtended by 1 from 2}} = \frac{|\hat{n}_1 \cdot \hat{\xi}| dA_1}{D^2} \quad \text{and} \quad d\Omega_{\text{Subtended by 2 from 1}} = \frac{|\hat{n}_2 \cdot \hat{\xi}| dA_2}{D^2}.$$

Eliminating D from these yields

$$|\hat{n}_2 \cdot \hat{\xi}| dA_2 d\Omega_{\text{Subtended by 1 from 2}} = |\hat{n}_1 \cdot \hat{\xi}| dA_1 d\Omega_{\text{Subtended by 2 from 1}}, \quad (2.8)$$

which leads directly to the field radiance in Eq. (2.7).

but since $\Delta\omega = \pi R^2/D^2$,

$$dP = \frac{J_0 dA_D}{D^2} \left[2 - 3 \left(\frac{\Delta\omega}{\pi} \right) - \dots \right].$$

Thus for Eq. (2.7) to provide the correct radiance $\Delta\omega \ll 1$. When this is satisfied the radiometer can provide the correct radiance using Eq. (2.7), even when the distance between the surface and the aperture of the radiometer is essentially zero.

Recall Eq. (2.6). It shows that as long as J_0 and θ are constant, the radiance is constant regardless of the distance from the surface. Thus, if the radiometer is aimed at a physical surface, because the radiance it measures is independent of the distance from the surface, it will record the actual radiance leaving the surface (in the absence of losses of power between the radiometer and the surface, the origin of which losses will be discussed in detail later in this chapter). For example, take a single lens reflex camera with a built-in exposure meter and aim the camera toward a blank wall (free of pictures, etc.). Note the exposure setting and walk toward the wall. You will find that the exposure remains unchanged as you move toward the wall (the apparent radiance is constant), and in the limit that you could physically place the camera on the surface without creating shadows etc., the exposure would still remain the same. The camera exposure meter is a radiometer and the aperture radiance is the same as the surface radiance.

The radiometric quantities we have been examining thus far all have counterparts when the polarization of the radiation is examined. Consider a detector that is equipped with a polarizer and quarter-wave plate, define a reference system specified by the unit vectors \hat{e}_r and \hat{e}_ℓ , where $\hat{\xi} = \hat{e}_r \times \hat{e}_\ell$, and perform the sequence of measurements described in the appendix to Chapter 1. Denoting by $P_\ell(\hat{\xi})$ the radiant power measured when the polarizer has pass direction along \hat{e}_ℓ , $P_r(\hat{\xi})$ when the pass direction is along \hat{e}_r , etc., then we can define the Stokes vector for intensity by⁷

$$\mathbf{I}_J(\hat{\xi}) = \begin{pmatrix} I_J(\hat{\xi}) \\ Q_J(\hat{\xi}) \\ U_J(\hat{\xi}) \\ V_J(\hat{\xi}) \end{pmatrix} = \frac{1}{d\Omega(\hat{\xi})} \begin{pmatrix} dP_\ell(\hat{\xi}) + dP_r(\hat{\xi}) \\ dP_\ell(\hat{\xi}) - dP_r(\hat{\xi}) \\ 2dP_p(\hat{\xi}) - dP_\ell(\hat{\xi}) - dP_r(\hat{\xi}) \\ 2dP_q(\hat{\xi}) - dP_\ell(\hat{\xi}) - dP_r(\hat{\xi}) \end{pmatrix} \quad (2.9)$$

The radiance measured by a similarly equipped radiometer is

$$\mathbf{I}_L(\hat{\xi}) = \begin{pmatrix} I_L(\hat{\xi}) \\ Q_L(\hat{\xi}) \\ U_L(\hat{\xi}) \\ V_L(\hat{\xi}) \end{pmatrix} = \frac{1}{dA_D d\omega} \begin{pmatrix} d^2P_\ell(\hat{\xi}) + d^2P_r(\hat{\xi}) \\ d^2P_\ell(\hat{\xi}) - d^2P_r(\hat{\xi}) \\ 2d^2P_p(\hat{\xi}) - d^2P_\ell(\hat{\xi}) - d^2P_r(\hat{\xi}) \\ 2d^2P_q(\hat{\xi}) - d^2P_\ell(\hat{\xi}) - d^2P_r(\hat{\xi}) \end{pmatrix} \quad (2.10)$$

⁷Recall that the subscript “p” stands for a measurement with a linear polarizer oriented with pass direction along $\hat{e}_p = (\hat{e}_r + \hat{e}_\ell)/\sqrt{2}$ in the beam. The subscript “q” references power measured with a quarter wave plate followed by a linear polarizer with pass direction \hat{e}_p in the beam.

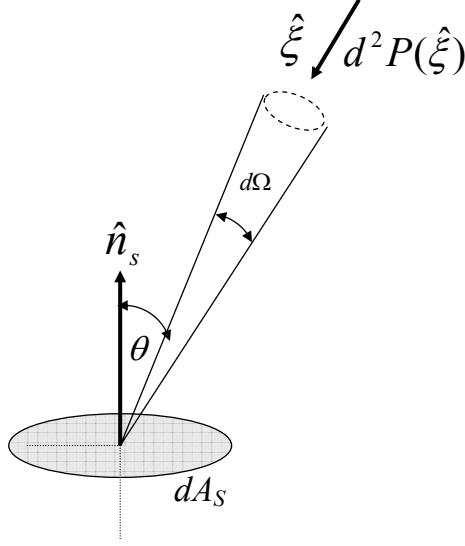


Figure 2.7: Radiant power $d^2P(\hat{\xi})$ within a solid angle $d\Omega(\hat{\xi})$ is incident on an element of surface area dA_S from the direction $\hat{\xi}$.

A radiometer that measures $\mathbf{I}_L(\hat{\xi})$ is often referred to as a polarimeter.

Thus far this discussion has been centered on describing the radiance leaving a surface due to emission (or reflection) of radiant power by (or from) the surface. It is of equal importance to be able to describe the angular distribution of radiant power incident on a surface. This is accomplished by defining the radiance incident on an element of the surface (of area dA_S) in a manner similar to that of the radiance emitted by an element of the surface. The geometrical setting is given in Figure 2.7, and the radiance incident on the surface is defined in analogy to Eq. (2.3) to be

$$L(\hat{\xi}) = \frac{d^2P(\hat{\xi})}{\cos \theta dA_S d\Omega(\hat{\xi})} \quad (2.11)$$

where $d^2P(\hat{\xi})$ is now the radiant power falling on dA_S from a range of directions contained in a solid angle $d\Omega(\hat{\xi})$ centered on the direction $\hat{\xi}$. This radiance can of course be measured by placing a radiometer at the position of interest on the surface (and replacing $\cos \theta dA_S$ by dA_D and $d\Omega$ by $d\omega$).

Radiance measured along a path depends on the refractive index along the path. The easiest way of showing this is to note that when passing through an interface between two

media of differing refractive index, such as air and water, the solid angle changes. Figure 2.8 shows radiant power propagating through an element of area dA_S on the interface. The radiance falling on dA_S from medium 2 is given by

$$L(\hat{\xi}_2) = \frac{d^2 P(\hat{\xi}_2)}{\cos \theta_2 dA_S d\Omega_2},$$

and the radiance leaving dA_S into medium 1 is

$$L(\hat{\xi}_1) = \frac{d^2 P(\hat{\xi}_1)}{\cos \theta_1 dA_S d\Omega_1},$$

where

$$\begin{aligned} d\Omega_1 &= \sin \theta_1 d\theta_1 d\phi_1, \\ d\Omega_2 &= \sin \theta_2 d\theta_2 d\phi_2. \end{aligned}$$

If the radiation is unpolarized, the transmitted power is related to the incident power through

$$d^2 P(\hat{\xi}_1) = t_f d^2 P(\hat{\xi}_2), \quad (2.12)$$

where t_f is the Fresnel transmittance of the interface for unpolarized light. The Fresnel reflectance r_f for unpolarized light is given by (Chapter 1)

$$r_f = \frac{1}{2} \left(\frac{\sin^2(\theta_1 - \theta_2)}{\sin^2(\theta_1 + \theta_2)} + \frac{\tan^2(\theta_1 - \theta_2)}{\tan^2(\theta_1 + \theta_2)} \right) = \frac{1}{2} (r_f^\perp + r_f^\parallel),$$

where θ_1 and θ_2 are related by Snell's Law,

$$m_1 \sin \theta_1 = m_2 \sin \theta_2.$$

The transmittance $t_f = 1 - r_f$. Hence,

$$L(\hat{\xi}_1) \cos \theta_1 dA_S d\Omega_1 = t_f L(\hat{\xi}_2) \cos \theta_2 dA_S d\Omega_2,$$

or

$$L(\hat{\xi}_1) \cos \theta_1 \sin \theta_1 d\theta_1 d\phi_1 = t_f L(\hat{\xi}_2) \cos \theta_2 \sin \theta_2 d\theta_2 d\phi_2.$$

Differentiating Snell's Law

$$m_1 \cos \theta_1 d\theta_1 = m_2 \cos \theta_2 d\theta_2,$$

and multiplying by $m_1 \sin \theta_1$ yields

$$m_1^2 \sin \theta_1 \cos \theta_1 d\theta_1 = m_2^2 \sin \theta_2 \cos \theta_2 d\theta_2.$$

Since $d\phi_1 = d\phi_2$, we have

$$m_1^2 \cos \theta_1 d\Omega_1 = m_2^2 \cos \theta_2 d\Omega_2, \quad (2.13)$$

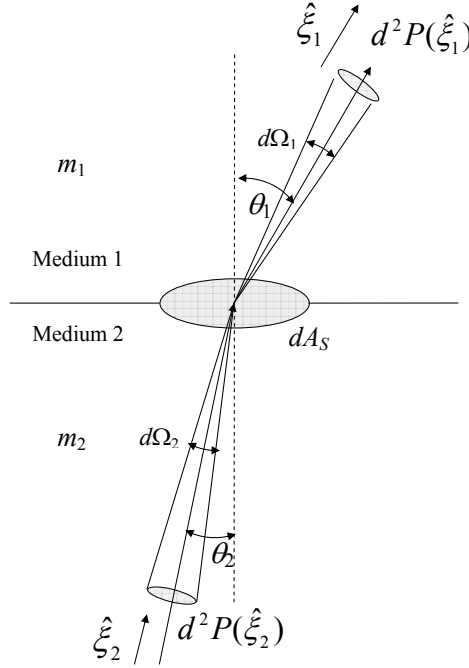


Figure 2.8: Radiance incident from medium 2 (refractive index m_2) is propagated across the interface into medium 1 (refractive index m_1).

which is known as Straubel's invariant, leading to

$$\frac{L(\hat{\xi}_1)}{m_1^2} = t_f \frac{L(\hat{\xi}_2)}{m_2^2}. \quad (2.14)$$

Thus in the absence of losses ($t_f = 1$) the quantity L/m^2 is constant along the path.

2.3.3 Irradiance

Irradiance incident on a surface element dA_S is defined to be the radiant power falling on (or emitted by) dA_S divided by dA_S , i.e.,

$$E = \frac{dP}{dA_S}.$$

For waves propagating in a single direction, $\hat{\xi}$, from Eq. (2.1) this is just the time-average Poynting vector toward the surface (energy flux within $\Delta\lambda$) divided by $\Delta\lambda$, i.e.,

$$E = \frac{\Delta \left(|\langle \vec{S}(\vec{r}) \rangle_{\text{Ave}}| \right)}{\Delta\lambda} |\hat{n}_S \bullet \hat{\xi}|,$$

where \hat{n}_S is the normal to the surface. For a collection of waves incident on the surface from all directions, $E\Delta\lambda$ is the sum of the time-average Poynting vectors of all of the spectral components within $\Delta\lambda$, with each weighted by $|\hat{n}_S \bullet \hat{\xi}|$. The irradiance on dA_S can easily be related to the radiance on dA_S by noting that

$$\frac{d^2 P(\hat{\xi})}{dA_S} = L(\hat{\xi}) \cos \theta d\Omega(\hat{\xi}).$$

Integrating this over all directions $\hat{\xi}$ in a *hemisphere* over dA_S yields

$$E = \frac{dP}{dA_S} = \int_{\text{All } \hat{\xi}} L(\hat{\xi}) \cos \theta d\Omega(\hat{\xi}). \quad (2.15)$$

(When power is being emitted by the surface, E is called the radiant emittance.) In cases where $L(\hat{\xi})$ is independent of $\hat{\xi}$, $E = \pi L$. If a parallel beam of radiance is incident on a Fresnel-reflecting surface, dividing Eq. (2.12) by dA_S tells us that $E_1(\hat{\xi}_1) = t_f E_2(\hat{\xi}_2)$, where $\hat{\xi}_1$ and $\hat{\xi}_2$ are related by Snells law and t_f is evaluated at the appropriate incident angle.

We cannot in general define a Stokes vector for uncollimated irradiance, because the radiation comes from a variety of directions, i.e., from any direction from which the detector surface can be seen. Since the reference vectors \hat{e}_ℓ and \hat{e}_r must satisfy $\hat{\xi} = \hat{e}_r \times \hat{e}_\ell$, and $\hat{\xi}$ is not unique (or not constrained to be within a well-defined small range of directions), it is not possible to define a unique reference system for defining the Stokes parameters. However, if the irradiance is in the form of a well-collimated beam, it is possible to define such a reference system, and in this case, the Stokes vector for irradiance is

$$\mathbf{I}_E(\hat{\xi}) = \begin{pmatrix} I_E(\hat{\xi}) \\ Q_E(\hat{\xi}) \\ U_E(\hat{\xi}) \\ V_E(\hat{\xi}) \end{pmatrix} = \frac{1}{dA_D} \begin{pmatrix} d^2 P_\ell(\hat{\xi}) + d^2 P_r(\hat{\xi}) \\ d^2 P_\ell(\hat{\xi}) - d^2 P_r(\hat{\xi}) \\ 2d^2 P_p(\hat{\xi}) - d^2 P_\ell(\hat{\xi}) - d^2 P_r(\hat{\xi}) \\ 2d^2 P_q(\hat{\xi}) - d^2 P_\ell(\hat{\xi}) - d^2 P_r(\hat{\xi}) \end{pmatrix}. \quad (2.16)$$

2.3.4 Radiance of a near parallel beam

Many times in this work we will require the radiance associated with a beam of nearly parallel light, e.g., the radiance associated with the irradiance from the Sun. In the case

of the Sun, as seen from the Earth the solar disk subtends a angle of $\sim 1/2^\circ$, so the rays of light from the Sun actually diverge at angles up to $1/2^\circ$. If we measure the radiance of the Sun as a function of the angle θ_d from the center of the disk, it will be constant up to $\theta_d = 1/4^\circ$ and then fall to zero and remain there at larger θ_d . (Actually, the radiance within the disk is not constant but decreases somewhat from the center to the edge — *limb darkening*.) Let irradiance from the Sun on a plane normal to the solar beam be F_0 . Then F_0 can be written

$$F_0 = \int L_s(\theta_d) \cos \theta_d d\omega \approx L_s \Delta\Omega_s, \quad (2.17)$$

where $\Delta\Omega_s$ is the solid angle subtended by the Sun at the Earth, and we have used the fact that $\cos \theta_d \approx 1$. Thus we have the relationship between the radiance and the irradiance of the Sun seen from the Earth

$$F_0 = L_s \Delta\Omega_s \quad \text{or} \quad L_s = \frac{F_0}{\Delta\Omega_s}.$$

At times it is useful to assume that the solar beam (or some other light beam) consists of radiation propagating in a single direction, i.e., in the example above $\Delta\Omega_s = 0$. Clearly the associated radiance is infinite, but the irradiance is still finite. A simple mathematical way of dealing with this is through the use of the Dirac delta function. For a beam propagating in the direction $\hat{\xi}_0$, the radiance is given by

$$L(\hat{\xi}) = F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0), \quad (2.18)$$

where F_0 is the irradiance on a plane with normal \hat{n} parallel to $\hat{\xi}_0$ (actually, $\hat{n} = -\hat{\xi}_0$), and $\hat{\xi}$ is a vector in any other direction. From the properties of $\delta^{(2)}(\hat{\xi} - \hat{\xi}_0)$, i.e., it is zero unless $\hat{\xi} = \hat{\xi}_0$, there is no radiance in any direction except $\hat{\xi}_0$, the irradiance on any surface with normal \hat{n}_s is

$$\begin{aligned} E(\hat{n}_s) &= \int |\hat{n}_s \bullet \hat{\xi}| L(\hat{\xi}) d\Omega(\hat{\xi}) \\ &= \int |\hat{n}_s \bullet \hat{\xi}| F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0) d\Omega(\hat{\xi}) \\ &= |\hat{n}_s \bullet \hat{\xi}_0| F_0 \\ &= F_0 \cos \theta_0, \end{aligned}$$

where θ_0 is the angle between the Sun's rays and the surface normal. In geophysical optics, the divergence of the solar beam can often be ignored, and the radiance of the Sun taken to be

$$L_s(\hat{\xi}) = F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0),$$

where F_0 is the extraterrestrial solar irradiance. The same approximation applies to any beam of small angular divergence.

2.3.5 An example: Is the Moon's surface a lambertian reflector?

For a simple example of the application of radiometry to the analysis of a physical problem, we examine the question — is the Moon's surface a lambertian reflector? By a lambertian reflector, we mean that the surface has the property that the reflected radiance is the same for all viewing directions and the reflected irradiance is proportional to the incident irradiance — the proportionality constant being called the *albedo*. A sheet of non-glossy white paper approximates a lambertian reflector.

Consider observing the Moon near full phase (the Earth between the Sun and the Moon, but without the Moon being eclipsed). The irradiance falling on a small area on the lunar surface will be $F_0|\hat{\xi}_0 \bullet \hat{n}|$, where $\hat{\xi}_0$ is the direction of propagation of the solar beam, and \hat{n} is the normal to the lunar surface at for the area in question. If the surface is lambertian, the reflected radiance is the reflected irradiance divided by π . The reflected irradiance is proportional to the incident irradiance, so the radiance of the lunar surface is

$$L_{\text{Moon}} = \frac{AF_0|\hat{\xi}_0 \bullet \hat{n}|}{\pi},$$

where A is the albedo. Thus, were the lunar surface lambertian, the radiance should vary as $|\hat{\xi}_0 \bullet \hat{n}|$, i.e., become darker from the center to the edge. Since this is not observed — the Moon looks more like a disk than a ball — the surface is not lambertian.

2.4 Inherent Optical Properties of Media

In order to study the transfer of radiation through absorbing-scattering-emitting media such as the atmosphere and ocean, it is necessary to define carefully the optically important quantities. These quantities are related to the scattering and absorption properties of the medium. First, we examine scattering. Consider a small volume dV that is illuminated by a beam of collimated irradiance $E_0(\hat{\xi}_0)$ as shown in Figure 2.9. We assume that the incident beam is unpolarized. If $dJ(\hat{\xi})$ is the intensity scattered in the direction $\hat{\xi}$, and measured with a detector *without* polarizers, then the *volume scattering function* $\beta(\hat{\xi}_0 \rightarrow \hat{\xi})$ is defined through

$$\beta(\hat{\xi}_0 \rightarrow \hat{\xi}) \triangleq \frac{dJ(\hat{\xi})}{E_0(\hat{\xi}_0) dV}, \quad (2.19)$$

where $dV = dA dl$. We note that in general the intensity is described by a Stokes vector $d\mathbf{I}_J$ and that when this vector is measured, the first element gives dJ , i.e., $dJ = dI_J$ in Eq. (2.9).

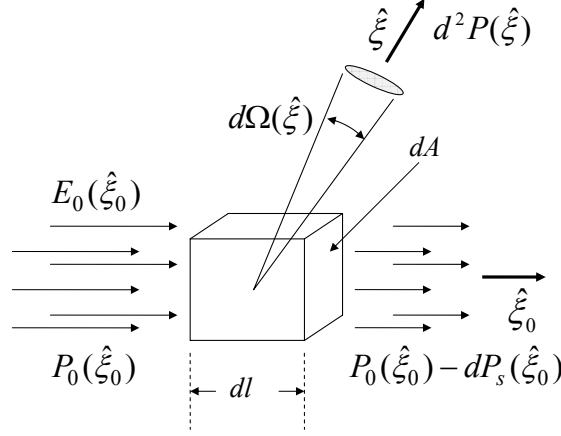


Figure 2.9: Radiant power $P_0(\hat{\xi}_0)$ is incident on a volume $dV = dA dl$. The transmitted power is $P_0(\hat{\xi}_0) - dP_s(\hat{\xi}_0)$, and an amount $d^2P(\hat{\xi})$ is scattered into a solid angle $d\Omega(\hat{\xi})$ around the direction $\hat{\xi}$.

In Chapter 1 we computed the fields scattered from a single atom. How can we use this information to compute β ? Let $dJ_1(\hat{\xi})$ be the intensity scattered by one atom. Then, if there are N such atoms distributed at random, and in random motion, within dV , the scattered radiation from each atom will add incoherently at the sensor and the total scattered intensity will be $dJ = N dJ_1$. Thus,

$$\beta(\hat{\xi}_0 \rightarrow \hat{\xi}) = \frac{N dJ_1(\hat{\xi})}{E_0(\hat{\xi}_0) dV} = \frac{n dJ_1}{E_0},$$

where n is the *number density* of atoms.

More generally, if the incident irradiance is polarized and described by a Stokes vector \mathbf{I}_E , and the scattered intensity is described by a Stokes vector $d\mathbf{I}_J$, then we define the 4×4 *volume scattering matrix* $\beta(\hat{\xi}_0 \rightarrow \hat{\xi})$ through

$$\begin{aligned} d\mathbf{I}_J(\hat{\xi}) &= \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) \mathbf{I}_E(\hat{\xi}_0) dV \\ &= \begin{pmatrix} \beta_{11}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{12}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{13}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{14}(\hat{\xi}_0 \rightarrow \hat{\xi}) \\ \beta_{21}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{22}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{23}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{24}(\hat{\xi}_0 \rightarrow \hat{\xi}) \\ \beta_{31}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{32}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{33}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{34}(\hat{\xi}_0 \rightarrow \hat{\xi}) \\ \beta_{41}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{42}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{43}(\hat{\xi}_0 \rightarrow \hat{\xi}) & \beta_{44}(\hat{\xi}_0 \rightarrow \hat{\xi}) \end{pmatrix} \begin{pmatrix} I_E(\hat{\xi}_0) \\ Q_E(\hat{\xi}_0) \\ U_E(\hat{\xi}_0) \\ V_E(\hat{\xi}_0) \end{pmatrix} dV, \end{aligned} \quad (2.20)$$

where β is a Mueller matrix taking $\mathbf{I}_E dV$ into $d\mathbf{I}_J$. Again, if we compute $d\mathbf{I}_{J1}(\hat{\xi})$ for 1 particle, then we can find β from $n d\mathbf{I}_{J1}(\hat{\xi}) = \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) \mathbf{I}_E(\hat{\xi}_0)$. If the incident irradiance is unpolarized, the Stokes vector of scattered intensity is

$$\begin{pmatrix} dI_J(\hat{\xi}) \\ dQ_J(\hat{\xi}) \\ dU_J(\hat{\xi}) \\ dV_J(\hat{\xi}) \end{pmatrix} = \begin{pmatrix} \beta_{11}(\hat{\xi}_0 \rightarrow \hat{\xi}) E_0(\hat{\xi}_0) \\ \beta_{21}(\hat{\xi}_0 \rightarrow \hat{\xi}) E_0(\hat{\xi}_0) \\ \beta_{31}(\hat{\xi}_0 \rightarrow \hat{\xi}) E_0(\hat{\xi}_0) \\ \beta_{41}(\hat{\xi}_0 \rightarrow \hat{\xi}) E_0(\hat{\xi}_0) \end{pmatrix} dV,$$

and if the scattered intensity is measured without regard for polarization, the result is

$$dI_J(\hat{\xi}) = \beta_{11}(\hat{\xi}_0 \rightarrow \hat{\xi}) E_0(\hat{\xi}_0) dV.$$

Thus, the volume scattering function β in Eq. (2.19) is the upper-left element of the volume scattering matrix in Eq. (2.20), i.e., β_{11} .

If we write the incident irradiance E_0 in Eq. (2.19) as $P_0(\hat{\xi}_0)/dA$, where dA is the area of the volume dV normal to $\hat{\xi}_0$ (so $dV = dA dl$) and $P_0(\hat{\xi}_0)$ is the radiant power incident on dA , then we can rewrite the definition in terms of radiant power:

$$\beta(\hat{\xi}_0 \rightarrow \hat{\xi}) = \frac{d^2 P(\hat{\xi})}{P_0(\hat{\xi}_0) dl d\Omega(\hat{\xi})},$$

or

$$\frac{1}{P_0(\hat{\xi}_0)} \frac{d^2 P(\hat{\xi})}{dl} = \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) d\Omega(\hat{\xi}).$$

Integrating over all directions $\hat{\xi}$, we obtain

$$\frac{1}{P_0(\hat{\xi}_0)} \frac{dP_s}{dl} = \int_{\text{All } \hat{\xi}} \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) d\Omega(\hat{\xi}).$$

On the left-hand-side of this equation dP_s is the total amount of radiant power scattered from dV (in all directions). This power is lost from the beam of irradiance $E_0(\hat{\xi}_0)$ and radiant power $P_0(\hat{\xi}_0)$. The fractional loss in power (or irradiance) per unit length of path from a collimated beam by scattering is called the scattering coefficient:

$$\begin{aligned} b &\triangleq \frac{1}{P_0(\hat{\xi}_0)} \frac{dP_s}{dl} \\ &= \int_{\text{All } \hat{\xi}} \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) d\Omega(\hat{\xi}), \end{aligned} \tag{2.21}$$

so the scattering coefficient is the integral of the volume scattering function over all solid angles. Notice that we have defined the scattering coefficient in terms of the total amount

of scattering from an *unpolarized* beam. It is also possible to define partial scattering coefficients for various incident irradiance polarizations.⁸ This is not carried out here because to our knowledge they are of no utility in geophysical optics at present (see Bohren and Huffman, *Absorption and Scattering of Light by Small Particles* for a further discussion of partial scattering coefficients).

As an example of the use of these definitions, we consider the scattering by a dilute atomic gas. In Chapter 1 it was found that the Stokes vector $\mathbf{I}_s(\Theta)$ scattered from an atom was related to the incident Stokes vector \mathbf{I}_0 through

$$\mathbf{I}_s(\Theta) = \mathbf{M}(\Theta)\mathbf{I}_0, \quad (2.22)$$

where

$$\mathbf{M}(\Theta) = \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0 r)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2}\sin^2 \Theta & 0 & 0 \\ -\frac{1}{2}\sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \quad (2.23)$$

In this equation, Θ is the scattering angle, and the reference vectors \hat{e}_ℓ and \hat{e}_r are taken, respectively, parallel and perpendicular to the scattering plane. It is important to note the difference between Eqs. (2.20) and (2.22). In Eq. (2.22) the elements of the Stokes vectors are simply various combinations of time-averaged Poynting vectors. The matrix \mathbf{M} relates these combinations of products of the scattered field to products of the incident field. It is dimensionless. In contrast, the elements of $d\mathbf{I}_J$ in Eq. (2.20) are time-averaged Poynting vectors times the detector area dA_D divided by the solid angle that the detector subtends at the atom $d\Omega_D$, and also divided by the spectral bandwidth $\Delta\lambda$ (see Eq. (2.9)), while the elements of \mathbf{I}_E are simply various combinations of Poynting vectors divided by $\Delta\lambda$ (see Eq. (2.16)), rendering the units of β , m^{-1} . However, Eq. (2.21) poses a dilemma. Although $d\Omega$ is dimensionless, we need to know its units⁹ in order to evaluate the integral. The standard unit for solid angle is the “steradian” (Sr), for which dA and r^2 in the defining equation $d\Omega \triangleq dA/r^2$ are measured in the same units. Thus, the units of β are taken to be $\text{m}^{-1}\text{Sr}^{-1}$, rather than m^{-1} , to remove the dilemma. The Sr^{-1} is added to *remind* us that standard units are being used for solid angle. To avoid a similar dilemma in the vector case, we associate the units $\text{m}^{-1}\text{Sr}^{-1}$ to β , and as we shall see later in this section, this requires assigning Sr^{-1} as the units of \mathbf{M} — the *dimensionless* Mueller matrix relating the incident to the scattered Stokes vector.

⁸Examples of b for various states of polarization of the incident light are provided in the footnote on page 221.

⁹Many dimensionless quantities have units. The simplest example is a plane angle, which is defined as the *length* of the arc of a circle divided by the *length* of the radius of the circle. It is typically given in units of radians (wherein the two lengths have the *same* units) or in degrees, but could be given in some hybrid unit such as cm/m. Clearly before using the number associated with a (dimensionless) angle, we need to be *reminded* of its units.

After moving the r^2 to the left-hand-side of Eq. (2.23), we have (in the notation of Chapter 1)

$$\begin{pmatrix} r^2 [\langle S_\ell \rangle + \langle S_r \rangle] \\ r^2 [\langle S_\ell \rangle - \langle S_r \rangle] \\ r^2 [2\langle S_p \rangle - \langle S_\ell \rangle - \langle S_r \rangle] \\ r^2 [2\langle S_p \rangle - \langle S_\ell \rangle - \langle S_r \rangle] \end{pmatrix} = \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2} \sin^2 \Theta & 0 & 0 \\ -\frac{1}{2} \sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \\ \times \begin{pmatrix} [\langle S_{0\ell} \rangle + \langle S_{0r} \rangle] \\ [\langle S_{0\ell} \rangle - \langle S_{0r} \rangle] \\ [2\langle S_{0p} \rangle - \langle S_{0\ell} \rangle - \langle S_{0r} \rangle] \\ [2\langle S_{0p} \rangle - \langle S_{0\ell} \rangle - \langle S_{0r} \rangle] \end{pmatrix}$$

Now, since $r^2 = dA_D/d\Omega_D$, when both sides are divided by $\Delta\lambda$, the equation becomes

$$d\mathbf{I}_J = \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2} \sin^2 \Theta & 0 & 0 \\ -\frac{1}{2} \sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \mathbf{I}_E.$$

This describes the scattering from one atom, so

$$\beta(\Theta) = n \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2} \sin^2 \Theta & 0 & 0 \\ -\frac{1}{2} \sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} = nr^2 \mathbf{M}(\Theta), \quad (2.24)$$

where n is the number density of atoms. It also provides the important *general* relationship $\beta(\Theta) = nr^2 \mathbf{M}(\Theta)$. The volume scattering function $\beta(\Theta) = \beta_{11}(\Theta) = nr^2 M_{11}(\Theta)$ is then

$$\beta(\Theta) = n \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right] \frac{1}{2}(1 + \cos^2 \Theta). \quad (2.25)$$

The scattering coefficient is

$$\begin{aligned} b &= \int_{\text{All } \hat{\xi}} \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) d\Omega(\hat{\xi}) \\ &= \int_0^{2\pi} d\Phi \int_0^\pi \beta(\Theta) \sin \Theta d\Theta \\ &= \frac{8\pi}{3} n \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right]. \end{aligned} \quad (2.26)$$

Thus, we see that the scattering coefficient is proportional to the number density of atoms (the scatterers). If we divide b by n , we get a number that is characteristic of the atom:

$\sigma_b = b/n$. The quantity σ_b is called the scattering cross section and has units of area (m^2).¹⁰

In addition to scattering it is possible for irradiance to be lost on passing through dV by absorption of radiant power *within* dV . The absorption coefficient associated with this is

$$a \triangleq \frac{1}{P_0(\hat{\xi}_0)} \frac{dP_a}{dl}, \quad (2.27)$$

where dP_a is the spectral radiant power absorbed from the beam within dV . Recognizing that the *change* in the radiant power in the beam after traversing dl is $-dP_a$, this formula agrees with Eq. (1.27) derived for an electromagnetic wave propagating in an absorbing medium.

Given the absorption and scattering coefficients it is natural to define the total attenuation coefficient c to be the sum $a + b$, i.e., a coefficient characteristic of *all* losses from dV regardless of the cause:

$$c \triangleq \frac{1}{P_0(\hat{\xi}_0)} \left[\frac{dP_s}{dl} + \frac{dP_a}{dl} \right]. \quad (2.28)$$

In the scattering process described above, there was no wavelength change between the incident and the scattered radiation. This is called *elastic* scattering.¹¹ In addition, there are processes similar to scattering, but in which the wavelength of the radiation is changed. Such processes are usually referred to as *inelastic scattering*. Inelastic scattering involves the absorption of a photon at one wavelength (called the exciting wavelength λ_E) and the subsequent emission of a photon at a longer wavelength λ_F , with a negligible time delay ($\sim 10^{-9}$ sec) between the absorption and the emission. The radiometry of the inelastic scattering process can be formulated in a manner analogous to that for elastic scattering, after recognizing that in general radiant power in a *band* wavelengths less than λ_F can inelastically scatter into λ_F . Consider a small volume dV , illuminated by an unpolarized

¹⁰Why is σ_b called a “cross section” with units of area? Consider a volume of atoms in the gaseous phase in a chamber with a macroscopic area A and a small thickness $\Delta\ell$. Associate a sphere of radius R with each atom such that if a photon passes within a distance R from the atom it will be scattered. Now illuminate the area A with a beam of N photons that uniformly fill, and are propagating normal to, the area A . Let the number density n of atoms be so small that no two of the volumes associated with atoms overlap. The number of atoms in the beam is $nA\Delta\ell$. But, from the point of view of the beam, the atoms are just a collection of $nA\Delta\ell$ *areas* each of size $\sigma_b = \pi R^2$. The total cross sectional area associated with the atoms in the beam is $\sigma_b nA\Delta\ell$. The fraction of photons $\Delta N/N$ that is scattered is the total cross sectional area associated with the atoms divided by the total area of the beam A , i.e., $\sigma_b n\Delta\ell$. But, from the definition of the scattering coefficient, $\Delta N/N = b\Delta\ell$, so $b = n\sigma_b$. Thus, σ_b is the cross sectional area of the imaginary volume centered on each atom, with the property that if a photon passes within the area it will be scattered.

¹¹This terminology (elastic) stems from elementary mechanics where collisions between two objects in which the energy is conserved are referred to as elastic collisions.

parallel beam propagating in a direction $\hat{\xi}_0$ with an irradiance $E(\hat{\xi}_0, \lambda_E)$ in a band of wavelengths between λ_E and $\lambda_E + \Delta\lambda_E$. If an intensity of inelastically scattered radiation $dJ(\hat{\xi}, \lambda_F)$ is observed at a wavelength λ_F propagating in a direction $\hat{\xi}$, the *inelastic* volume scattering function $\beta_{\text{In}}(\hat{\xi}_0 \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda_F)$ is defined by,

$$\beta_{\text{In}}(\hat{\xi}_0 \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda_F) \triangleq \frac{dJ(\hat{\xi}, \lambda_F)}{dV \int_{\Delta\lambda_E} E(\hat{\xi}_0, \lambda_E) d\lambda_E}. \quad (2.29)$$

The units of β_{In} are *not* the same as the units of β because of the integration over the excitation wavelengths $\Delta\lambda_E$. The latter integration provides the *total* radiant power incident on the sample from wavelengths within $\Delta\lambda_E$. (Note that the measurement of $dJ(\hat{\xi}, \lambda_F)$ is made with a detector equipped with a filter of bandwidth $\Delta\lambda_F$.) The inelastic volume scattering function is sometimes referred to as the *trans-spectral* scattering function, although this is non-standard terminology. In a manner similar to the scattering coefficient, one can also define the inelastic scattering coefficient $b_{\text{In}}(\lambda_E \rightarrow \lambda_F)$ through

$$b_{\text{In}}(\lambda_E \rightarrow \lambda_F) \triangleq \int_{\text{All } \hat{\xi}} \beta_{\text{In}}(\hat{\xi}_0 \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda_F) d\Omega(\hat{\xi}). \quad (2.30)$$

Inelastic scattering can be formulated for polarized incident irradiance in a manner similar to *elastic* scattering above, i.e., scattering without any change in wavelength as discussed earlier in this section, simply by replacing $E(\hat{\xi}_0, \lambda_E)$, $dJ(\hat{\xi}, \lambda_F)$ and $\beta_{\text{In}}(\hat{\xi}_0 \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda_F)$, by their Stokes vector and Mueller matrix counterparts.

From the definitions of a , β , and β_{In} , it is clear that if the absorption, elastic, and inelastic scattering functions result from N separate components (species) within dV , that if acting independently the components would have coefficients a_i , etc., we can write

$$\begin{aligned} a &= \sum_{n=1}^N a_i \\ \beta &= \sum_{n=1}^N \beta_i \\ (\beta_{\text{In}}) &= \sum_{n=1}^N (\beta_{\text{In}})_i \end{aligned}$$

The inherent optical properties are *summable* (additive) over those of the individual components.

We now have defined all of the inherent optical properties (IOPs) needed to develop an equation governing the transport of radiance in an absorbing-scattering medium — *the radiative transfer equation*.

2.5 The Radiative Transfer Equation

In this section we provide a derivation of the radiative transfer equation and examine some of its properties. In later sections we provide some methods of solution.

2.5.1 A Phenomenological Derivation of the Radiative Transfer Equation

The propagation of radiance in an absorbing-scattering medium is governed by the radiative transfer equation (RTE). We know that radiation traversing a scattering medium for any significant distance will be incident on a given point in the medium from all directions. We will “derive” the RTE by keeping track of how the radiance in a given direction changes from one position to another.¹² The change in radiance propagating in a direction $\hat{\xi}$ between two positions \vec{r}_1 and \vec{r}_2 within a medium is determined by placing a radiometer at \vec{r}_1 to measure $L(\vec{r}_1, \hat{\xi})$ and then moving it to \vec{r}_2 to measure $L(\vec{r}_2, \hat{\xi})$. This is shown schematically in Figure 2.10, where in position 1 the radiometer, with a field of view restricted to a solid angle $d\omega$ by a Gershun tube, measures $L_1(\hat{\xi})$. When the radiometer is moved a distance dl in the direction $\hat{\xi}$, it records $L_2(\hat{\xi})$. If P_1 and P_2 are the radiant powers measured by the detector at positions 1 and 2, respectively, some radiant power will have been removed between positions 1 and 2 because absorption within the volume $dA_s dl$ and because of scattering *out* of $dA_s dl$. Thus,

$$P_2 = P_1 - dP = P_1(1 - c dl),$$

or dividing by $d\omega dA_d$ to form the radiances,

$$L_2(\hat{\xi}) = L_1(\hat{\xi})(1 - c dl) \quad \text{or} \quad \frac{L_2(\hat{\xi}) - L_1(\hat{\xi})}{dl} = -cL_1(\hat{\xi}),$$

which shows that

$$\frac{dL(\hat{\xi})}{dl} = -cL(\hat{\xi}).$$

Thus, absorption and scattering cause the radiance to *decrease* as the radiometer is moved in the direction of propagation. However, there may be sources of radiant power within

¹²We neglect changes in radiance with time as they occur on vastly different time scales from those of interest, e.g. time scales of the order of the time it takes light to propagate from the top of the atmosphere to the water surface ($\sim 100 \mu s$). Most processes encountered in environmental optics can be considered steady state processes in radiative transfer. (An exception is LIDAR which will be discussed in a later chapter.)

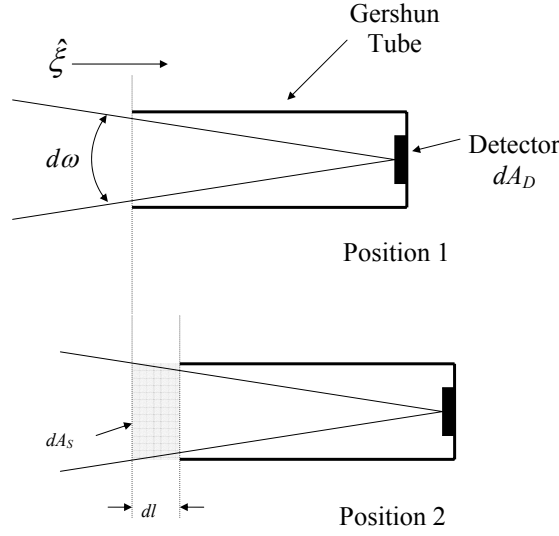


Figure 2.10: A radiometer in position 1 is measuring the radiance $L_1(\hat{\xi})$ propagating in the $\hat{\xi}$ direction. It is then moved a distance dl in the direction $\hat{\xi}$ to position 2 and measures a radiance $L_2(\hat{\xi})$ that has been modified from $L_1(\hat{\xi})$ by the presence of the additional volume of medium $dV = dA_S dl$. (Note that the area of the open end of the Gershun tube is dA_S)

the volume $dV = dA_S dl$, e.g., scattering of radiant power incident on the volume from some other direction, $\hat{\xi}'$, into the direction $\hat{\xi}$. How do we include such sources? The simplest way is to consider the case of a single point source within dV as shown in Figure 2.11. This source causes the detector to record a radiant power $J_0(\hat{\xi})d\Omega_D$, where $J_0(\hat{\xi})$ is the intensity of the source in the direction $\hat{\xi}$. If there are N such sources within dV , the total power will be $NJ_0(\hat{\xi})d\Omega_D$, which leads to a radiance

$$dL(\hat{\xi}) = \frac{NJ_0(\hat{\xi})d\Omega_D}{d\omega dA_D} = \frac{NJ_0(\hat{\xi})}{dA_S},$$

where in the last equality we used $d\omega dA_D = d\Omega_D dA_S$. Now, $NJ_0(\hat{\xi})$ is just the total intensity $dJ(\hat{\xi})$ generated within the volume $dA_S dl = dV$, so

$$dL(\hat{\xi}) = \frac{dJ(\hat{\xi})}{dV} dl, \quad \text{or} \quad \frac{dL(\hat{\xi})}{dl} = \frac{dJ(\hat{\xi})}{dV}.$$

Combining this gain by sources and the loss by absorption and scattering,

$$\frac{dL(\hat{\xi})}{dl} = -cL(\hat{\xi}) + \frac{dJ(\hat{\xi})}{dV}. \quad (2.31)$$

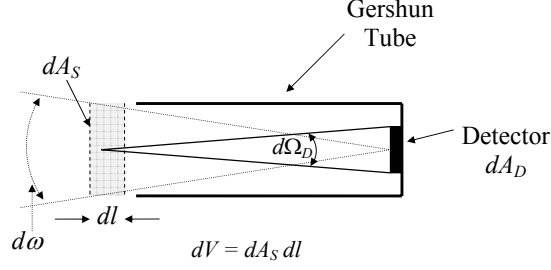


Figure 2.11: A radiometer in position 1 is measuring the radiance $L_1(\hat{\xi})$ propagating in the $\hat{\xi}$ direction. It is then moved a distance dl in the direction $\hat{\xi}$ to position 2 and measures a radiance $L_2(\hat{\xi})$ that has been modified from $L_1(\hat{\xi})$ by the presence of the additional volume of medium $dV = dA_S dl$. (Note that the area of the open end of the Gershun tube is dA_S .)

The quantity $dJ(\hat{\xi})/dV$ is the intensity density of sources within the medium.

When the intensity density of sources within dV is caused by scattering from some other direction $\hat{\xi}'$, Eq. (2.19) shows that it is related to the volume scattering function by

$$\frac{dJ(\hat{\xi})}{dV} = \beta(\hat{\xi}' \rightarrow \hat{\xi}) E(\hat{\xi}'),$$

where $E(\hat{\xi}')$ is the irradiance incident on dV from the direction $\hat{\xi}'$ (with E measured on a plane normal to $\hat{\xi}'$). In general, irradiance will be incident from many directions (many different $\hat{\xi}'$ s), and this can be accommodated by noting that the irradiance on a plane normal to $\hat{\xi}'$ resulting from radiance $L(\hat{\xi}')$, contained in a solid angle $d\Omega(\hat{\xi}')$ around $\hat{\xi}'$, is $L(\hat{\xi}') d\Omega(\hat{\xi}')$ — recall the radiance of a near-parallel beam. Because $\hat{\xi}'$ can be in any direction,

$$\frac{dJ(\hat{\xi})}{dV} = \int_{\text{All } \hat{\xi}'} \beta(\hat{\xi}' \rightarrow \hat{\xi}) L(\hat{\xi}') d\Omega(\hat{\xi}').$$

Thus, in an absorbing-scattering medium, the radiance in the direction $\hat{\xi}$ changes with distance l (measured along $\hat{\xi}$) according to

$$\frac{dL(\hat{\xi})}{dl} = -cL(\hat{\xi}) + \int_{\text{All } \hat{\xi}'} \beta(\hat{\xi}' \rightarrow \hat{\xi}) L(\hat{\xi}') d\Omega(\hat{\xi}').$$

Recalling from the Mathematical Appendix that the directional derivative of a scalar function $f(\vec{r})$ of position along the direction $\hat{\xi}$ is

$$\frac{df(\vec{r})}{dl} = \hat{\xi} \bullet \nabla f(\vec{r}),$$

and noting that $L(\hat{\xi})$ is a function of position in the medium, we have (restoring all of the functional dependencies, including wavelength)

$$\hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) = -c(\vec{r}, \lambda) L(\vec{r}, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(\vec{r}, \hat{\xi}', \lambda) d\Omega(\hat{\xi}'). \quad (2.32)$$

If the medium can also undergo *inelastic* scattering from λ_E to λ , with $\lambda_E < \lambda$, there is an additional contribution to the intensity density given by Eq. (2.29) resulting in

$$\begin{aligned} \hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) &= -c(\vec{r}, \lambda) L(\vec{r}, \hat{\xi}, \lambda) \\ &+ \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(\vec{r}, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') \\ &+ \int_{\lambda_E < \lambda} \int_{\text{All } \hat{\xi}'} \beta_{\text{In}}(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda) L(\vec{r}, \hat{\xi}', \lambda_E) d\Omega(\hat{\xi}') d\lambda_E. \end{aligned} \quad (2.33)$$

Finally, if the measurements performed in the derivation of Eq. (2.32) are carried out with a radiometer equipped with polarizers to measure the Stokes vector associated with L , i.e., \mathbf{I}_L , the same sequence of steps clearly leads to¹³

$$\hat{\xi} \bullet \nabla \mathbf{I}_L(\vec{r}, \hat{\xi}, \lambda) = -c(\vec{r}, \lambda) \mathbf{I}_L(\vec{r}, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) \mathbf{I}_L(\vec{r}, \hat{\xi}', \lambda) d\Omega(\hat{\xi}'), \quad (2.34)$$

where $c(\vec{r}, \lambda)$ is the 4×4 extinction Mueller matrix (see Chapter 3, Section 3.6.1, page 269). Note that in Eq. (2.34) the notation on the left-hand-side means

$$\hat{\xi} \bullet \nabla \mathbf{I}_L(\vec{r}, \hat{\xi}, \lambda) \triangleq \begin{pmatrix} \hat{\xi} \bullet \nabla I_L(\vec{r}, \hat{\xi}, \lambda) \\ \hat{\xi} \bullet \nabla Q_L(\vec{r}, \hat{\xi}, \lambda) \\ \hat{\xi} \bullet \nabla U_L(\vec{r}, \hat{\xi}, \lambda) \\ \hat{\xi} \bullet \nabla V_L(\vec{r}, \hat{\xi}, \lambda) \end{pmatrix}.$$

Why not derive the RTE directly from Maxwell's equations, i.e., let a plane electromagnetic wave fall on a collection of absorbing and scattering atoms, molecules, or physical particles, and directly compute the Poynting vector that would be measured by a radiometer? Since each scatterer is subjected to the scattered fields of *all* of the other scatterers as well as to the incident field, this would be an incredibly difficult task; however, it actually has been accomplished by M. Mishchenko and co-workers under certain restrictions (see

¹³Note that the transfer equation written in this fashion is not really as simple as it appears because β is referenced to directions relative to the plane of scattering, which change with the integration variable $\hat{\xi}'$, while \mathbf{I}_L is usually referenced to a set of directions relative to the plane containing the propagation direction $\hat{\xi}$ and some *fixed* direction. This state of affairs is remedied in Appendix 1 (Section 2.11.3 Eq. (2.113)) for a plane parallel medium by referencing both β and \mathbf{I}_L to the *same* plane. The more useful form of the radiative transfer equation, including polarization, is then given in Eq. (2.113).

Bibliographic Notes). The final result is identical to the RTE as derived above. Thus, even though the development here of the radiative transfer equation is a significant departure from the methods of electromagnetic theory, the pedagogical advantage of the phenomenological derivation is undeniable.¹⁴

2.5.2 Uniqueness of RTE Solutions

If we write the RTE in the form

$$\hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) = -c(\vec{r}, \lambda)L(\vec{r}, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda)L(\vec{r}, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') + Q(\vec{r}, \hat{\xi}, \lambda)$$

where Q is the intensity density of sources in the medium, e.g., inelastic scattering or actual light sources within the medium, then the following theorem can be proved:

Consider a volume V bounded by a surface S . Let \vec{r} represent the position of any point in the medium, $\vec{\rho}$ the position of any point on the boundary, and \hat{n}_s the *outward* normal to S . Then, if $a(\vec{r}, \lambda) > 0$ throughout V , the radiance $L(\vec{r}, \hat{\xi}, \lambda)$ is uniquely determined by the radiance $L(\vec{\rho}, \hat{\xi}, \lambda)$ incident on the boundary from the outside (i.e., $\hat{\xi} \bullet \hat{n}_s < 0$) and the sources $Q(\vec{r}, \hat{\xi}, \lambda)$ within V . If $a(\vec{r}) = 0$ everywhere in the medium, the radiance exiting V (i.e., $L(\vec{\rho}, \hat{\xi}, \lambda)$ with $\hat{\xi} \bullet \hat{n}_s > 0$) is still unique, but the uniqueness the interior radiance has not been proved.

As all geophysical media will have some absorption at each wavelength, we can assume that the interior and exiting radiance for a volume are uniquely determined by the radiance incident on the volume and the sources within the volume.

2.5.3 Superposition of Solutions

Noting that the radiative transfer equation is linear in the radiance, it is clear that

given a volume V in which the inherent optical properties c and β are specified, if L_1 satisfies the RTE in V with source Q_1 subject to a radiance L_{1B} incident

¹⁴In Chapter 3 we will study the scattering of electromagnetic radiation from single particles. As a prelude to the phenomenological derivation of the RTE, we operationally defined $\beta(\hat{\xi}' \rightarrow \hat{\xi})$ as the scattered intensity density from a small volume divided by the irradiance of the incident (collimated) beam. In order that $\beta(\hat{\xi}' \rightarrow \hat{\xi})$ be derivable from electromagnetic theory, we must require that the volume be sufficiently small and the density of scatterers be sufficiently low, that each scatterer is subjected to *only* the external (incident) field, i.e., there is negligible multiple scattering within the volume. In addition, we must assume that all of the scatterers within the volume scatter independently (incoherently).

on the boundary from the outside, and L_2 satisfies the RTE in V with source Q_2 subject to a radiance L_{2B} incident on the boundary from the outside, then, $L_1 + L_2$ will satisfy the RTE in V with source $Q_1 + Q_2$ subject to $L_{1B} + L_{2B}$ incident on the boundary from the outside.

This allows one, for example, to find the radiance in a volume in the presence of both radiance incident from the outside and an internal source by first solving the problem in the absence of the internal source, and then adding to this the solution of the problem in the presence of the internal source with *no* radiance incident from the outside. This will be useful in the discussion of inelastic processes.

2.5.4 Continuously Varying Refractive Index

Equation (2.14) shows that across an interface between two lossless dielectric media of different refractive indices, L/m^2 decreases along the direction of propagation only through Fresnel reflection at the interface. This reflection results from a *discontinuous* change in the refractive index. If the refractive index is a *continuous* function of position, then $t_F = 1$ and L/m^2 is invariant along the propagation path, i.e., along a ray. Applying this, we find that in a medium in which m is a continuous function of position, Eq. (2.31) becomes

$$\frac{d}{dl} \left[\frac{L(\hat{\xi})}{m^2} \right] = -c \left[\frac{L(\hat{\xi})}{m^2} \right] + \frac{1}{m^2} \frac{dJ(\hat{\xi})}{dV},$$

where $\hat{\xi}$ is a unit vector tangent to the (generally curved) propagation path. When this is taken into account, the radiative transfer equations for a medium in which m is a continuous function of position are those given in Eq. (2.32), (2.33), and (2.34), with all of the radiances replaced by radiance divided by $m^2(\vec{r})$.

2.5.5 The Reciprocity Principle

Consider two different radiative transfer problems on the same medium. Let the medium have a volume V bounded by a surface S that has a normal (outward) \hat{n} . Let any point on the boundary of the medium be indicated by the position vector $\vec{\rho}$. In Problem 1 the radiance incident on the boundary from the outside is $L_1(\vec{\rho}, \hat{\xi}, \lambda)$ and the internal sources are specified by $Q_1(\vec{r}, \hat{\xi}, \lambda)$. Likewise, in Problem 2 the radiance incident on the boundary from the outside is $L_2(\vec{\rho}, \hat{\xi}, \lambda)$ and the internal sources are specified by $Q_2(\vec{r}, \hat{\xi}, \lambda)$. The inherent optical properties of the medium are identical in both problems. Under the

assumption that the volume scattering function is time reversal invariant, i.e.,

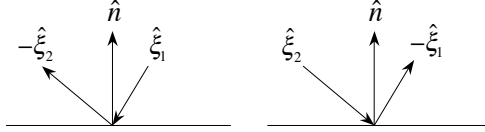
$$\beta(\vec{r}, -\hat{\xi}' \rightarrow -\hat{\xi}, \lambda) = \beta(\vec{r}, \hat{\xi} \rightarrow \hat{\xi}', \lambda),$$

it is possible to derive the *reciprocity principle* that relates the solutions $L_1(\vec{r}, \hat{\xi}, \lambda)$ and $L_2(\vec{r}, \hat{\xi}, \lambda)$ of these two problems to one another:

$$\begin{aligned} & \int_S dS \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| \left[\frac{L_1(\vec{\rho}, \hat{\xi}, \lambda) L_2(\vec{\rho}, -\hat{\xi}, \lambda)}{m^2(\vec{\rho})} - \frac{L_1(\vec{\rho}, -\hat{\xi}, \lambda) L_2(\vec{\rho}, \hat{\xi}, \lambda)}{m^2(\vec{\rho})} \right] d\Omega(\hat{\xi}) \\ &= \int_{4\pi} d\Omega(\hat{\xi}) \int_V \left[\frac{Q_2(\vec{r}, -\hat{\xi}, \lambda) L_1(\vec{r}, \hat{\xi}, \lambda)}{m^2(\vec{r})} - \frac{Q_1(\vec{r}, \hat{\xi}, \lambda) L_2(\vec{r}, -\hat{\xi}, \lambda)}{m^2(\vec{r})} \right] dV. \end{aligned} \quad (2.35)$$

In this equation, $L_1(\vec{\rho}, -\hat{\xi}, \lambda)$ and $L_2(\vec{\rho}, -\hat{\xi}, \lambda)$ are the radiances *exiting* the medium at the boundaries, and the $\hat{\xi} \bullet \hat{n} < 0$ on the solid angle integral on the left-hand-side indicates that the integration is restricted to directions *into* the medium.¹⁵ The meaning of the

¹⁵As a simple and useful example of the application of the reciprocity principle, consider two radiative transfer problems on a plane parallel medium bounded on the top. Radiance is incident from the outside on the boundary and there are no sources (Q 's) inside the medium. In the first problem, $L_1(\hat{\xi}) = F_0 \delta(\hat{\xi} - \hat{\xi}_1)$, while in the second $L_2(\hat{\xi}) = F_0 \delta(\hat{\xi} - \hat{\xi}_2)$. Thus, in problem one a parallel beam of radiance is incident in the direction $\hat{\xi}_1$, while in the second



the beam is incident in the direction $\hat{\xi}_2$. Applying the reciprocity principle, Eq. (2.35),

$$\begin{aligned} 0 &= \int_S dS \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| \left[F_0 \delta(\hat{\xi} - \hat{\xi}_1) L_2(-\hat{\xi}) - L_1(-\hat{\xi}) F_0 \delta(\hat{\xi} - \hat{\xi}_2) \right] d\Omega(\hat{\xi}) \\ &= \int_S dS F_0 \left[|\hat{\xi}_1 \bullet \hat{n}| L_2(-\hat{\xi}_1) - |\hat{\xi}_2 \bullet \hat{n}| L_1(-\hat{\xi}_2) \right] \implies \left[|\hat{\xi}_1 \bullet \hat{n}| L_2(-\hat{\xi}_1) - |\hat{\xi}_2 \bullet \hat{n}| L_1(-\hat{\xi}_2) \right] = 0, \end{aligned} \quad (2.36)$$

$$\text{or} \quad \frac{L_1(-\hat{\xi}_2)}{|\hat{\xi}_1 \bullet \hat{n}|} = \frac{L_2(-\hat{\xi}_1)}{|\hat{\xi}_2 \bullet \hat{n}|}.$$

If we use the notation $L(\hat{\xi}_a \rightarrow \hat{\xi}_b)$ to indicate the radiance propagating in the direction $\hat{\xi}_b$ when the source is a beam propagating in the direction $\hat{\xi}_a$, then in this more explicit notation:

$$\frac{L(\hat{\xi}_1 \rightarrow -\hat{\xi}_2)}{|\hat{\xi}_1 \bullet \hat{n}|} = \frac{L(\hat{\xi}_2 \rightarrow -\hat{\xi}_1)}{|\hat{\xi}_2 \bullet \hat{n}|}.$$

assumption of time reversal invariance of the volume scattering function is that if we viewed the scattering process backward in time, the same volume scattering function would apply. We shall not provide a proof of the reciprocity principle here; however, proofs can be found in references provided in the Bibliographic Notes. This relationship will be useful by relating difficult-to-solve radiative transfer problems to those that are easier to solve.

2.5.6 Plane-Parallel Media and One-Dimensional Radiative Transfer

In nearly all of geophysical optics we can neglect the curvature of the earth, the oceans, and the atmosphere.¹⁶ Thus, for example, we can treat the ocean and the atmosphere as *plane-parallel media*, by which we mean surfaces of constant pressure in the atmosphere and ocean become flat surfaces of constant altitude and depth, respectively. In addition, because the variation in the inherent optical properties of both media with distance are much more rapid in the vertical than in the horizontal, we assume that these properties vary *only* in the vertical, i.e., the inherent optical properties are uniform in the horizontal. We set up the coordinate axes with z directed *into* the medium from the boundary and the $x - y$ plane on the surface. The corresponding unit vectors are \hat{e}_z , \hat{e}_x and \hat{e}_y . Then we have

$$c(\vec{r}, \lambda) = c(z, \lambda) \quad \text{and} \quad \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) = \beta(z, \hat{\xi}' \rightarrow \hat{\xi}, \lambda).$$

In addition, if the boundary radiances (top and bottom) are independent of position in the horizontal and internal sources are dependent only on the vertical coordinate, the interior and exiting radiance cannot depend on horizontal position and

$$L(\vec{r}, \hat{\xi}, \lambda) = L(z, \hat{\xi}, \lambda) \quad \text{and} \quad \hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) = \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial z} L(z, \hat{\xi}, \lambda).$$

Thus, the RTE in this setting is

$$\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial z} L(z, \hat{\xi}, \lambda) = -c(z, \lambda) L(z, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(z, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(z, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') + Q(z, \hat{\xi}, \lambda), \quad (2.37)$$

and the incident boundary radiances are specified at two values of z , the top and the bottom of the medium. Equation (2.37) is called the *one-dimensional* (or 1-d) radiative transfer equation.

In many applications that we will consider, the 1-d medium will be illuminated by a beam, e.g., the solar beam. In this case the radiance due to the beam is often treated separately in the following manner. The radiance incident on the top of the medium is

¹⁶Exceptions to this occur when the Sun is near the horizon, when a satellite sensor views near the limb of the Earth, or both simultaneously.

given by Eq. (2.18), i.e., $L(0, \hat{\xi}, \lambda) = F_0(\lambda) \delta^{(2)}(\hat{\xi} - \hat{\xi}_0)$, where F_0 is the incident irradiance on a plane normal to $\hat{\xi}_0$. The beam itself will propagate through the medium according to

$$L_s(z, \hat{\xi}, \lambda) = L(0, \hat{\xi}, \lambda) \exp\left[-\int_0^l c(l', \lambda) dl'\right],$$

where l is the distance along the path ($\hat{\xi}_0$) from the surface to a depth z : $l = z/|\hat{\xi}_0 \bullet \hat{e}_z|$. (Note that the solar beam is assumed here to have zero solid angle, so no radiance is ever scattered *into* the beam.) Thus, at all points in the medium, the solar beam produces a radiance

$$L_s(z, \hat{\xi}, \lambda) = F_0(\lambda) \delta^{(2)}(\hat{\xi} - \hat{\xi}_0) \exp\left[-\int_0^z \frac{c(z', \lambda) dz'}{|\hat{\xi}_0 \bullet \hat{e}_z|}\right].$$

The radiance generated by this beam through scattering is just

$$\int_{\text{All } \hat{\xi}'} \beta(z, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L_s(z, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') = F_0(\lambda) \beta(z, \hat{\xi}_0 \rightarrow \hat{\xi}, \lambda) \exp\left[-\int_0^z \frac{c(z', \lambda) dz'}{|\hat{\xi}_0 \bullet \hat{e}_z|}\right].$$

Including this as a source, the RTE becomes

$$\begin{aligned} \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial z} L(z, \hat{\xi}, \lambda) &= -c(z, \lambda) L(z, \hat{\xi}, \lambda) \\ &+ \int_{\text{All } \hat{\xi}'} \beta(z, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(z, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') \\ &+ F_0(\lambda) \beta(z, \hat{\xi}_0 \rightarrow \hat{\xi}, \lambda) \exp\left[-\int_0^z \frac{c(z', \lambda) dz'}{|\hat{\xi}_0 \bullet \hat{e}_z|}\right] \\ &+ Q(z, \hat{\xi}, \lambda), \end{aligned} \tag{2.38}$$

where, now at the upper boundary we have $L(0, \hat{\xi}, \lambda) = 0$ because the solar beam has been explicitly included as an additional source of radiance *within* the medium, i.e., as a Q . Equations (2.37) and (2.38) are completely equivalent. The only virtue of Eq. (2.38) over Eq. (2.37) is that at times it is computationally preferable to have the simpler boundary condition; $L(0, \hat{\xi}, \lambda) = 0$.

2.5.7 Irradiances in One-Dimensional Media

In plane-parallel media such as the ocean and the atmosphere, the irradiances are useful, particularly when referenced to a horizontal surface. The *downwelling* irradiance $E_d(z, \lambda)$ and the *upwelling* irradiance $E_u(z, \lambda)$ are given by

$$E_d(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z > 0} |\hat{\xi} \bullet \hat{e}_z| L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \tag{2.39}$$

and

$$E_u(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z < 0} |\hat{\xi} \bullet \hat{e}_z| L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (2.40)$$

respectively.¹⁷ These can be combined to form the *vector* irradiance $E_V(z, \lambda)$:

$$E_V(z, \lambda) \triangleq E_d(z, \lambda) - E_u(z, \lambda) = \int_{\text{All } \hat{\xi}} \hat{\xi} \bullet \hat{e}_z L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}). \quad (2.41)$$

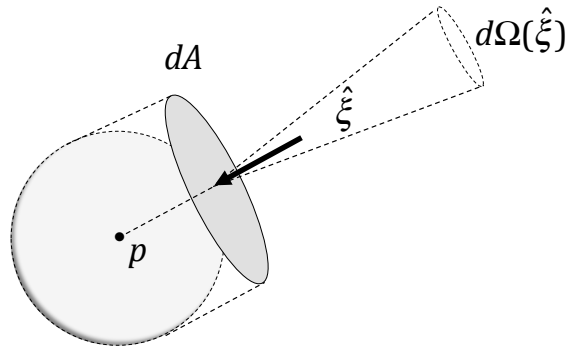
In addition, the downwelling (E_{0d}) and upwelling (E_{0u}) *scalar* irradiances are defined as above but without the $\hat{\xi} \bullet \hat{e}_z$ factor:

$$E_{0d}(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z > 0} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (2.42)$$

$$E_{0u}(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z < 0} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (2.43)$$

$$E_0(z, \lambda) \triangleq E_{0d}(z, \lambda) + E_{0u}(z, \lambda) = \int_{\text{All } \hat{\xi}} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}). \quad (2.44)$$

The sum E_0 is referred to as the *scalar* irradiance. The plane irradiances E_d and E_u are the energy fluxes (in the interval $\Delta\lambda$ and divided by $\Delta\lambda$) across a horizontal surface at z propagating in the downward and upward directions, respectively. The vector irradiance is the *net* downward energy flux. How do we interpret the scalar irradiance? Consider the figure below showing radiance from the direction $\hat{\xi}$, within the solid angle $d\Omega(\hat{\xi})$, incident



¹⁷Sometimes these are referred to as *plane* irradiances to contrast them with the *vector* irradiance or the *scalar* irradiance defined next.

on a detector of area dA . The radiant power incident on dA from directions within $d\Omega(\hat{\xi})$ is

$$d^3\mathcal{P}(\hat{\xi}, \lambda) = L(\hat{\xi}, \lambda) dA d\Omega(\hat{\xi}) d\lambda = d^2 |\langle \vec{S}(\hat{\xi}, \lambda) \rangle_{\text{Ave}}| dA,$$

where $\vec{S}(\hat{\xi}, \lambda)$ is the sum of the Poynting vectors from all of the (independent) sources of radiant power within the field of view of the radiometer, and within the spectral interval $d\lambda$. Thus,

$$L(\hat{\xi}, \lambda) d\Omega(\hat{\xi}) = \frac{d}{d\lambda} \left[d |\langle \vec{S}(\hat{\xi}, \lambda) \rangle_{\text{Ave}}| \right] = \frac{d}{d\lambda} \left[c_\ell \langle du_{em}(\hat{\xi}, \lambda) \rangle \right],$$

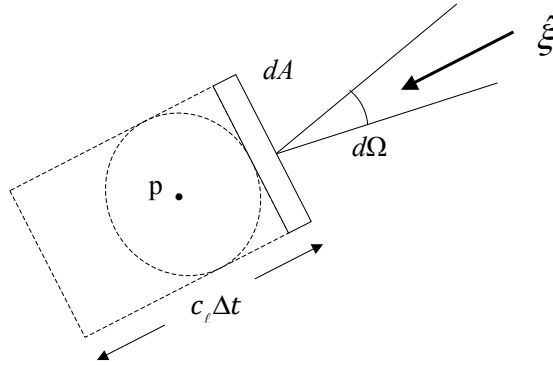
where c_ℓ is the speed of light¹⁸ and $du_{em}(\hat{\xi}, \lambda)$ is the electromagnetic energy density in the region around the point “p” due to radiation with propagation directions within $d\Omega(\hat{\xi})$ and wavelengths within $d\lambda$. Now, if we integrate over $d\Omega$, i.e., sweep the radiometer over the surface of the sphere centered on the point “p,” so as to cover the whole imaginary sphere around “p,” and sum the radiance-solid angle product for each position, we have

$$\int_{\text{All } \hat{\xi}} L(\hat{\xi}, \lambda) d\Omega(\hat{\xi}) = E_0(p, \lambda) = \frac{d}{d\lambda} \left[\langle u_{em}(\hat{\xi}, \lambda) \rangle \right],$$

where $\langle u_{em} \rangle$ is the time-averaged electromagnetic energy density at “p” from radiation within $d\lambda$ centered on λ , and $d\langle u_{em} \rangle / d\lambda$ is the time-averaged *spectral* energy density at “p.” Therefore, the scalar irradiance at a point is the speed of light times the time-averaged spectral energy density at that point.¹⁹ Relabeling the spectral energy density

¹⁸We will use the notation c_ℓ for the speed of light in situations where it might be confused with the beam attenuation coefficient c .

¹⁹One can also reach the same conclusion using the photon concept. In the figure below, radiance from the direction $\hat{\xi}$ in $d\Omega$ falls on a detector of area dA .

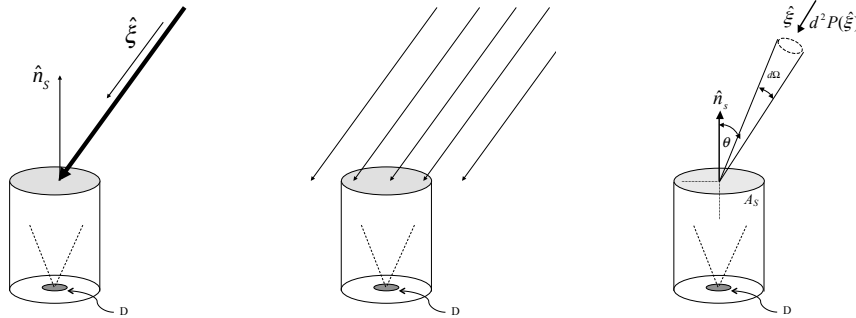


$d\langle u_{em} \rangle / d\lambda \triangleq \langle u_{em}^s \rangle$, we obtain the final result

$$\int_{\text{All } \hat{\xi}} L(\hat{\xi}, \lambda) d\Omega(\hat{\xi}) = E_0(p, \lambda) = c_\ell \langle u_{em}^s(p, \lambda) \rangle.$$

2.5.8 Measurement of Irradiances

Equations (6.1) – (6.6) show how the various irradiances can be determined from the radiance distribution $L(z, \hat{\xi}, \lambda)$; however, in practice the radiance distribution is very rarely measured for the full range of directions $\hat{\xi}$. Is there a way of measuring the various irradiances without requiring the full radiance distribution? Most measurements of plane irradiance are made with relatively simple devices called *irradiance meters*. Here's how they work. We want to measure the spectral irradiance falling on the surface of area A_S (shaded) in the figure below. In the figure, the detector D (including the appropriate spectral filter) views the surface A_S from below within a cylindrical container. The surface of area A_S is made of an optical diffusing material that is usually translucent plastic or a milky-looking glass (opal glass). Ideally, such a material has the property that if any point



During a time interval Δt , the number, $\Delta n(\hat{\xi}, \lambda)$, of photons in $d\lambda$ that stream through dA are all contained within the cylinder of volume $c_\ell \Delta t dA$. If the irradiance is independent of time (as we are assuming) then the spatial distribution of photons within the cylinder is uniform and the density of photons within the cylinder (and also in the sphere shown in the figure), $d^2 u_p(\hat{\xi}, \lambda) = \Delta n(\hat{\xi}, \lambda) / (c_\ell \Delta t dA)$, is uniform. The differential $d^2 u_p(\hat{\xi}, \lambda)$ is of second order, because $\hat{\xi}$ is within $d\Omega$ and λ is within $d\lambda$. The radiant power falling on dA is

$$d^3 \mathcal{P} = \frac{hc_\ell}{\lambda} \frac{\Delta n(\hat{\xi}, \lambda)}{\Delta t} = \frac{hc_\ell}{\lambda} [c_\ell d^2 u_p(\hat{\xi}, \lambda)] dA = L(\hat{\xi}, \lambda) dA d\Omega d\lambda.$$

Therefore $L(\hat{\xi}, \lambda) d\Omega = (hc_\ell / \lambda) c_\ell d^2 u_p(\hat{\xi}, \lambda) / d\lambda$. If we now integrate over all solid angles around “p,” we find

$$E_0(p, \lambda) = \frac{hc_\ell}{\lambda} \frac{d}{d\lambda} [c_\ell u_p(\lambda)],$$

but since $(hc_\ell / \lambda) u_p(\lambda) = u_{em}(\lambda)$, we arrive at the result in the text.

on the surface is illuminated by a thin beam of parallel light propagating in the direction $\hat{\xi}$ (left panel), and having a cross sectional area $A_B \ll A_S |\hat{n}_s \bullet \hat{\xi}|$, the output of the detector D is independent of $\hat{\xi}$,²⁰ and proportional to $P_B(\hat{\xi})$, the incident spectral radiant power in the beam,²¹ i.e., the current in the detector is $i \propto P(\hat{\xi})$, and independent of $\hat{\xi}$. When the sensor is illuminated by a broad beam (center panel), with $A_B \gg A_S |\hat{n}_s \bullet \hat{\xi}|$, and that can be imagined to be made up of a large number of thin beams, the power from the photons actually striking the surface is proportional to $|\hat{n}_s \bullet \hat{\xi}|$, so the current in the detector is proportional to $|\hat{n}_s \bullet \hat{\xi}|$.²² Finally, when the surface is illuminated by an extended source (right panel), if we consider the power only within the solid angle $d\Omega(\hat{\xi})$, i is proportional to $L(\hat{\xi}) |\hat{n} \bullet \hat{\xi}| A_S d\Omega(\hat{\xi})$. Therefore, when the surface can view the entire upward hemisphere, the actual current will be

$$i \propto P_{\text{Total}} = \int_{\Omega} L(\hat{\xi}) |\hat{n} \bullet \hat{\xi}| A_S d\Omega(\hat{\xi}) = A_S E,$$

where E is the *plane* irradiance on the horizontal surface of which A_S is a part. In the orientation shown, the current in the detector is proportional to the downwelling irradiance: $i \propto E_d$. Turn the whole assembly upside down and $i \propto E_u$. The whole assembly, diffuser, detector, and (hopefully watertight) container is called an *irradiance meter*. The proportionality constant relating i and E is referred to as the *calibration constant* of the irradiance meter.²³ By placing several detectors with different filters (or some equivalent arrangement) the full spectrum of the irradiance can be obtained.

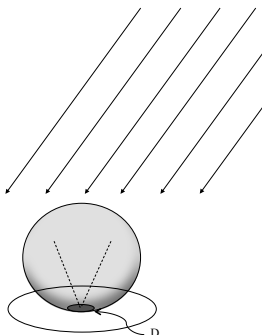
Is there a way that we can use such an approach to measure the scalar irradiance? The answer is yes. Consider rather than a plane diffuser, a spherical diffuser with the detector

²⁰A sheet of white paper is an example of a diffuser that is close to ideal. This can be verified by illuminating the sheet from behind with a narrow beam formed by passing sunlight through a small hole (or by illuminating the sheet with a laser beam, e.g., from a laser pointer) and viewing the transmitted beam (a bright spot) from the other side. As the angle between the surface normal to the beam is varied, it will be observed that the brightness of the spot on the back surface of the paper varies only slightly, and with most of the variation occurring at large angles (e.g., $\gtrsim 80^\circ$). What variation there is can be reduced by using two sheets, i.e., a thicker diffuser, etc., but with a decrease in overall brightness.

²¹Note, P_B is proportional to the number of photons falling on the surface per second (N_B).

²²For this dependence on angle for a broad beam, such a diffuser is often referred to as a *cosine collector*.

²³It is important to note that if the radiometric calibration of the instrument is carried out in air (for convenience), the effect of placing the instrument in water must be investigated. The performance of a diffuser changes when it is passed from air to water because the change in the refractive index contrast at its surface, and this effect (the immersion effect) must be assessed. In addition, the diffuser is usually not placed flush with the surface of the container as shown in the drawing, but usually protrudes a little above it, allowing light to come in from the sides in order to better approximate $|\hat{n}_s \bullet \hat{\xi}|$ in the response to a broad beam at low values of $|\hat{n}_s \bullet \hat{\xi}|$. Details of the design and calibration of such an instrument can be found in [Tyler and Smith \[1970\]](#).



inside as shown in the figure above. Such a diffuser/detector arrangement would behave as the plane diffuser-detector arrangement when illuminated by a narrow pencil of light, i.e., the response of the detector would be independent of the beam's direction *and* of the position where it struck the sphere,²⁴ but when illuminated by a broad beam, the response would be independent of the direction of the incident light. Viewing a diffuse scene, the power collected by the sphere from $\hat{\xi}$ would be given by $L(\hat{\xi})\pi R^2 d\Omega(\hat{\xi})$, where R is the radius of the sphere and πR^2 is the *projected* area of the sphere in the $\hat{\xi}$ direction. The total spectral radiant power is found by addition (integration) and the current is therefore proportional to

$$\int_{\Xi} L(\hat{\xi}) d\Omega(\hat{\xi}),$$

where Ξ is the range of directions $\hat{\xi}$ that can influence the sphere. To provide for a well defined Ξ a baffle can be placed over the disk on which the detector sits in the diagram, restricting the field of view of the sphere to angles between the vertical and horizontal, i.e., blocking most light propagating in upward directions. In that case, the current in the detector, $i \propto E_{0d}$. Turn the assembly upside down and $i \propto E_{0u}$. Sometimes the sphere is at the end of a thin post protruding from a small container. In this case Ξ is nearly 4π Sr and the current in the detector is approximately proportional to E_0 . This is usually the case for measurements of *quantum irradiance* (see Chapter 6). As with the plane irradiance meter, calibration must be carried out to ensure that the scalar instrument performs as desired. Of the two instruments, the plane irradiance meter is much more common than the scalar irradiance meter.

²⁴Some additional optics within the sphere would be required for the position independence, but clearly the response at the center of the sphere would be independent of position on the surface.

2.5.9 Auxiliary Optical Properties

There are several auxiliary optical properties that are useful in radiative transfer theory. The first is the *scattering phase function* defined according to²⁵

$$P(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) \triangleq \frac{4\pi\beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda)}{b(\vec{r}, \lambda)},$$

the second is the *albedo of single scattering* (also referred to as the single scattering albedo),

$$\omega_0(\vec{r}, \lambda) \triangleq \frac{b(\vec{r}, \lambda)}{c(\vec{r}, \lambda)}.$$

These are important, because for a collection of identical scatterers, both quantities are *independent* of concentration. In addition, ω_0 is the *probability* of scattering of a photon, $(1 - \omega_0)$ is the probability of absorption, and given that scattering takes place,

$$\frac{P(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda)}{4\pi} d\Omega(\hat{\xi})$$

is that *probability* of scattering of a photon at \vec{r} from $\hat{\xi}'$ to $\hat{\xi}$ (within the range of solid angles $d\Omega(\hat{\xi})$). For atomic scattering (also called *Rayleigh scattering*), Eqs. (2.25) and (2.26) show that

$$P(\Theta) = \frac{3}{4} [1 + \cos^2 \Theta], \quad (2.45)$$

where Θ is the scattering angle.

The reduction to plane-parallel media in which the inherent optical properties depend only on z is made by replacing \vec{r} by z as before. In addition, in such a medium, another useful quantity is the *optical depth* defined through

$$d\tau(z, \lambda) = c(z, \lambda) dz \quad \text{or} \quad \tau(z, \lambda) = \int_0^z c(z, \lambda) dz.$$

In terms of these auxiliary optical properties the RTE for a 1-d medium becomes

$$\hat{\xi} \cdot \hat{e}_z \frac{\partial}{\partial \tau} L(\tau, \hat{\xi}, \lambda) = -L(\tau, \hat{\xi}, \lambda) + \frac{\omega_0(\tau, \lambda)}{4\pi} \int_{\text{All } \hat{\xi}'} P(\tau, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(\tau, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') + \frac{Q(\tau, \hat{\xi}, \lambda)}{c(\tau, \lambda)}. \quad (2.46)$$

²⁵Some authors define the phase function as we do here, while others define it *without* the 4π factor, i.e., β/b . This causes endless confusion. The reader should always make sure which convention is being used when consulting the radiative transfer literature.

2.6 Methods of Solving the RTE

There are only a few situations in which analytical solutions of the RTE have been found. Generally, all are cases in which $\omega_0(\tau, \lambda) = 0$ or $P(\tau, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) \propto \delta^{(2)}(\hat{\xi}' - \hat{\xi})$, i.e., there is no scattering or only forward scattering. However, there are some special situations in which analytical solutions can be found, namely deep within a 1-d semi-infinite medium far from the boundary.²⁶ When the boundary is illuminated from outside the medium, the radiance can be found to within a multiplicative constant for scattering with the isotropic and Rayleigh phase functions for all values of ω_0 . Analytic solutions can also be found for such a medium when the source is internal with a uniform, isotropic, intensity density, i.e., $Q(\tau, \hat{\xi})$ independent of τ and $\hat{\xi}$. Unfortunately, for settings of interest in remote sensing, we must rely on numerical solutions.

There are several methods for obtaining numerical solutions of the RTE. The two that have been most useful for our purposes are the successive order of scattering method, used in developing the atmospheric correction algorithm, and the Monte Carlo method used to provide solutions to the transfer equation in the water. These are discussed in this section.

2.6.1 Successive Order of Scattering

Consider a slab of a 1-d medium that extends from $z = 0$ to $z = z_1$ (Figure 2.12). Assume that there are no internal sources, i.e., $Q = 0$. Also assume that the inherent optical properties of the slab are independent of position, noting that if they do depend on depth, such stratification could be represented by a set of homogeneous slabs. Under these conditions, the RTE becomes

$$\hat{\xi} \cdot \hat{e}_z \frac{\partial}{\partial \tau} L(\tau, \hat{\xi}) = -L(\tau, \hat{\xi}) + \frac{\omega_0}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \quad (2.47)$$

and the total optical thickness of the medium is $\tau_1 = cz_1$. Note that for compactness we are omitting wavelength λ from the argument list. To find the radiance $L(\tau, \hat{\xi})$ this equation must be solved subject to boundary conditions at the top and bottom of the slab:

$$L(0, \hat{\xi}) = L_{\text{Top}}(\hat{\xi}) \quad \text{and} \quad L(\tau_1, \hat{\xi}) = L_{\text{Bottom}}(\hat{\xi}), \quad (2.48)$$

where in these expressions $\hat{\xi} \cdot \hat{n}_{\text{Top}} < 0$, $\hat{\xi} \cdot \hat{n}_{\text{Bottom}} < 0$, and \hat{n}_{Top} and \hat{n}_{Bottom} are the *outward* normals of the top and bottom, respectively.

²⁶Later we will refer to this as the *asymptotic regime* (see page ??).

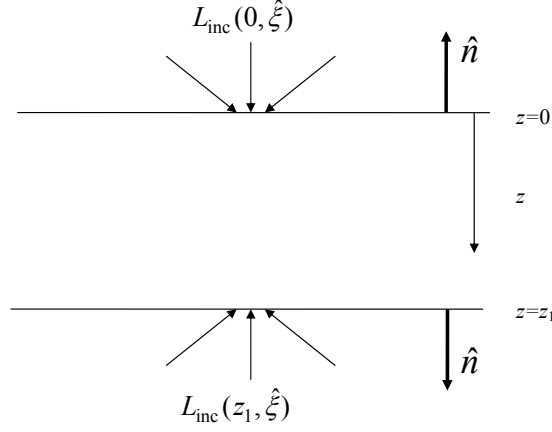


Figure 2.12: Geometry of a 1-d slab. The (outward) normals, \hat{n} , for the top and bottom surfaces are shown. Depth z is measured downward, and the radiances $L_{\text{inc}}(0, \hat{\xi})$ and $L_{\text{inc}}(z_1, \hat{\xi})$ are incident on the top and bottom of the slab, respectively.

We now assume that the radiance can be expanded in a series in ω_0 :

$$L(\tau, \hat{\xi}) = L^{(0)}(\tau, \hat{\xi}) + \omega_0 L^{(1)}(\tau, \hat{\xi}) + \omega_0^2 L^{(2)}(\tau, \hat{\xi}) \cdots = \sum_{n=0}^{\infty} \omega_0^n L^{(n)}(\tau, \hat{\xi}). \quad (2.49)$$

Inserting this into the RTE, and rearranging the terms yields

$$\begin{aligned} & \left[\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(0)}(\tau, \hat{\xi}) + L^{(0)}(\tau, \hat{\xi}) \right] + \\ & \omega_0 \left[\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(1)}(\tau, \hat{\xi}) + L^{(1)}(\tau, \hat{\xi}) - \frac{1}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L^{(0)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}') \right] + \\ & \omega_0^2 \left[\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(2)}(\tau, \hat{\xi}) + L^{(2)}(\tau, \hat{\xi}) - \frac{1}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L^{(1)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}') \right] + \\ & \cdots \\ & = 0. \end{aligned}$$

Since ω_0 is arbitrary, satisfying this equation (for any finite number of terms) requires that

the terms in the square brackets vanish individually, i.e.,

$$\begin{aligned}
\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(0)}(\tau, \hat{\xi}) + L^{(0)}(\tau, \hat{\xi}) &= 0, \\
\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(1)}(\tau, \hat{\xi}) + L^{(1)}(\tau, \hat{\xi}) &= \frac{1}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L^{(0)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \\
\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(2)}(\tau, \hat{\xi}) + L^{(2)}(\tau, \hat{\xi}) &= \frac{1}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L^{(1)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \\
&\dots \\
\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(n)}(\tau, \hat{\xi}) + L^{(n)}(\tau, \hat{\xi}) &= \frac{1}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L^{(n-1)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \\
&\dots
\end{aligned} \tag{2.50}$$

Thus, solving all of the individual equations in Eq. (2.50) provides a solution to Eq. (2.47). However, to solve the problem we posed, the solution must also satisfy the boundary conditions Eq. (2.48). There is a considerable amount of freedom on how these boundary conditions can be spread over the various $L^{(n)}$'s: the simplest choice is

$$L^{(0)}(0, \hat{\xi}) = L_{\text{Top}}(\hat{\xi}) \quad \text{and} \quad L^{(0)}(\tau_1, \hat{\xi}) = L_{\text{Bottom}}(\hat{\xi}),$$

with

$$L^{(n)}(0, \hat{\xi}) = 0 \quad \text{and} \quad L^{(n)}(\tau_1, \hat{\xi}) = 0 \quad \text{for} \quad n > 0.$$

Clearly satisfying all of the equations in Eq. (2.50) subject to these boundary conditions provides a solution to Eq. (2.47) subject to Eq. (2.48). All of this hinges of course on the assumption that the series Eq. (2.49) converges. Preisendorfer [1965] has been able to prove convergence of the series under the condition that $\omega_0 < 1$, i.e., under the same conditions that one can prove uniqueness of the solutions of the RTE.

The method described above provides what is called the *successive order of scattering* solution. The reason for this is that $L^{(0)}$ is the radiance from the outside that has not been scattered, $L^{(1)}$ is the radiance in the medium that has undergone one scattering in the medium, etc.²⁷

²⁷This interpretation allows us compute the average number of scatterings that have been undergone to yield the radiance at a given point in the medium. Equation (2.49) shows that the probability of n scatterings (P_n) is just $\omega_0^n L_n / L$. Therefore, the mean number of scatterings is given by

$$\langle n \rangle = \sum_0^\infty n P_n = \sum_0^\infty n \frac{\omega_0^n L_n}{L} = \frac{\omega_0}{L} \frac{dL}{d\omega_0}.$$

Note that the average number of scatterings depends on the depth in the medium as well as the direction of propagation of the radiance.

Examination of the final equations of the successive order of scattering solution suggests that stratified media can be approached in a similar manner. If we let the phase function and ω_0 be functions of τ , and formally write

$$\begin{aligned}
 \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(0)}(\tau, \hat{\xi}) + L^{(0)}(\tau, \hat{\xi}) &= 0, \\
 \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(1)}(\tau, \hat{\xi}) + L^{(1)}(\tau, \hat{\xi}) &= \frac{\omega_0(\tau)}{4\pi} \int_{\text{All } \hat{\xi}'} P(\tau, \hat{\xi}' \rightarrow \hat{\xi}) L^{(0)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \\
 \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(2)}(\tau, \hat{\xi}) + L^{(2)}(\tau, \hat{\xi}) &= \frac{\omega_0(\tau)}{4\pi} \int_{\text{All } \hat{\xi}'} P(\tau, \hat{\xi}' \rightarrow \hat{\xi}) L^{(1)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \\
 &\dots \\
 \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L^{(n)}(\tau, \hat{\xi}) + L^{(n)}(\tau, \hat{\xi}) &= \frac{\omega_0(\tau)}{4\pi} \int_{\text{All } \hat{\xi}'} P(\tau, \hat{\xi}' \rightarrow \hat{\xi}) L^{(n-1)}(\tau, \hat{\xi}') d\Omega(\hat{\xi}'), \\
 &\dots
 \end{aligned} \tag{2.51}$$

Then solving these subject to the boundary conditions

$$L^{(0)}(0, \hat{\xi}) = L_{\text{Top}}(\hat{\xi}) \quad \text{and} \quad L^{(0)}(\tau_1, \hat{\xi}) = L_{\text{Bottom}}(\hat{\xi}),$$

with

$$L^{(n)}(0, \hat{\xi}) = 0 \quad \text{and} \quad L^{(n)}(\tau_1, \hat{\xi}) = 0 \quad \text{for} \quad n > 0,$$

and summing the individual solutions

$$L(\tau, \hat{\xi}) = L^{(0)}(\tau, \hat{\xi}) + L^{(1)}(\tau, \hat{\xi}) + L^{(2)}(\tau, \hat{\xi}) \dots = \sum_{n=0}^{\infty} L^{(n)}(\tau, \hat{\xi}), \tag{2.52}$$

clearly satisfies Eq. (2.47) and Eq. (2.48) and therefore is a solution to the problem of a *stratified* slab. Note that in this case

$$\tau(z) = \int_0^z c(z) dz, \quad \text{so} \quad \tau_1(z_1) = \int_0^{z_1} c(z) dz.$$

The extension of the successive order of scattering method to the case when the polarization of the radiance is considered, i.e., Eq. (2.34) (or really, Eq. (2.113)), is immediate — no new concepts are required.

In general the successive order of scattering method is useful whenever the product of ω_0 and τ_1 is small, in which case only a small number of terms in the series is required. It is most useful for radiative transfer in the atmosphere in the visible region of the spectrum. In the ocean, where τ_1 is essentially ∞ , the technique is only valuable for $\omega_0 \ll 1$.

2.6.2 The Monte Carlo Method

The Monte Carlo solution is a direct simulation of the radiative transfer process wherein a large number of individual photons are followed through the medium.²⁸ Decisions concerning where an individual photon interacts with the medium, what type of interaction is involved, and the fate of photons after the interaction, are based on the probability that these various events will occur. To explain the procedure we will discuss a solution to the problem described in the previous section — a 1-d slab of optical thickness τ_1 , with a given $\omega_0(\tau)$ and $P(\tau, \hat{\xi}' \rightarrow \hat{\xi})$. For simplicity, we will assume that the radiance on the upper boundary is in a narrow range of solid angles $\Delta\Omega(\hat{\xi}_0)$ around a single direction $\hat{\xi}_0$, i.e., a beam, and that there is no radiance incident on the lower boundary from the outside. This simplifying assumption results in no loss in generality because the solution for any incident distribution of radiance on the top of the slab can be developed from a simple superposition of such beams. Let $L(0, \hat{\xi}_0)$ be the radiance incident on the upper boundary. We need to relate this to a photon rate. This is straightforward as the radiant power falling on a surface is proportional to the rate N at which photons strike the surface, i.e., Eq. (2.1). Recall that the radiance incident on an area dA_S of a surface, with normal \hat{n} , from a direction $\hat{\xi}$ within a range of solid angles $d\Omega_S(\hat{\xi})$ is (Eq. (2.11))

$$L(\hat{\xi}) = \frac{d^2 P(\hat{\xi})}{|\hat{\xi} \bullet \hat{n}| dA_S d\Omega_S(\hat{\xi})} \propto \frac{d^2 N(\hat{\xi})}{|\hat{\xi} \bullet \hat{n}| dA_S d\Omega_S(\hat{\xi})}.$$

Letting the proportionality constant be C , and the rate at which photons are incident on an area A_S at the top of the slab ($\tau = 0$) be N_0 , we have

$$L(0, \hat{\xi}_0) = C \frac{N_0}{|\hat{\xi}_0 \bullet \hat{e}_z| A_S d\Omega_S(\hat{\xi}_0)}.$$

In a similar manner, within the medium, if $N(\tau, \hat{\xi})$ is the rate at which photons propagating in the direction $\hat{\xi}$ within $d\Omega(\hat{\xi})$ cross a horizontal area A (normal to \hat{e}_z) at an optical depth τ is

$$L(\tau, \hat{\xi}) = C \frac{N(\tau, \hat{\xi})}{|\hat{\xi} \bullet \hat{e}_z| A d\Omega(\hat{\xi})}.$$

Note that we have chosen horizontal areas because L depends only on the vertical coordinate z . Practically, this means that $N_0 \propto A_S$ and $N(\tau, \hat{\xi}) \propto A$. Thus, if we inject a number N_0 of photons within $d\Omega(\hat{\xi}_0)$ into the slab from the top and trace their life histories, the number that cross a horizontal surface at τ within $d\Omega(\hat{\xi})$ is proportional to the radiance $L(\tau, \hat{\xi})$ on that surface. Since the photon rates are proportional

²⁸In this section we make extensive use of the photon concept, as the Monte Carlo method was originally developed to solve problems that focussed on the transport of neutrons in nuclear reactors, i.e., the multiple scattering of physical *particles*. A development of the Monte Carlo method of solving the RTE without the introduction of photons is provided in Appendix 2 to this chapter.

to the areas, one must choose $A = A_S$. It is customary to define F_0 to be the irradiance on a plane normal to the incident beam, i.e., Eq. (2.17) $F_0 \triangleq L(0, \hat{\xi}_0) d\Omega(\hat{\xi}_0)$, so $F_0 = L(0, \hat{\xi}_0) d\Omega(\hat{\xi}_0) = CN_0/(|\hat{\xi}_0 \bullet \hat{e}_z|A)$. Then if $\Delta N(\tau, \hat{\xi})$ is the number of photons that cross a horizontal surface at optical depth τ within a *finite* range of solid angles $\Delta\Omega(\hat{\xi})$, the radiance falling on this surface can be estimated by

$$L(\tau, \hat{\xi}) = C \frac{\Delta N(\tau, \hat{\xi})}{|\hat{\xi} \bullet \hat{e}_z| A \Delta\Omega(\hat{\xi})},$$

or, eliminating C

$$\frac{L(\tau, \hat{\xi})}{F_0} = \frac{|\hat{\xi}_0 \bullet \hat{e}_z|}{|\hat{\xi} \bullet \hat{e}_z|} \frac{1}{N_0} \frac{\Delta N(\tau, \hat{\xi})}{\Delta\Omega(\hat{\xi})}. \quad (2.53)$$

For $\Delta\Omega(\hat{\xi})$ sufficiently small, we expect that $\Delta N(\tau, \hat{\xi}) \propto \Delta\Omega(\hat{\xi})$, so $\Delta N(\tau, \hat{\xi})/\Delta\Omega(\hat{\xi})$ will have a well defined limit. However, if $\Delta\Omega(\hat{\xi})$ is too small, $\Delta N(\tau, \hat{\xi})$ may fluctuate wildly with $\Delta\Omega(\hat{\xi})$ unless N_0 is extremely large. In a Monte Carlo simulation the unit sphere is divided into a finite number of solid angles $\Delta\Omega(\hat{\xi}_i)$, each around a given direction $\hat{\xi}_i$. The number of directions is chosen to provide an adequate $\Delta N(\tau, \hat{\xi}_i)$ for each $\Delta\Omega(\hat{\xi}_i)$.

Given Eq. (2.53), the irradiances are also easy to estimate:

$$\begin{aligned} E_d(\tau) &= \int_{\hat{\xi} \bullet \hat{e}_z > 0} |\hat{\xi} \bullet \hat{e}_z| L(\tau, \hat{\xi}) d\Omega(\hat{\xi}) \\ &\approx \sum_{i, \hat{\xi}_i \bullet \hat{e}_z > 0} |\hat{\xi}_i \bullet \hat{e}_z| L(\tau, \hat{\xi}_i) \Delta\Omega(\hat{\xi}_i) \\ &= F_0 \sum_{i, \hat{\xi}_i \bullet \hat{e}_z > 0} \frac{\Delta N(\tau, \hat{\xi}_i)}{N_0}, \end{aligned} \quad (2.54)$$

with a similar expression for $E_u(\tau)$, but with $\hat{\xi}_i \bullet \hat{e}_z < 0$. Thus, the irradiances $E_d(\tau)$ and $E_u(\tau)$ are simply F_0 times the fraction of injected photons that pass a surface at optical depth τ traveling in the downward and upward directions, respectively. Likewise, the scalar irradiance is

$$E_0(\tau) = F_0 \sum_i \frac{1}{|\hat{\xi}_i \bullet \hat{e}_z|} \frac{\Delta N(\tau, \hat{\xi}_i)}{N_0}. \quad (2.55)$$

Equations (2.53), (2.54), and (2.55) provide the basis for Monte Carlo estimates of the important radiometric quantities. It should be noted that if only irradiances are required in the problem, $\hat{\xi}$ can be allowed to vary continuously, and the sums can be accumulated during the computation, e.g., each time a photon passes τ we add $1/|\hat{\xi} \bullet \hat{e}_z|$ to the sum for $E_0(\tau)$ and divide by N_0 at the end.

Thus far, we have shown that, given the life histories of a large number of photons, we can estimate the interesting radiometric quantities: $L(\tau, \hat{\xi})$, $E_d(\tau)$, $E_u(\tau)$, and $E_0(\tau)$. Now we consider how to develop these life histories. After injecting a photon into the medium in the direction $\hat{\xi}_0$, we need to determine where it will interact. From the definition of c , we know that the radiant power in a collimated beam satisfies the differential equation

$$dP = -Pc \, dl, \quad \text{so} \quad P(l) = P(0) \exp(-cl).$$

Therefore, if there are N_0 photons per second in the beam at $l = 0$, the number *remaining* at l is

$$N(l) = N_0 \exp(-cl).$$

This means that the probability that a photon reaches l is $\exp(-cl)$. Having reached l , the probability of an interaction (absorption or scattering) within dl is $c \, dl$ (recall $dN/N = -c \, dl$, where dN is the number interacting within dl). Therefore, starting from $l = 0$, the probability of an interaction between l and $l + dl$, $d\mathcal{P}_l(dl)$ is

$$d\mathcal{P}_l(dl) = \frac{dN}{N_0} = \frac{dN}{N} \exp(-cl) = c \, dl \exp(-cl) \triangleq p_l(l) \, dl,$$

where $p_l(l)$ is the *probability density* for an interaction at l . Now, assume that we have a set of *random numbers* ρ *uniformly distributed* on the interval 0 to 1. Uniformly distributed means that the probability of finding a given random number (ρ_i) between ρ and $\rho + d\rho$ (which is $p_\rho(\rho) d\rho$) is independent of ρ , i.e., $p_\rho(\rho) = 1$, and the probability of the random number being in $d\rho$ is just $d\rho$. So, if we divided the entire interval $[0,1]$ into a set of equal-width subdivisions and picked a large number (\mathcal{N}) of uniformly distributed random numbers, the number in each subdivision would be the same (except for statistical fluctuations that approach zero as $\mathcal{N} \rightarrow \infty$). Such a distribution is usually labeled $U[0,1]$. Sequences of random numbers distributed according to $U[0,1]$ are approximated by *pseudo* random number generators available for most computers. Given the sequence of random numbers we can generate photon interactions as follows. If we want a random number in $d\rho$ to represent an interaction in dl , then we must require

$$p_\rho(\rho) d\rho = p_l(l) dl, \quad \text{or} \quad d\rho = p_l(l) dl.$$

Integrating

$$\int_0^{\rho_0} d\rho = \int_0^{l_0} p_l(l) dl, \quad \text{or} \quad \rho_0 = 1 - \exp(-cl_0),$$

so

$$l_0 = -\frac{1}{c} \ln(1 - \rho_0).$$

The path length l_0 is measured along the path of the photon $\hat{\xi}_0$. In terms of depth the photon interacts at z_0 given by

$$z_0 = -\frac{|\hat{\xi}_0 \bullet \hat{e}_z|}{c} \ln(1 - \rho_0).$$

If the medium is stratified, i.e., $c = c(z)$, then a similar argument shows that

$$\tau_0 = -|\hat{\xi}_0 \bullet \hat{e}_z| \ell n(1 - \rho_0). \quad (2.56)$$

A little thought shows that these equations must also provide the distances between any two interactions Δl , Δz , and $\Delta \tau$.

If $\tau_0 > \tau_1$ (the optical depth of the slab), the photon leaves the medium and a new photon is injected after the various quantities in Eq. s(2.53), (2.54), and (2.55) are updated. If $\tau_0 < \tau_1$ the photon interacts with the medium at τ_0 , and we next need to determine the nature of the interaction. Given that an interaction takes place at $z = z_0$ or $\tau = \tau_0$, what kind of interaction will be, absorption or scattering? The probability of scattering is $\omega_0 = b/c$. Thus if we choose a second random number ρ_1 distributed according to $U[0,1]$, we have at $\tau = \tau_0$

$$\begin{aligned} \rho_1 \leq \omega_0(\tau) & \text{ implies scattering, and} \\ \rho_1 > \omega_0(\tau) & \text{ implies absorption.} \end{aligned} \quad (2.57)$$

Clearly these inequalities hold for *any* interaction, not just the first.

If the photon is absorbed, it is terminated, and a new photon is injected into the medium. If it is scattered, we need to determine the new direction of the photon's propagation. The new direction of propagation is provided by the scattering phase function, because

$$\frac{1}{4\pi} P(\tau, \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi})$$

is the probability that the scattered direction will be within $d\Omega(\hat{\xi})$. Thus, if the unit sphere around the interaction point is divided into a number n of solid angles $\Delta\Omega(\hat{\xi}_i)$ around a set of directions $\hat{\xi}_i$, the probability of scattering from a direction $\hat{\xi}_j$ into $\Delta\Omega(\hat{\xi}_i)$ is

$$p(\hat{\xi}_j \rightarrow \hat{\xi}_i) \Delta\Omega(\hat{\xi}_i) = \frac{1}{4\pi} \int_{\Delta\Omega(\hat{\xi}_i)} P(\tau, \hat{\xi}_j \rightarrow \hat{\xi}) d\Omega(\hat{\xi}). \quad (2.58)$$

Dividing the interval $[0,1]$ into $\rho_1, \rho_2 - \rho_1, \rho_3 - \rho_2, \dots, \rho_i - \rho_{i-1}, \dots, 1 - \rho_{n-1}$, so that for each $\hat{\xi}_j$

$$\rho_i - \rho_{i-1} = p(\hat{\xi}_j \rightarrow \hat{\xi}_i) \Delta\Omega(\hat{\xi}_i), \quad (2.59)$$

given $\hat{\xi}_j$ and a $U[0,1]$ random number ρ , the scattered direction is $\hat{\xi}_i$ when $\rho_{i-1} < \rho < \rho_i$. This is shown schematically in Figure 2.13, in which a mapping into various directions $\hat{\xi}_i$ for a particular $\hat{\xi}_j$ is provided. There is a separate mapping for each $\hat{\xi}_j$. In the mapping shown here, the scattering $\hat{\xi}_j \rightarrow \hat{\xi}_3$ is highly probable, but the scattering $\hat{\xi}_j \rightarrow \hat{\xi}_4$ is relatively improbable.

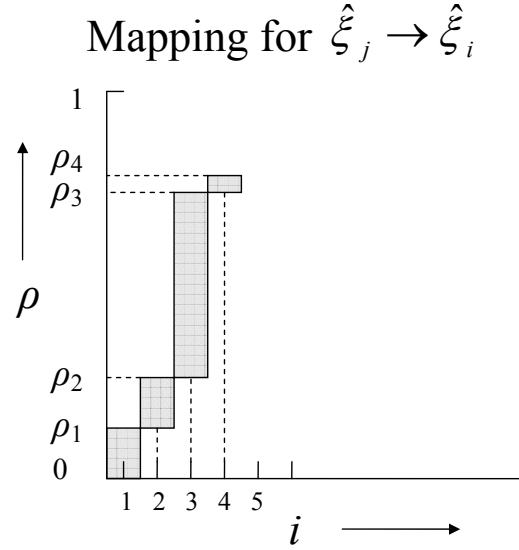


Figure 2.13: This shows the mapping between the random number ρ and the scattered direction $\hat{\xi}_i$ for scattering from $\hat{\xi}_j \rightarrow \hat{\xi}_i$. There is a mapping such as this for each $\hat{\xi}_j$. For the cases shown, scattering from $\hat{\xi}_j \rightarrow \hat{\xi}_3$ is highly probable (the probability being $\rho_3 - \rho_2$), but scattering from $\hat{\xi}_j \rightarrow \hat{\xi}_4$ is relatively improbable.

In all cases of interest to us here, the scattering phase function depends only on the angle between the incident and the scattered directions. This considerably simplifies the determination of the direction of propagation of the scattered photon. Consider a scattering event from $\hat{\xi}_j$ to $\hat{\xi}_i$ depicted in Figure 2.14, in which the scattering is specified by the angles Θ and Φ . The scattering phase function depends only on Θ , being independent of Φ . Therefore,

$$\frac{P(\hat{\xi}_j \rightarrow \hat{\xi}_i)}{4\pi} d\Omega(\hat{\xi}_i) = \frac{P(\Theta, \Phi)}{4\pi} d\Omega = \frac{P(\Theta)}{4\pi} \sin \Theta d\Theta d\Phi = p_\Theta(\Theta) p_\Phi(\Phi) d\Theta d\Phi,$$

where $p_\Theta(\Theta)$ and $p_\Phi(\Phi)$ are the probability densities for Θ and Φ , respectively. Since all the Φ 's are equally probable,

$$p_\Phi(\Phi) = \frac{1}{2\pi}.$$

It follows then that

$$p_\Theta(\Theta) = \frac{1}{2} P(\Theta) \sin \Theta.$$

Thus, given random numbers ρ distributed according to $U[0,1]$, for ρ in $d\rho$ to represent Φ in $d\Phi$ or Θ in $d\Theta$, we require that

$$d\rho_\Phi = \frac{d\Phi}{2\pi} \quad \text{and} \quad d\rho_\Theta = P(\Theta) \sin \Theta d\Theta$$

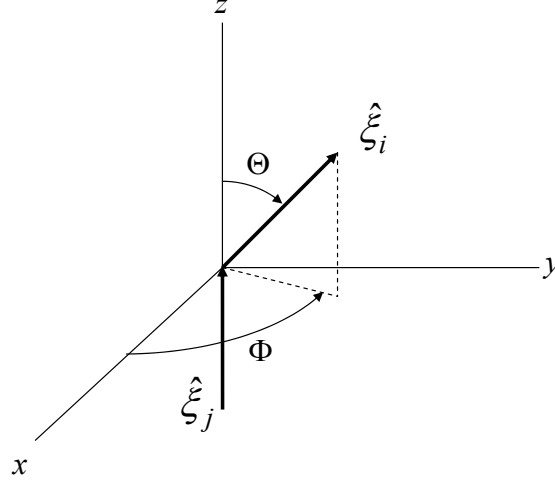


Figure 2.14: The scattering angle Θ from $\hat{\xi}_j$ to $\hat{\xi}_i$. The angle Φ is the azimuth of the scattering plane.

or

$$\Phi = 2\pi\rho_\Phi \quad \text{and} \quad \rho_\Theta = \frac{1}{2} \int_0^\Theta P(\Theta') \sin \Theta' d\Theta'. \quad (2.60)$$

Given the initial photon direction $\hat{\xi}_j$ and the scattering angles Θ and Φ , we next need to determine the direction of propagation of the scattered photon, $\hat{\xi}_i$. Figure 2.15 provides the geometry of the scattering in a fixed coordinate system in which z is into the slab (medium) and the x - y plane is parallel to the surface. The simplest way of analytically specifying $\hat{\xi}_j$ and $\hat{\xi}_i$ is through their direction cosines:

$$\hat{\xi}_j = \xi_j^x \hat{e}_x + \xi_j^y \hat{e}_y + \xi_j^z \hat{e}_z \quad \text{and} \quad \hat{\xi}_i = \xi_i^x \hat{e}_x + \xi_i^y \hat{e}_y + \xi_i^z \hat{e}_z,$$

where

$$\begin{aligned} \xi_j^x &= \cos \phi_j \sin \theta_j & \xi_i^x &= \cos \phi_i \sin \theta_i \\ \xi_j^y &= \sin \phi_j \sin \theta_j & \xi_i^y &= \sin \phi_i \sin \theta_i \\ \xi_j^z &= \cos \theta_j & \xi_i^z &= \cos \theta_i. \end{aligned} \quad \text{and}$$

Given ξ_i^x , ξ_i^y , and ξ_i^z (the incident photon's direction) and the scattering angles Θ and Φ

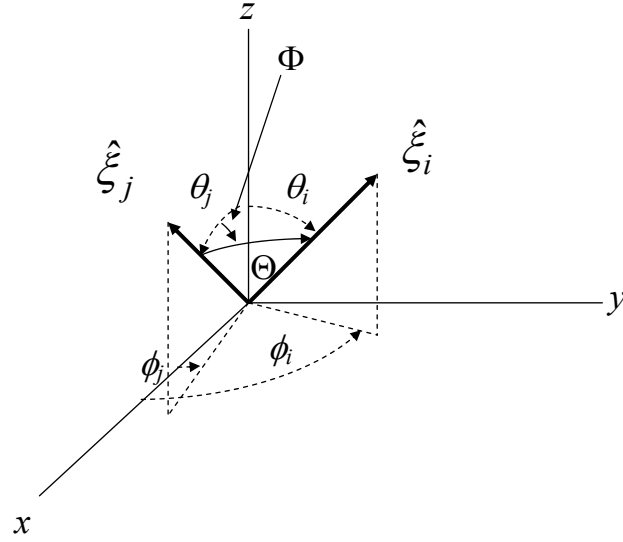


Figure 2.15: The scattering angle Θ ($\cos \Theta = \hat{\xi}_j \bullet \hat{\xi}_i$) for scattering from $\hat{\xi}_j \rightarrow \hat{\xi}_i$. The angles θ_i and ϕ_i are the polar and azimuth angle for $\hat{\xi}_i$, and θ_j and ϕ_j are the polar and azimuth angle for $\hat{\xi}_j$. Φ is the angle between the plane formed by the z -axis and $\hat{\xi}_j$ and the scattering plane.

the scattered photon's direction is found to be²⁹

$$\begin{aligned}\xi_i^x &= \xi_j^x \cos \Theta - \frac{\sin \Theta}{\sqrt{1 - (\xi_j^z)^2}} \left(\xi_j^y \sin \Phi + \xi_j^x \xi_j^z \cos \Phi \right) \\ \xi_i^y &= \xi_j^y \cos \Theta + \frac{\sin \Theta}{\sqrt{1 - (\xi_j^z)^2}} \left(\xi_j^x \sin \Phi - \xi_j^y \xi_j^z \cos \Phi \right) \\ \xi_i^z &= \xi_j^z \cos \Theta + \sqrt{1 - (\xi_j^z)^2} \sin \Theta \cos \Phi.\end{aligned}$$

When ξ_j^z is near unity (θ_j near zero), these should be replaced by

$$\begin{aligned}\xi_i^x &= \sin \Theta \cos \Phi \\ \xi_i^y &= \sin \Theta \sin \Phi \\ \xi_i^z &= \cos \Theta,\end{aligned}$$

to avoid computational difficulties.

²⁹The demonstration of these equations is tedious, but not difficult. One notes that in Figure 2.15 the unit vectors $\hat{\xi}_j$, $\hat{\xi}_i$, and \hat{e}_z all end on a unit sphere. If we apply the law of sines and the law of cosines to the spherical triangle formed by the ends of these vectors, the result follows after some algebra.

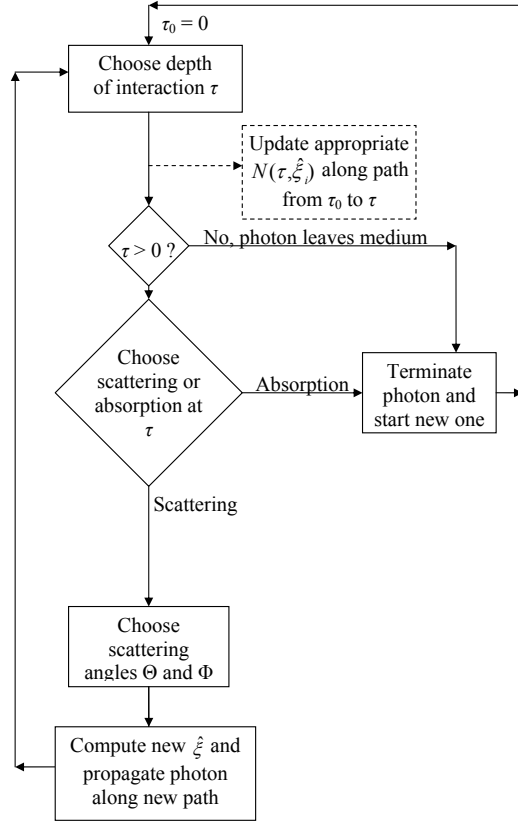


Figure 2.16: Flow chart of Monte Carlo radiative transfer process for the case of a semi-infinite medium, i.e., $\tau_1 \rightarrow \infty$. Photon enters the medium at the top. Detectors are continually updated along the photon's path. The photon is terminated on absorption, or whenever $\tau < 0$, and a new photon is introduced and followed.

The entire simulation process proceeds as follows. Given the new scattering direction, one chooses the distance to the next interaction, types it (absorption or scattering), and if scattering, finds the new $\hat{\xi}$, etc. The simulation's "detectors" are placed at various depths within the medium, and whenever a photon passes a given depth propagating along $\hat{\xi}$ within one of the $\Delta\Omega(\hat{\xi}_i)$'s it increases $\Delta N(\tau, \hat{\xi}_i)$ in Eq. (2.53) by one unit. When the photon is absorbed or leaves the medium, a new photon is started into the medium. The process is shown schematically in Figure 2.16 for the case of a semi-infinite medium, i.e., $\tau_1 \rightarrow 0$.

The number $\Delta N(\tau, \hat{\xi}_i)$ is subject to stochastic fluctuations, i.e., if the simulation were run twice with different sequences of random numbers, the results would be different in the two runs. The expected (standard) deviation in the number of photons in a detector's

collection range, i.e., $\Delta\Omega(\hat{\xi}_i)$, is approximately the square root of the number of photons collected. Thus, in a Monte Carlo simulation there is always a tradeoff between resolution (smaller $\Delta\Omega(\hat{\xi}_i)$'s) and accuracy (larger $\Delta N(\tau, \hat{\xi}_i)$'s).

There are many ways that have been developed to reduce the statistical fluctuations. These are usually referred to as *variance reduction techniques*. A simple example will suffice to provide the flavor of these techniques. Consider the slab problem above in the case where ω_0 is small but τ_1 is large. We will be faced with serious difficulties if we want to estimate $E_d(\tau_1)$ the irradiance exiting the bottom of the slab — few, if any, photons will make their way to the bottom, and the error in E_d will be large. To reduce this error, i.e., to make a more statistically significant estimate of $E_d(\tau_1)$, imagine that each photon not as a single quantity, but as a group of a large number of photons. At each collision, a fraction $(1 - \omega_0)$ of these are absorbed and a fraction ω_0 are scattered and go on to the next collision. After n collisions, a fraction ω_0^n survive. When the group of photons pass a detector depth, the number $\Delta N(\tau, \hat{\xi}_i)$ is *not* augmented by one as described above, rather it is augmented by ω_0^n , where n is the number of collisions prior to passing the detector depth. Thus, photons are *never* terminated in the medium, and each contributes either to $E_u(0)$ or $E_d(\tau_1)$. Although we will not attempt to quantify, it should be clear to the reader that the estimate of $E_d(\tau_1)$ obtained in this manner is more accurate than that in the direct simulation. Many other variance reduction techniques have been presented in the literature.

2.7 Additional Solution Methods

In this section, we will examine three additional solution techniques for problems in plane-parallel geometry. Our goal is to describe the basic ideas of the methods. We direct the reader wanting more details to the appropriate sources in the literature. These methods have been used in the analysis of radiative transfer in the ocean-atmosphere system; however, they are not used in this work³⁰ and may be omitted if desired. We begin by describing how the integral term in the RTE can be discretized, i.e., written as a finite or infinite sum, but first some preliminaries.

³⁰An exception is the proof of the existence of the asymptotic light field using the discrete-ordinate method.

2.7.1 Preliminaries

To simplify the discussion, we deal with some mathematical preliminaries that apply to all of the methods we will examine. These involve first writing the phase function as a Fourier series in the azimuth, which allows solving the transfer equation for each Fourier component separately, and second replacing the integral term in the RTE by a discrete sum. But before that, it is useful to make a notational change.

2.7.1.1 A Change in Notation

It is helpful to make a notational change when dealing with problems in plane-parallel geometry. We have been specifying the direction of propagation by the unit vector $\hat{\xi}$. In a spherical coordinate system with the z axis normal to the surface (and pointed downward in geophysical optics) and the x and y axes parallel to the surfaces of constant pressure, $\hat{\xi}$ is given by

$$\hat{\xi} = \hat{e}_x \cos \phi \sin \theta + \hat{e}_y \sin \phi \sin \theta + \hat{e}_z \cos \theta,$$

so specification of θ and ϕ completely determines $\hat{\xi}$. We will find it convenient not to use θ and ϕ , but rather to use $\cos \theta \triangleq u$ and ϕ . (Later in the Chapter, when $u < 0$ we will use $\mu = |u|$.) Thus, in this notation the solid angle is $d\Omega = du d\phi$, so Eq. s(2.37) and (2.46) become

$$\begin{aligned} u \frac{\partial}{\partial z} L(z, u, \phi, \lambda) = & -c(z, \lambda) L(z, u, \phi, \lambda) \\ & + \int_{-1}^1 du' \int_0^{2\pi} d\phi' \beta(z, u' \rightarrow u, \phi' \rightarrow \phi, \lambda) L(z, u', \phi', \lambda) \\ & + Q(z, u, \phi, \lambda), \end{aligned} \quad (2.61)$$

and

$$\begin{aligned} u \frac{\partial}{\partial \tau} L(\tau, u, \phi, \lambda) = & -L(\tau, u, \phi, \lambda) \\ & + \frac{\omega_0(\tau)}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' P(\tau, u' \rightarrow u, \phi' \rightarrow \phi, \lambda) L(\tau, u', \phi', \lambda) \\ & + \frac{Q(\tau, u, \phi, \lambda)}{c(\tau)}. \end{aligned} \quad (2.62)$$

Note that in slab geometry, radiance exiting the top of the slab has $u < 0$, while those exiting the bottom of the slab have $u > 0$. The irradiances in Eqs. (6.1) to (6.6) are given

by

$$\begin{aligned}
E_d(z, \lambda) &= \int_0^1 du \int_0^{2\pi} u L(z, u, \phi, \lambda) d\phi, \\
E_u(z, \lambda) &= - \int_{-1}^0 du \int_0^{2\pi} u L(z, u, \phi, \lambda) d\phi, \\
E_d(z, \lambda) - E_u(z, \lambda) &= \int_{-1}^1 du \int_0^{2\pi} u L(z, u, \phi, \lambda) d\phi, \\
E_0(z, \lambda) &= \int_{-1}^1 du \int_0^{2\pi} L(z, u, \phi, \lambda) d\phi.
\end{aligned} \tag{2.63}$$

2.7.1.2 Phase Function Symmetries and Fourier Analysis

When the volume scattering function only depends on the angle Θ between $\hat{\xi}'$ and $\hat{\xi}$, where $\cos \Theta = \hat{\xi}' \bullet \hat{\xi}$, e.g., scattering by a collection of atoms, the scattering phase function possesses many useful symmetries. Expressing $\hat{\xi}'$ and $\hat{\xi}$ in spherical coordinates,

$$\begin{aligned}
\cos \Theta &= \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi') \\
&= uu' + \sqrt{(1-u^2)(1-u'^2)} \cos(\phi - \phi').
\end{aligned} \tag{2.64}$$

This shows that Θ , and therefore $P(\Theta)$ is invariant under the following operations:

- interchanging u and u' : $P(u' \rightarrow u, \phi' \rightarrow \phi) = P(u \rightarrow u', \phi' \rightarrow \phi)$,
- interchanging ϕ and ϕ' : $P(u' \rightarrow u, \phi' \rightarrow \phi) = P(u' \rightarrow u, \phi \rightarrow \phi')$,
- interchanging both u 's and ϕ 's: $P(u' \rightarrow u, \phi' \rightarrow \phi) = P(u \rightarrow u', \phi \rightarrow \phi')$,
- changing signs of all angles: $P(u' \rightarrow u, \phi' \rightarrow \phi) = P(-u' \rightarrow -u, -\phi' \rightarrow -\phi)$.

Combining the last two symmetries shows that $P(\hat{\xi}' \rightarrow \hat{\xi}) = P(-\hat{\xi} \rightarrow -\hat{\xi}')$, a requirement for the validity of the reciprocity principle. In addition, the fact that Θ depends on $\phi - \phi'$ only through $\cos(\phi - \phi')$ implies that it can be expanded in a Fourier cosine series³¹ in $\phi - \phi'$, i.e.,

$$P(u' \rightarrow u, \phi' \rightarrow \phi) = p_0(u' \rightarrow u) + 2 \sum_{m=1}^{\infty} p_m(u' \rightarrow u) \cos m(\phi - \phi'). \tag{2.65}$$

³¹Another way of justifying this is that Eq. (2.64) shows that Θ is an even function of $\phi - \phi'$, while terms like $\sin m(\phi - \phi')$ that might occur in the Fourier series are odd functions of $\phi - \phi'$.

For the Rayleigh scattering phase function, Eq. (2.45), this series is easy to construct as follows.

$$\begin{aligned}
 P(\Theta) &= \frac{3}{4} [1 + \cos^2 \Theta] \\
 &= \frac{3}{4} \left[1 + \left(uu' + \sqrt{(1-u^2)(1-u'^2)} \cos(\phi - \phi') \right)^2 \right] \\
 &= \frac{3}{4} \left[1 + u^2 u'^2 + (1-u^2)(1-u'^2) \cos^2(\phi - \phi') + 2uu' \sqrt{(1-u^2)(1-u'^2)} \cos(\phi - \phi') \right].
 \end{aligned}$$

Noting that $\cos^2 x = (1 + \cos 2x)/2$, we get

$$\begin{aligned}
 P(\Theta) &= \frac{3}{4} \left(1 + u^2 u'^2 + \frac{1}{2} (1-u^2)(1-u'^2) \right) \\
 &\quad + \frac{3}{2} uu' \sqrt{(1-u^2)(1-u'^2)} \cos(\phi - \phi') \\
 &\quad + \frac{3}{8} (1-u^2)(1-u'^2) \cos 2(\phi - \phi').
 \end{aligned} \tag{2.66}$$

The Fourier coefficients in the expansion can then be read off directly and the series terminates after 3 terms, i.e., $p_m = 0$ for $m > 2$.

In most cases, a simple analytical expression for $P(u' \rightarrow u, \phi' \rightarrow \phi)$ is not available; however, we can use standard techniques of Fourier analysis to find the p_m 's. First, note that³²

$$\begin{aligned}
 \int_0^{2\pi} \cos m(\phi - \phi_1) \cos l(\phi - \phi_0) d\phi &= \pi \delta_{ml} \cos m(\phi_1 - \phi_0), & m \neq 0 \\
 &= 2\pi \delta_{ml}, & m = 0, \\
 \int_0^{2\pi} \sin m(\phi - \phi_1) \sin l(\phi - \phi_0) d\phi &= \pi \delta_{ml} \cos m(\phi_1 - \phi_0), & m \neq 0 \\
 &= 0, & m = 0, \\
 \int_0^{2\pi} \cos m(\phi - \phi_1) \sin l(\phi - \phi_0) d\phi &= \pi \delta_{ml} \sin m(\phi_1 - \phi_0), & m \neq 0 \\
 &= 0, & m = 0,
 \end{aligned} \tag{2.67}$$

where ϕ_0 and ϕ_1 are constants. Second, multiply Eq. (2.65) by $\cos l(\phi - \phi')$ and integrate over all ϕ (or over all ϕ'). Then using the first of Eqs. (2.67), we find

$$p_l(u' \rightarrow u) = \frac{1}{2\pi} \int_0^{2\pi} P(u' \rightarrow u, \phi' \rightarrow \phi) \cos l(\phi - \phi') d\phi = \frac{1}{2\pi} \int_0^{2\pi} P(u' \rightarrow u, \phi' \rightarrow \phi) \cos l(\phi - \phi') d\phi'.$$

³²The function δ_{ij} is called the Kronecker Delta. Its properties are: $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$.

The motivation for expanding the phase function in a Fourier series in azimuth lies in the simplification that results if a similar expansion is made for the radiance. Consider

$$L(\tau, u, \phi) = L_0(\tau, u) + 2 \sum_{m=1}^{\infty} L_m^{(c)}(\tau, u) \cos m(\phi - \phi_1) + 2 \sum_{m=1}^{\infty} L_m^{(s)}(\tau, u) \sin m(\phi - \phi_1), \quad (2.68)$$

where

$$L_m^{(c)}(\tau, u) = \frac{1}{2\pi} \int_0^{2\pi} L(\tau, u, \phi) \cos m(\phi - \phi_1) d\phi,$$

$$L_m^{(s)}(\tau, u) = \frac{1}{2\pi} \int_0^{2\pi} L(\tau, u, \phi) \sin m(\phi - \phi_1) d\phi,$$

and ϕ_1 is a constant reference angle. Inserting the Fourier series into the 1-d RTE, Eq. (2.46) with $Q = 0$ yields

$$\begin{aligned} \left[u \frac{\partial}{\partial \tau} + 1 \right] & \left[L_0(\tau, u) + 2 \sum_{m=1}^{\infty} L_m^{(c)}(\tau, u) \cos m(\phi - \phi_1) + 2 \sum_{m=1}^{\infty} L_m^{(s)}(\tau, u) \sin m(\phi - \phi_1) \right] \\ &= \frac{\omega_0(\tau)}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \left[p_0(u' \rightarrow u) + 2 \sum_{m=1}^{\infty} p_m(u' \rightarrow u) \cos m(\phi - \phi') \right] \\ & \times \left[L_0(\tau, u') + 2 \sum_{m=1}^{\infty} L_m^{(c)}(\tau, u') \cos m(\phi' - \phi_1) + 2 \sum_{m=1}^{\infty} L_m^{(s)}(\tau, u') \sin m(\phi' - \phi_1) \right]. \end{aligned}$$

The integral over ϕ' can now be evaluated, and the right-hand-side becomes

$$\begin{aligned} & \frac{\omega_0(\tau)}{4\pi} \int_{-1}^1 du' \left[2\pi p_0(u' \rightarrow u) L_0(\tau, u') \right. \\ & \left. + 4\pi \sum_{m=1}^{\infty} p_m(u' \rightarrow u) \left(L_m^{(c)}(\tau, u') \cos m(\phi - \phi_1) + L_m^{(s)}(\tau, u') \sin m(\phi - \phi_1) \right) \right]. \end{aligned}$$

Now, multiply the full equation by $\cos n(\phi - \phi_1) d\phi$, integrate over ϕ from 0 to 2π , and use Eqs. (2.67). This results in

$$\begin{aligned} \left[u \frac{\partial}{\partial \tau} + 1 \right] L_m^{(c)}(\tau, u) &= \frac{\omega_0}{2} \int_{-1}^1 du' p_m(u' \rightarrow u) L_m^{(c)}(\tau, u'), \\ \left[u \frac{\partial}{\partial \tau} + 1 \right] L_m^{(s)}(\tau, u) &= \frac{\omega_0}{2} \int_{-1}^1 du' p_m(u' \rightarrow u) L_m^{(s)}(\tau, u'), \end{aligned} \quad (2.69)$$

where the second equation is derived by the same procedure, but replacing cosine by sine in the last step. Performing the same analysis on the boundary conditions, yields for a slab

$$L(0, u, \phi) = L_0(0, u) + 2 \sum_{m=1}^{\infty} L_m^{(c)}(0, u) \cos m(\phi - \phi_1) + 2 \sum_{m=1}^{\infty} L_m^{(s)}(0, u) \sin m(\phi - \phi_1), \quad u > 0,$$

and

$$L(\tau_1, u, \phi) = L_0(\tau_1, u) + 2 \sum_{m=1}^{\infty} L_m^{(c)}(\tau_1, u) \cos m(\phi - \phi_1) + 2 \sum_{m=1}^{\infty} L_m^{(s)}(\tau_1, u) \sin m(\phi - \phi_1), \quad u < 0,$$

where

$$L_m^{(c)}(0, u) = \frac{1}{2\pi} \int_0^{2\pi} L(0, u, \phi) \cos m(\phi - \phi_1) d\phi,$$

etc. These provide the boundary conditions for each Fourier component $L_m^{(c)}(\tau, u)$ and $L_m^{(s)}(\tau, u)$ of the radiance. Thus, each Fourier component of the radiance satisfies a radiative transfer equation in which the scattering term contains only the Fourier component of the phase function with the same Fourier order. The associated boundary conditions are the Fourier components of the boundary conditions (of the same Fourier order). Immediate consequences of this are

- if $L_m^{(i)}(0, u) = 0$ for $u > 0$ and $L_m^{(i)}(\tau_1, u) = 0$ for $u < 0$, then $L_m^{(i)}(\tau, u) = 0$, where $i = c$ or s ,³³
- if $p_m(u' \rightarrow u) = 0$, then $L_m^{(i)}(\tau, u) = L_m^{(i)}(0, u) \exp[-\tau/u]$ for $u > 0$ and $L_m^{(i)}(\tau, u) = L_m^{(i)}(\tau_1, u) \exp[-(\tau - \tau_1)/u]$ for $u < 0$.

An additional simplification takes place in the case of a slab illuminated by a beam source beam source, e.g., the Sun. A beam source has the property that the radiance is confined to small solid angle, e.g., $\Delta\Omega = \Delta u \Delta\phi$. Assume that the radiance incident on the top of the slab is $L(0, u, \phi) = L_{\text{Sun}}(u)$ for ϕ in the range $\phi_0 - \Delta\phi/2$ to $\phi_0 + \Delta\phi/2$. Then it is easy to show that, if ϕ_1 above is taken to be ϕ_0 , i.e., the reference azimuth is taken to be the azimuth of the beam,

$$L_m^{(c)}(0, u) = \frac{\cos m\Delta\phi}{2\pi m} L_{\text{Sun}}(u) \quad \text{and} \quad L_m^{(s)}(0, u) = 0,$$

and so $L_m^{(s)}(\tau, u) = 0$. In this case the radiance is symmetric with respect to ϕ_0 , i.e.,

$$L(\tau, u, \phi - \phi_0) = L(\tau, u, -\phi + \phi_0).$$

Finally, when the Fourier expansion of the radiances is carried out, the irradiances in

³³The easiest way to show this is to note that if there are no internal sources of a particular Fourier component and if no radiance for that component is incident on the boundary, then applying the successive-order of scattering method to Eq. (2.69) yields zero internal radiance as well for that component.

Eq. (2.63) take a particularly simple form:

$$\begin{aligned}
 E_d(z, \lambda) &= 2\pi \int_0^1 u L_0(z, u, \lambda) du, \\
 E_u(z, \lambda) &= -2\pi \int_{-1}^0 u L_0(z, u, \lambda) du, \\
 E_d(z, \lambda) - E_u(z, \lambda) &= 2\pi \int_{-1}^1 u L_0(z, u, \lambda) du, \\
 E_0(z, \lambda) &= 2\pi \int_{-1}^1 L_0(z, u, \lambda) du.
 \end{aligned} \tag{2.70}$$

Thus, the irradiances depend only on the lowest order Fourier coefficient L_0 , and if only irradiances are required, only L_0 need be computed.

2.7.1.3 Gaussian Quadratures

It is straightforward to show that the individual RTE's for the Fourier components can be solved using the successive-order-of-scattering technique, however, one must still evaluate the integral of the appropriate phase function times the appropriate radiance (in one order lower), i.e., an integral of the form

$$\int_{-1}^1 du' p(u' \rightarrow u) L(\tau, u'),$$

where p and L are Fourier components. For completeness, we explain how this is effected in a typical radiative transfer computation. The integral is always evaluated through a sum. The most efficient way of doing this is to use *Gaussian quadratures*. In this case,

$$\int_{-1}^1 du' p(u' \rightarrow u) L(\tau, u') = \sum_{i=-n}^n a_i p(u_i \rightarrow u) L(\tau, u_i),$$

where the *division* points or *quadrature* points u_i are determined from $P_{2n}(u) = 0$, where P_{2n} is the Legendré polynomial of order $2n$.³⁴ It can be shown that the *weights* a_i can be chosen so that the sum is exactly equal to the integral if the integrand $p(u' \rightarrow u) L(\tau, u')$ is

³⁴The Legendré polynomials are given by

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} [(1-x^2)^n].$$

They are polynomials that are even or odd functions of x as n is even or odd. All are of the form $P_n(x) = a_n x^n + a_{n-2} x^{n-2} + \dots$.

a polynomial of order $2n - 1$ or less.³⁵ Since $2n$ is even, $u_i = -u_{-i}$, and $a_i = a_{-i}$. Tables of u_i and a_i for various n 's are readily available.

Other methods are obviously available for turning the integral into a sum. For example, one could take $n + 1$ values of u separated by equal increments of Δu and employ the trapezoidal rule,

$$\int_{-1}^1 du' p(u' \rightarrow u) L(\tau, u') = \sum_{i=0}^n a_i p(u_i \rightarrow u) L(\tau, u_i),$$

where $a_0 = a_n = \Delta u/2$ and $a_i = \Delta u$ for $i \neq 0$ or n . However, regardless of how the discretization is accomplished, the result is always the sum of the product of the integrand and weights at the division points. Gaussian quadratures provide the maximum accuracy for a given number of division points.³⁶

2.7.2 The Discrete-Ordinate Method

The discrete ordinate method uses Fourier analysis in azimuth and Gaussian quadratures in the polar direction to treat the integral term in the RTE. Consider a homogeneous slab as before, i.e., Figure 2.12, and a particular Fourier coefficient of the radiance. The appropriate RTE for say $L_m^{(c)}$ is

$$\left[u \frac{\partial}{\partial \tau} + 1 \right] L_m^{(c)}(\tau, u) = \frac{\omega_0}{2} \int_{-1}^1 du' p_m(u' \rightarrow u) L_m^{(c)}(\tau, u').$$

We can rewrite the integral as a sum using Gaussian quadratures to give

$$\left[u \frac{\partial}{\partial \tau} + 1 \right] L_m^{(c)}(\tau, u) = \frac{\omega_0}{2} \sum_{j=-n}^{j=n} a_j p_m(u_j \rightarrow u) L_m^{(c)}(\tau, u_j).$$

³⁵The weights are given by

$$a_j = \frac{2}{(1 - u_j^2)[P'_n(u_j)]^2},$$

where $P'_n(x) = dP_n(x)/dx$.

³⁶It can be shown that, if $f(x)$ can be represented by a polynomial of degree $4n - 1$ (or less), then

$$\int_{-1}^1 f(x) dx = \sum_{i=-n}^{i=+n} a_i f(x_i), \quad \text{exactly.}$$

This equation holds for all u , and therefore must hold for $u = u_i$, i.e.,³⁷

$$\left[u_i \frac{d}{d\tau} + 1 \right] L_m^{(c)}(\tau, u_i) = \frac{\omega_0}{2} \sum_{j=-n}^{j=n} a_j p_m(u_j \rightarrow u_i) L_m^{(c)}(\tau, u_j). \quad (2.71)$$

We will now drop the superscript “ c ” and subscript “ m ” on L for economy of notation, and write

$$\mathbf{L} = \begin{pmatrix} L_n \\ \vdots \\ L_1 \\ L_{-n} \\ \vdots \\ L_{-1} \end{pmatrix},$$

where $L_i = L_m^{(c)}(\tau, u_i)$, etc. This is a set of coupled linear differential equations that can be written

$$\frac{d\mathbf{L}}{d\tau} + \mathbf{S}\mathbf{L} = 0,$$

where the square matrix \mathbf{S} has elements

$$S_{i,j} = \frac{1}{u_i} \left[\delta_{ij} - \frac{\omega_0}{2} a_j p_m(u_j \rightarrow u_i) \right],$$

and so

$$S_{i,j} = -S_{-i,-j} \quad \text{and} \quad S_{-i,j} = -S_{i,-j}.$$

This can be solved by letting

$$\mathbf{L} = \mathbf{C} \exp[k\tau], \quad (2.72)$$

where \mathbf{C} is a constant vector; however, it is more illuminating to make a change that exploits the symmetry of \mathbf{S} . We let

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}^+ \\ \mathbf{L}^- \end{pmatrix}, \quad \text{where} \quad \mathbf{L}^+ = \begin{pmatrix} L_n \\ L_{n-1} \\ \vdots \\ L_1 \end{pmatrix}, \quad \text{and} \quad \mathbf{L}^- = \begin{pmatrix} L_{-n} \\ L_{-n+1} \\ \vdots \\ L_{-1} \end{pmatrix}$$

Then the RTE can be written

$$\frac{d\mathbf{L}^+}{d\tau} + \mathbf{S}^{++}\mathbf{L}^+ + \mathbf{S}^{-+}\mathbf{L}^- = 0$$

³⁷Here we have replaced $\partial L_m^{(c)}/\partial \tau$ by $dL_m^{(c)}/d\tau$ because $L_m^{(c)}(\tau, u_i)$ now depends on only one variable: τ .

and

$$\frac{d\mathbf{L}^-}{d\tau} + \mathbf{S}^{+-}\mathbf{L}^+ + \mathbf{S}^{--}\mathbf{L}^- = 0,$$

where the \mathbf{S} 's can be read directly from \mathbf{S} :

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}^{++} & \mathbf{S}^{-+} \\ \mathbf{S}^{+-} & \mathbf{S}^{--} \end{pmatrix}$$

Now, the symmetry of \mathbf{S} demands that

$$\mathbf{S}^{--} = -\mathbf{S}^{++} \quad \text{and} \quad \mathbf{S}^{+-} = -\mathbf{S}^{-+},$$

so

$$\frac{d\mathbf{L}^+}{d\tau} + \mathbf{S}^{++}\mathbf{L}^+ + \mathbf{S}^{-+}\mathbf{L}^- = 0$$

and

$$\frac{d\mathbf{L}^-}{d\tau} - \mathbf{S}^{-+}\mathbf{L}^+ - \mathbf{S}^{++}\mathbf{L}^- = 0.$$

Now, let

$$\mathbf{L}^+ = \mathbf{C}^+ \exp[k\tau] \quad \text{and} \quad \mathbf{L}^- = \mathbf{C}^- \exp[k\tau],$$

where

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}^+ \\ \mathbf{C}^- \end{pmatrix}$$

and insert these into the above equations yielding

$$\begin{aligned} \mathbf{C}^+ + \mathbf{S}^{++}\mathbf{C}^+ + \mathbf{S}^{-+}\mathbf{C}^- &= 0 \\ k\mathbf{C}^- - \mathbf{S}^{-+}\mathbf{C}^+ - \mathbf{S}^{++}\mathbf{C}^- &= 0. \end{aligned}$$

Adding and subtracting these it is easy to show that

$$k^2 \mathbf{I} - (\mathbf{S}^{++} - \mathbf{S}^{-+})(\mathbf{S}^{++} + \mathbf{S}^{-+}) = 0 \quad (2.73)$$

where \mathbf{I} is the unit matrix. This is called the eigenvalue equation and provides the allowable values of k so that Eq. (2.72) is a solution of Eq. (2.71).³⁸ There are $2n$ such values of k (*eigenvalues*), which we label $\pm k_1, \pm k_1, \dots, \pm k_n$. If we put the individual eigenvalues back into Eq. (2.73) all of the $2n$ \mathbf{C} 's (*eigenvectors*) can be found individually; however, since Eq. (2.73) is homogeneous, the components of the individual \mathbf{C} 's can only be found to within an arbitrary constant. That constant can be determined by imposing the condition that \mathbf{C} be normalized, i.e., $\tilde{\mathbf{C}}\mathbf{C} = 1$. Then, the general solution is a linear combination of the individual solutions

$$\mathbf{L} = \sum_{j=1}^n \left[\alpha_{+j} \mathbf{C}_{+j} \exp[k_j \tau] + \alpha_{-j} \mathbf{C}_{-j} \exp[-k_j \tau] \right]. \quad (2.74)$$

³⁸Eigenvalue problems are discussed in the Mathematical Appendix, Section 14.5.3.

The $2n$ arbitrary constants $\alpha_{\pm j}$ are then determined using the boundary conditions — the \mathbf{C} 's are known at this point. In the case of a slab, we are given \mathbf{L}^+ for $\tau = 0$ and \mathbf{L}^- for $\tau = \tau_1$. If the incident radiance on the slab is along one of the u_i 's, e.g., along u_I , and no radiance incident on the bottom, we have

$$L(0, u_i) = L(0, u_I)\delta_{iI}, \quad i = 1 \rightarrow n \quad \text{and} \quad L(\tau_1, u_i) = 0, \quad i = -1 \rightarrow -n.$$

It should be recalled that this solution is for a particular Fourier coefficient in the expansion of the radiance, and the above procedure must be carried out for each Fourier coefficient. We note that since the α_i are linear combinations of the incident boundary radiances, Eq. (2.74) shows that the radiances exiting the slab depend *linearly* on the radiances incident on the slab. This fact will be used later in discussing the *doubling or adding* method.

We note that in the case of a slab illuminated by a beam of radiance, the discrete ordinate method seems to require that the incident beam be along one of the u_i 's, the positions of which depend on the order of approximation. This is of no real consequence, as we can choose n to be as large as is required to put the beam as close to a quadrature point as desired. However, there is another approach. Since the incident radiance in the case of a beam source is

$$L(0, u, \phi) = F_0\delta(u - u_0)\delta(\phi - \phi_0)$$

we can include its influence as an internal source as in Eq. (2.38). Then we need to solve, not Eq. (2.71) but,

$$\left[u_i \frac{\partial}{\partial \tau} + 1 \right] L_m^{(c)}(\tau, u_i) = \frac{\omega_0}{2} \sum_{j=-n}^{j=n} a_j p_m(u_j \rightarrow u_i) L_m^{(c)}(\tau, u_j) + \frac{\omega_0}{2} F_0 p_m(u_0 \rightarrow u) \exp \left[-\frac{\tau}{u_0} \right], \quad (2.75)$$

subject to the boundary radiances $\mathbf{L}^+ = 0$ at $\tau = 0$ and $\mathbf{L}^- = 0$ at $\tau = \tau_1$. This requires finding a single solution \mathbf{L}_p (the *particular* solution) of this inhomogeneous equation and adding it to the general solution of the homogeneous RTE given in Eq. (2.74), i.e.,

$$\mathbf{L} = \sum_{j=1}^n \left[\alpha_{+j} \mathbf{C}_{+j} \exp[k_j \tau] + \alpha_{-j} \mathbf{C}_{-j} \exp[-k_j \tau] \right] + \mathbf{L}_p.$$

The boundary radiances, $\mathbf{L}^+ = 0$ at $\tau = 0$ and $\mathbf{L}^- = 0$ at $\tau = \tau_1$, are then applied to this equation to effect the solution. Note there is now no requirement that u_0 equal any u_i .

A interesting conclusion can be drawn from Eq. (2.74) in the case of a semi-infinite medium illuminated from the top, e.g., the ocean illuminated at the surface. To prevent the radiance from becoming infinite as $\tau \rightarrow \infty$, we must require that $\alpha_{+j} = 0$, so

$$\mathbf{L} = \sum_{i=j}^n \alpha_{-j} \mathbf{C}_{-j} \exp[-k_j \tau].$$

Now, at great depth one term in this equation is dominant, the one with the smallest $|k_i|$, call it k_s . Thus, as $\tau \rightarrow \infty$

$$\mathbf{L} \rightarrow \alpha_{-s} \mathbf{C}_{-s} \exp[-k_s \tau] \quad (2.76)$$

and the angular distribution of the radiance for this Fourier coefficient is given by a single eigenvector \mathbf{C}_{-s} corresponding to the smallest $|k_i|$. Thus, deep within a homogeneous medium $L_m^{(c)}(\tau, u)$ becomes independent of the boundary radiance $L_m^{(c)}(0, u)$ and is characteristic of the medium. This is of course true for all Fourier coefficients; however, of these only one can be dominant for $\tau \rightarrow \infty$, the one with the smallest of the k_s 's. So, in fact, deep within a medium the angular distribution of radiance must become independent of depth. It is clear on physical grounds that the $m = 0$ Fourier coefficient must be the one that remains,³⁹ i.e., the $m = 0$ coefficient has the smallest $|k_s|$. This is called the *asymptotic radiance distribution*. This distribution can only depend on ω_0 and $p_0(u' \rightarrow u)$ and Eq. (2.76) shows that it can be written in the form

$$L_\infty(\tau, u) = U(u) \exp[-K\tau]. \quad (2.77)$$

Also, $L_\infty(\tau, u)$ must satisfy the homogeneous RTE with $m = 0$,

$$\left[u \frac{\partial}{\partial \tau} + 1 \right] L_0(\tau, u) = \frac{\omega_0}{2} \int_{-1}^1 du' p_0(u' \rightarrow u) L_0(\tau, u'),$$

and inserting Eq. (2.77) yields

$$[1 - Ku]U(u) = \frac{\omega_0}{2} \int_{-1}^1 du' p_0(u' \rightarrow u) U(u'). \quad (2.78)$$

from which K and U can be found. Equation (2.78) provides the form of the radiance and its decay with depth in a 1-d medium far from the boundary. These are characteristic of ω_0 and the phase function. Equation (2.78) can be solved numerically for any phase function, and analytically for simple very simple phase functions, e.g., isotropic or Rayleigh scattering.⁴⁰

³⁹As one proceeds into the medium, the radiance distribution must become smoother because of scattering. Therefore higher-order Fourier components of L must decay more rapidly than lower-order Fourier components.

⁴⁰In the case of isotropic scattering ($p_0 = 1$), the solution is very easy. For $p_0 = 1$, the right-hand side of Eq. (2.78) is a constant, call it " A ." Then $U = A/(1 - Ku)$. Inserting this into Eq. (2.78) and removing A yields:

$$1 = \frac{\omega_0}{2} \int_{-1}^1 \frac{du}{1 - Ku} = \frac{\omega_0}{2K} \ln \left[\frac{1 + K}{1 - K} \right],$$

which provides K , given ω_0 . The radiance is then

$$L(\tau, u) = C \frac{\exp(-K\tau)}{1 - Ku} \quad \text{as } \tau \rightarrow \infty.$$

, where C is an arbitrary constant.

2.7.3 Matrix Operator, Doubling, or Adding Method

In discussing these methods, for simplicity we restrict our attention to Eq. (2.71) rather than Eq. (2.75). These methods are all based on the observation in the last section that the solution, Eq. (2.74), of Eq. (2.71) subject to the boundary conditions depends *linearly* on the boundary radiances, i.e., if we consider a slab from τ_a to τ_b , with $\tau_b > \tau_a$, the radiances $\mathbf{L}^+(\tau_b)$ and $\mathbf{L}^-(\tau_a)$ for a given Fourier component, exiting the the bottom and the top of the slab, respectively, are

$$\begin{aligned}\mathbf{L}^+(\tau_b) &= \mathbf{t}_{ab}\mathbf{L}^+(\tau_a) + \mathbf{r}_{ba}\mathbf{L}^-(\tau_b) \\ \mathbf{L}^-(\tau_a) &= \mathbf{r}_{ab}\mathbf{L}^+(\tau_a) + \mathbf{t}_{ba}\mathbf{L}^-(\tau_b)\end{aligned}$$

where matrices \mathbf{r} and \mathbf{t} are, respectively, the $n \times n$ reflectance and transmittance matrices for the layer with the first element of the subscript denoting the side of the slab on which the radiance is incident, e.g., \mathbf{t}_{ba} is the transmittance matrix for radiance propagating through the slab from τ_b to τ_a , etc., and as in the discrete ordinate method,

$$\mathbf{L}^+(\tau) = \begin{pmatrix} L(\tau, u_n) \\ L(\tau, u_{n-1}) \\ \vdots \\ L(\tau, u_1) \end{pmatrix}, \quad \text{and} \quad \mathbf{L}^-(\tau) = \begin{pmatrix} L(\tau, u_{-n}) \\ L(\tau, u_{-n+1}) \\ \vdots \\ L(\tau, u_{-1}) \end{pmatrix}.$$

If the layer is homogeneous, $\mathbf{r}_{ab} = \mathbf{r}_{ba}$ and $\mathbf{t}_{ab} = \mathbf{t}_{ba}$, because it is irrelevant if such a layer is illuminated from above or below. Consider a second slab from τ_b to τ_c , with $\tau_c > \tau_b$, then

$$\begin{aligned}\mathbf{L}^+(\tau_c) &= \mathbf{t}_{bc}\mathbf{L}^+(\tau_b) + \mathbf{r}_{cb}\mathbf{L}^-(\tau_c) \\ \mathbf{L}^-(\tau_b) &= \mathbf{r}_{bc}\mathbf{L}^+(\tau_b) + \mathbf{t}_{cb}\mathbf{L}^-(\tau_c)\end{aligned}$$

Now for the combined layer from τ_a to τ_c we have

$$\begin{aligned}\mathbf{L}^+(\tau_c) &= \mathbf{t}_{ac}\mathbf{L}^+(\tau_a) + \mathbf{r}_{ca}\mathbf{L}^-(\tau_c) \\ \mathbf{L}^-(\tau_a) &= \mathbf{r}_{ac}\mathbf{L}^+(\tau_a) + \mathbf{t}_{ca}\mathbf{L}^-(\tau_c)\end{aligned}$$

If we eliminate $\mathbf{L}^\pm(\tau_b)$ from these expressions, we can relate the \mathbf{r} 's and \mathbf{t} 's for the combined layer in terms of those for the individual layers. The result is

$$\begin{aligned}\mathbf{t}_{ac} &= \mathbf{t}_{bc}(\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})^{-1}\mathbf{t}_{ab}, \\ \mathbf{r}_{ca} &= \mathbf{r}_{cb} + \mathbf{t}_{bc}(\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})^{-1}\mathbf{r}_{ba}\mathbf{t}_{cb}, \\ \mathbf{t}_{ca} &= \mathbf{t}_{ba}(\mathbf{I} - \mathbf{r}_{bc}\mathbf{r}_{ba})^{-1}\mathbf{t}_{cb}, \\ \mathbf{r}_{ac} &= \mathbf{r}_{ab} + \mathbf{t}_{ba}(\mathbf{I} - \mathbf{r}_{bc}\mathbf{r}_{ba})^{-1}\mathbf{r}_{bc}\mathbf{t}_{ab},\end{aligned}\tag{2.79}$$

where \mathbf{I} is the unit matrix, and $(\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})^{-1}$ is the inverse of $(\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})$. Thus, if we know the \mathbf{r} and \mathbf{t} matrices for two slabs, we can use Eq. (2.79) to find the \mathbf{r} and \mathbf{t}

matrices for the combined (thicker) slab. As before, these equations apply to each Fourier component individually. It should be noted that these equations provide more than just the radiances exiting the slab. It is easy to show that, given the boundary radiances into the combined slab, i.e., $\mathbf{L}^+(\tau_a)$ and $\mathbf{L}^-(\tau_c)$, one can manipulate the set of equations to find the interior radiances $\mathbf{L}^+(\tau_b)$ and $\mathbf{L}^-(\tau_b)$ as well:

$$\begin{aligned}\mathbf{L}^+(\tau_b) &= (\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})^{-1} [\mathbf{t}_{ab}\mathbf{L}^+(\tau_a) + \mathbf{r}_{ba}\mathbf{t}_{cb}\mathbf{L}^-(\tau_c)], \\ \mathbf{L}^-(\tau_b) &= (\mathbf{I} - \mathbf{r}_{bc}\mathbf{r}_{ba})^{-1} [\mathbf{r}_{bc}\mathbf{t}_{ab}\mathbf{L}^+(\tau_a) + \mathbf{t}_{cb}\mathbf{L}^-(\tau_c)].\end{aligned}$$

Equations (2.79) are easy to understand, if $(\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})^{-1}$ is expanded using the binomial theorem:

$$(\mathbf{I} - \mathbf{r}_{ba}\mathbf{r}_{bc})^{-1} = \mathbf{I} + \mathbf{r}_{ba}\mathbf{r}_{bc} + \mathbf{r}_{ba}\mathbf{r}_{bc}\mathbf{r}_{ba}\mathbf{r}_{bc} + \mathbf{r}_{ba}\mathbf{r}_{bc}\mathbf{r}_{ba}\mathbf{r}_{bc}\mathbf{r}_{ba}\mathbf{r}_{bc} + \cdots.$$

The term $\mathbf{r}_{ba}\mathbf{r}_{bc}$ represents reflection from the slab bc toward the slab ab then from the slab ab back toward the slab bc . Each subsequent term in the sum represents this process multiple times. So in the last equation in Eq. (2.79) the first term represents direct reflection of radiance from slab ab , while the second term represents reflection from slab bc including all orders of multiple reflections between the two slabs.

Equations (2.79) represent basis of the *Matrix Operator* method for solving the RTE. Their value lies in the fact that for very thin layers analytical expressions can be developed for these matrices by assuming that photons can scatter only once. Thus, if we want to solve the RTE for a thick homogeneous slab, we need only use single scattering to find \mathbf{r} and \mathbf{t} for a thin slab, then combine two such slabs to find \mathbf{r} and \mathbf{t} for a slab twice as thick. Combining two of the latter slabs results in a still thicker slab, etc. Thus, at each step in the process, the thickness of the slab for which the RTE is solved is doubled. In this manner \mathbf{r} and \mathbf{t} for homogeneous slabs of any desired thickness can be generated. This is referred to as the *doubling* method. Since an inhomogeneous medium can be approximated by homogeneous slabs, and \mathbf{r} and \mathbf{t} for each slab can be generated through doubling, the \mathbf{r} and \mathbf{t} for an inhomogeneous medium can be generated by combining such homogeneous slabs. This is sometimes called the *adding* method.

All of the methods described in the previous paragraph start with \mathbf{t} and \mathbf{r} for a thin layer in which single scattering is extant. But before developing the formulae for such a layer, it is desirable to make a small notational change that will aid in understanding some of the equations. We shall let $\mu_i = |u_i|$ in which case

$$\mathbf{L}^+(\tau) = \begin{pmatrix} L(\tau, \mu_n) \\ L(\tau, \mu_{n-1}) \\ \vdots \\ L(\tau, \mu_1) \end{pmatrix}, \quad \text{and} \quad \mathbf{L}^-(\tau) = \begin{pmatrix} L(\tau, -\mu_n) \\ L(\tau, -\mu_{n-1}) \\ \vdots \\ L(\tau, -\mu_1) \end{pmatrix}.$$

In this way, the direction (upward or downward) is readily apparent from the sign attached to μ_i . Then using the results of the single-scattering analysis in Appendix 1 in this chapter, we have, for a homogeneous layer of thickness $\Delta\tau$, after replacing the u 's by μ 's,

$$\begin{aligned}
 {}_{ij} &= \frac{\omega_0 \Delta\tau}{2\mu_i} a_j p_m(\mu_j \rightarrow -\mu_i), \\
 [\mathbf{r}_{ba}(\Delta\tau)]_{ij} &= \frac{\omega_0 \Delta\tau}{2\mu_i} a_j p_m(-\mu_j \rightarrow \mu_i), \\
 [\mathbf{t}_{ab}(\Delta\tau)]_{ij} &= \left[1 - \frac{\Delta\tau}{\mu_i}\right] \delta_{ij} + \frac{\omega_0 \Delta\tau}{2\mu_i} a_j p_m(\mu_j \rightarrow \mu_i), \\
 [\mathbf{t}_{ba}(\Delta\tau)]_{ij} &= \left[1 - \frac{\Delta\tau}{\mu_i}\right] \delta_{ij} + \frac{\omega_0 \Delta\tau}{2\mu_i} a_j p_m(-\mu_j \rightarrow -\mu_i),
 \end{aligned} \tag{2.80}$$

Note that the signs attached to μ_i and μ_j are the same in the \mathbf{t} 's but opposite in the \mathbf{r} 's. The symmetry properties of the phase function imply

$$\begin{aligned}
 [\mathbf{r}_{ab}(\Delta\tau)]_{ij} &= [\mathbf{r}_{ba}(\Delta\tau)]_{ij}, \\
 [\mathbf{t}_{ab}(\Delta\tau)]_{ij} &= [\mathbf{t}_{ba}(\Delta\tau)]_{ij}.
 \end{aligned} \tag{2.81}$$

This symmetry does *not* apply to inhomogeneous media of finite optical thickness.

We have developed these matrix methods by starting with Eq. (2.71) and assuming that in the case of a beam source the beam was directed along a single quadrature direction. Similar equations can be developed starting from Eq. (2.75), and the resulting equations for \mathbf{r} and \mathbf{t} also include a source term for each layer. As the goal of this discussion is to acquaint the reader with the “flavor” of the various methods of solving the RTE, as opposed to all of the technical details, we shall not develop such equations here.

2.7.4 Invariant Imbedding

The final method that we shall describe for solving the RTE is called *invariant imbedding*. It can be developed from the matrix operator method in a simple manner. Consider a 1-d medium consisting of a slab of finite optical thickness T . If we add a second slab on the *top* of the first slab, then the reflection and transmission matrices for the combined slab are given by Eq. (2.79). Let the upper slab have a small (infinitesimal) optical thickness ΔT . Then using the last of Eq. (2.79), we have

$$\begin{aligned}
 \mathbf{r}_{ac}(T + \Delta T) &= \mathbf{r}_{ab}(\Delta T) + \mathbf{t}_{ba}(\Delta T)(\mathbf{I} - \mathbf{r}_{bc}(T)\mathbf{r}_{ba}(\Delta T))^{-1}\mathbf{r}_{bc}(T)\mathbf{t}_{ab}(\Delta T) \\
 &= \mathbf{r}_{ab}(\Delta T) + \mathbf{t}_{ba}(\Delta T)\mathbf{r}_{bc}(T)\mathbf{t}_{ab}(\Delta T) \\
 &\quad + \mathbf{t}_{ba}(\Delta T)\mathbf{r}_{bc}(T)\mathbf{r}_{ba}(\Delta T)\mathbf{r}_{bc}(T)\mathbf{t}_{ab}(\Delta T) + \cdots,
 \end{aligned} \tag{2.82}$$

where we have explicitly indicated the optical thicknesses associated with the various operators. Note that in the present notation, T is measured from the *bottom* of slab bc (which is why we changed the notation for optical depth from τ to T). Now, from Eq. (2.80) it is clear that we can write

$$\begin{aligned} \mathbf{r}_{ab}(\Delta T) &= \mathbf{m}_{ab}^r \times \Delta T, \\ \mathbf{r}_{ba}(\Delta T) &= \mathbf{m}_{ba}^r \times \Delta T, \\ \mathbf{t}_{ab}(\Delta T) &= \mathbf{I} + \mathbf{m}_{ab}^t \times \Delta T, \\ \mathbf{t}_{ba}(\Delta T) &= \mathbf{I} + \mathbf{m}_{ba}^t \times \Delta T, \end{aligned} \tag{2.83}$$

where the \mathbf{m} matrices are independent of (ΔT) , and can be directly read from Eq. (2.80). Thus, inserting Eq. (2.83) into Eq. (2.82) and retaining only first order terms in ΔT , we have

$$\mathbf{r}_{ac}(T + \Delta T) = \mathbf{m}_{ab}^r \Delta T + \mathbf{r}_{bc}(T) + \mathbf{m}_{ba}^t \mathbf{r}_{bc}(T) \Delta T + \mathbf{r}_{bc}(T) \mathbf{m}_{ab}^t \Delta T + \mathbf{r}_{bc}(T) \mathbf{m}_{ba}^r \mathbf{r}_{bc}(T) \Delta T.$$

Rearranging,

$$\frac{\mathbf{r}_{ac}(T + \Delta T) - \mathbf{r}_{bc}(T)}{\Delta T} = \mathbf{m}_{ab}^r + \mathbf{m}_{ba}^t \mathbf{r}_{bc}(T) + \mathbf{r}_{bc}(T) \mathbf{m}_{ab}^t + \mathbf{r}_{bc}(T) \mathbf{m}_{ba}^r \mathbf{r}_{bc}(T).$$

Now $\mathbf{r}_{bc}(T)$ is the reflectance $\mathbf{r}(T)$ of a slab of thickness T illuminated from the top (recall that the illumination is from b to c). We find it useful for clarity to change in notation here and write this as \mathbf{r}_T^\downarrow . In this notation, \mathbf{r}_T^\downarrow is the reflection matrix of the original slab for radiance incident from the top (in the direction of the arrow superscript), while $\mathbf{r}_{ac}(T + \Delta T)$ is the reflectance $\mathbf{r}_{T+\Delta T}^\downarrow$ of a similar slab that is thicker by ΔT . Thus, without confusion we can drop the ac and bc subscripts from \mathbf{r} , take the limit as $\Delta T \rightarrow 0$ and get

$$\frac{d\mathbf{r}_T^\downarrow}{dT} = \mathbf{m}_{ab}^r + \mathbf{m}_{ba}^t \mathbf{r}_T^\downarrow + \mathbf{r}_T^\downarrow \mathbf{m}_{ab}^t + \mathbf{r}_T^\downarrow \mathbf{m}_{ba}^r \mathbf{r}_T^\downarrow.$$

By using the symmetry properties of the phase function, we can also drop the subscripts on the \mathbf{m} 's (see, e.g., Eq. (??)). This gives

$$\frac{d\mathbf{r}_T^\downarrow}{dT} = \mathbf{m}^r + \mathbf{m}^t \mathbf{r}_T^\downarrow + \mathbf{r}_T^\downarrow \mathbf{m}^t + \mathbf{r}_T^\downarrow \mathbf{m}^r \mathbf{r}_T^\downarrow.$$

Note that nowhere in the analysis have we assumed that the \mathbf{m} 's are independent of optical depth. In fact, we only assumed that the infinitesimal layer of thickness ΔT was homogeneous, but then took the limit as $\Delta T \rightarrow 0$. Thus, the \mathbf{m} 's may vary with optical depth — through any dependence of ω_0 and p_m on optical depth — and it would be more proper to display this dependence explicitly and replace \mathbf{m}^r and \mathbf{m}^t by $\mathbf{m}^r(T)$ and $\mathbf{m}^t(T)$, respectively:

$$\frac{d\mathbf{r}_T^\downarrow}{dT} = \mathbf{m}^r(T) + \mathbf{m}^t(T) \mathbf{r}_T^\downarrow + \mathbf{r}_T^\downarrow \mathbf{m}^t(T) + \mathbf{r}_T^\downarrow \mathbf{m}^r(T) \mathbf{r}_T^\downarrow, \tag{2.84}$$

where $\mathbf{m}^r(T)$ and $\mathbf{m}^t(T)$ are evaluated at the top of the layer of thickness T . It is important to note that by virtue of the way Eq. (2.84) was derived, the zero for the variable T is at the *bottom* of the layer and increases *upward*. In our usual notation, the zero of optical depth τ is at the top of the medium and increases downward. Thus, for a layer from $\tau = \tau_1$ to $\tau = \tau_2$, ($\tau_2 > \tau_1$), the variable $T = \tau_2 - \tau$ and $dT = -d\tau$.

Equation (2.84) allows one to compute the reflectance matrix \mathbf{r}^\downarrow for a slab illuminated from the top in the following manner. First, start the bottom (a slab of zero thickness) with $\mathbf{r}_0^\downarrow = 0$. Next, integrate the equation upward using the known \mathbf{m}^t and \mathbf{m}^r matrices as a function of position in the medium. Finally, stop when the desired top of the slab is reached. Such a procedure is amenable to numerical computation.

Applying the same reasoning to the first of the four equations in Eq. (2.79) results in an equation for \mathbf{t}^\downarrow :

$$\frac{d\mathbf{t}_T^\downarrow}{dT} = \mathbf{t}_T^\downarrow \mathbf{m}^t(T) + \mathbf{t}_T^\downarrow \mathbf{m}^r(T) \mathbf{r}_T^\downarrow. \quad (2.85)$$

This equation can be integrated to find \mathbf{t}^\downarrow by (1) starting with at the bottom of the slab with $\mathbf{t}_0^\downarrow = \mathbf{I}$ and $\mathbf{r}_0^\downarrow = 0$, (2) integrating upward using the known \mathbf{m}^t and \mathbf{m}^r matrices as a function of position, and, after solving Eq. (2.84), the known \mathbf{r}_T^\downarrow matrices, and (3) stopping when the top of the slab is reached.

When the medium is homogeneous, $\mathbf{t}^\uparrow = \mathbf{t}^\downarrow$ and $\mathbf{r}^\uparrow = \mathbf{r}^\downarrow$; however, if it is inhomogeneous (i.e., vertically stratified) this is no longer true — think about reflection from a slab with a thin scattering layer above a thick absorbing layer and then reverse the order of the layers. For inhomogeneous media, the simplest approach to finding \mathbf{t}^\uparrow and \mathbf{r}^\uparrow is to invert the medium and compute \mathbf{t}^\downarrow and \mathbf{r}^\downarrow using Eq. s(2.84) and (2.85).

The reader may be wondering why this method goes by the name “invariant imbedding.” In this technique we have considered a slab to be imbedded in a medium, and asked how the reflection and transmission matrices vary if the slab’s *boundaries* are changed. Invariant comes from consideration of situations in which the reflectance and transmittance matrices cannot change. For example, consider a semi-infinite homogeneous medium. If a thin layer is added to the top of such a medium, it is clear at the reflectance matrix $\mathbf{r}_\infty^\downarrow$ will not change. Therefore for this,

$$\frac{d\mathbf{r}_\infty^\downarrow}{dT} = 0$$

and so

$$\mathbf{m}^r + \mathbf{m}^t \mathbf{r}_\infty^\downarrow + \mathbf{r}_\infty^\downarrow \mathbf{m}^t + \mathbf{r}_\infty^\downarrow \mathbf{m}^r \mathbf{r}_\infty^\downarrow = 0,$$

which can be solved for $\mathbf{r}_\infty^\downarrow$.

2.8 Influence of Reflecting and Transmitting Boundaries

In discussing the uniqueness theorem and the several methods of solving the RTE, the boundary radiances play a central role. In geophysical optics we encounter three boundaries: the top of the atmosphere that is illuminated only by a collimated beam from the sun; the surface of a water body that reflects and transmits radiance incident from above and below; and the land surface or the sea bottom that reflects some radiance back into the medium and absorbs the rest.

Consider a general reflecting and transmitting surface. If the radiance is incident from a direction $\hat{\xi}_i$ at a position $\vec{\rho}$ on the surface, the radiances reflected in the direction $\hat{\xi}_r$ and transmitted in the direction $\hat{\xi}_t$ are, respectively, given by

$$L(\vec{\rho}, \hat{\xi}_r) = \int_{\text{All } \hat{\xi}_i} r(\vec{\rho}, \hat{\xi}_i \rightarrow \hat{\xi}_r) L(\vec{\rho}, \hat{\xi}_i) d\Omega(\hat{\xi}_i)$$

and

$$L(\vec{\rho}, \hat{\xi}_t) = \int_{\text{All } \hat{\xi}_i} t(\vec{\rho}, \hat{\xi}_i \rightarrow \hat{\xi}_t) L(\vec{\rho}, \hat{\xi}_i) d\Omega(\hat{\xi}_i),$$

where r and t define the reflection and transmission functions. Given r and t these are straightforward to evaluate. If the surface were the boundary of the medium with $L(\vec{\rho}, \hat{\xi}_i)$ the radiance incident on the boundary from the *inside*, then $L(\vec{\rho}, \hat{\xi}_r)$ would be part of the radiance incident *into* the medium, and $L(\vec{\rho}, \hat{\xi}_t)$ the totality of radiance exiting the medium at $\vec{\rho}$.

2.8.1 Determination of the Reflection and Transmission Functions

Including boundaries in the radiative transfer process when r depends on *both* $\hat{\xi}_i$ and $\hat{\xi}_r$ and t depends on *both* $\hat{\xi}_i$ and $\hat{\xi}_t$ is difficult. Only in the Monte Carlo method is the process straightforward. Consider the reflectance and a collimated incident beam of radiance

$$L(\hat{\xi}_i) = F_b \delta^{(2)}(\hat{\xi} - \hat{\xi}_b),$$

where F_b is the irradiance that would fall on a plane normal to $\hat{\xi}_b$. The radiance reflected from the surface is then

$$L(\hat{\xi}_r) = F_b r(\hat{\xi}_b \rightarrow \hat{\xi}_r). \quad (2.86)$$

Now, the irradiance incident on the reflecting surface is $F_b |\hat{\xi}_b \bullet \hat{n}_S|$, where \hat{n}_S is the normal to the reflecting surface. Recalling that the radiant power is proportional to the photon

rate N , and using the definitions of radiance and irradiance, i.e., those leading up to Eq. (2.53), we can rewrite Eq. (2.86) as

$$\frac{\Delta N(\hat{\xi}_r)}{A_S |\hat{\xi}_r \bullet \hat{n}_S| \Delta \Omega(\hat{\xi}_r)} = \frac{N(\hat{\xi}_b)}{A_S |\hat{\xi}_b \bullet \hat{n}_S|} r(\hat{\xi}_b \rightarrow \hat{\xi}_r),$$

where $N(\hat{\xi}_b)$ is the photon rate incident on an area A_S of the surface and $\Delta N(\hat{\xi}_r)$ is the rate at which photons are reflected from A_S in the direction $\hat{\xi}_r$ in $\Delta \Omega(\hat{\xi}_r)$. (Note that a prescription for measuring $r(\hat{\xi}_b \rightarrow \hat{\xi}_r)$ follows immediately from this equation.) The quantity $\Delta N(\hat{\xi}_r)/N(\hat{\xi}_b)$ is the probability of reflection from $\hat{\xi}_b$ into $\Delta \Omega(\hat{\xi}_r)$, i.e.,

$$\Delta \mathcal{P}(\hat{\xi}_b \rightarrow \hat{\xi}_r) = \frac{\Delta N(\hat{\xi}_r)}{N(\hat{\xi}_b)} = \frac{|\hat{\xi}_r \bullet \hat{n}_S|}{|\hat{\xi}_b \bullet \hat{n}_S|} r(\hat{\xi}_b \rightarrow \hat{\xi}_r) \Delta \Omega(\hat{\xi}_r). \quad (2.87)$$

The probability of reflection (as opposed to absorption or transmission) from the surface in *any* direction is

$$R(\hat{\xi}_b) = \int_{\text{All } \hat{\xi}_r} d\mathcal{P}(\hat{\xi}_b \rightarrow \hat{\xi}_r) = \frac{1}{|\hat{\xi}_b \bullet \hat{n}|} \int_{\text{All } \hat{\xi}_r} |\hat{\xi}_r \bullet \hat{n}_S| r(\hat{\xi}_b \rightarrow \hat{\xi}_r) d\Omega(\hat{\xi}_r) \leq 1. \quad (2.88)$$

Now, with these preliminaries, it is clear that in the Monte Carlo method, given a sequence of random numbers ρ_i distributed $U[0,1]$, a photon incident on the surface in a direction $\hat{\xi}_b$ will be reflected if the chosen $\rho_i \leq R(\hat{\xi}_b)$. Thus, $R(\hat{\xi}_b)$ for a surface reflection event plays the same role as ω_0 for a scattering event. If the photon is reflected, its direction can be determined from the probability density⁴¹

$$p_r(\hat{\xi}_b \rightarrow \hat{\xi}_r) = \frac{1}{R(\hat{\xi}_b)} \frac{|\hat{\xi}_r \bullet \hat{n}_S|}{|\hat{\xi}_b \bullet \hat{n}_S|} r(\hat{\xi}_b \rightarrow \hat{\xi}_r),$$

in a manner identical to determining the change in a photon's direction on scattering, i.e., Eqs. (2.58) and (2.59) and Figure 2.13. So $p_r(\hat{\xi}_b \rightarrow \hat{\xi}_r)$ for a surface plays the same role as $p(\hat{\xi}_j \rightarrow \hat{\xi}_i)$ for a scattering event.

A particularly simple example of this is a lambertian surface. Recall that for such a surface the radiance reflected when a collimated beam is incident, is independent of $\hat{\xi}_r$ and the *irradiance* that is reflected is $\mathcal{A} |\hat{\xi}_b \bullet \hat{n}_S| F_b$, where \mathcal{A} is the surface *albedo*. Since the irradiance for a uniform radiance is πL , the reflected radiance is

$$L(\hat{\xi}_r) = \frac{\mathcal{A}}{\pi} |\hat{\xi}_b \bullet \hat{n}_S| F_b,$$

⁴¹The R^{-1} factor in this equation is required to normalize p_r , i.e., so that

$$\int_{\text{All } \hat{\xi}_r} p_r(\hat{\xi}_b \rightarrow \hat{\xi}_r) d\Omega(\hat{\xi}_r) = 1.$$

so, using Eq. (2.86)

$$r(\hat{\xi}_b \rightarrow \hat{\xi}_r) = \frac{\mathcal{A}}{\pi} |\hat{\xi}_b \bullet \hat{n}_S|. \quad (2.89)$$

Then from Eq. (2.87) and (2.88)

$$\begin{aligned} \Delta\mathcal{P}(\hat{\xi}_b \rightarrow \hat{\xi}_r) &= \frac{\mathcal{A}}{\pi} |\hat{\xi}_r \bullet \hat{n}_S| \Delta\Omega(\hat{\xi}_r), \\ R(\hat{\xi}_b) &= \mathcal{A}, \\ p_r(\hat{\xi}_b \rightarrow \hat{\xi}_r) &= \frac{|\hat{\xi}_r \bullet \hat{n}_S|}{\pi}. \end{aligned}$$

Now, if we let θ_r and ϕ_r be the polar and azimuth angles for $\hat{\xi}_r$, i.e.,

$$\hat{\xi}_r = \hat{e}_x \cos \phi_r \sin \theta_r + \hat{e}_y \sin \phi_r \sin \theta_r + \hat{e}_z \cos \theta_r,$$

we have

$$p_r(\theta_r, \phi_r) = \frac{\cos \theta_r}{\pi} = p_\theta(\theta_r) p_\phi(\phi_r).$$

Since p_r is independent of ϕ_r , all $0 \leq \phi_r \leq 2\pi$ are equally probable, so

$$p_\phi(\phi_r) = \frac{1}{2\pi} \quad \text{and} \quad p_\theta(\theta_r) = 2 \cos \theta_r.$$

Again, following the procedure leading to Eq. (2.60), given random numbers ρ_θ and ρ_ϕ from $U[0,1]$, we find the direction of the reflected photon from

$$\phi_r = 2\pi\rho_\phi \quad \text{and} \quad \rho_\theta = 2 \int_0^{\theta_r} \cos \theta'_r \sin \theta'_r d\theta'_r.$$

The above integral is easily evaluated to yield the reflected polar angle

$$\cos \theta_r = -\sqrt{1 - \rho_\theta}.$$

where the negative sign for the square root is chosen because θ_r must be $> \pi/2$ for an upward propagating photon ($\theta_b < \pi/2$). Of course the reflected direction is determined only *after* it has been determined that the photon is not absorbed, i.e., by testing a separate random number against $R(\hat{\xi}_b) = \mathcal{A}$.

Another simple boundary is a flat Fresnel-reflecting surface, e.g., the surface of a flat water body. For such a surface we have

$$L(z, u_r, \phi_r) = \int_0^{2\pi} d\phi_i \int_0^1 r(u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) L(z, u_i, \phi_i) du_i,$$

and a similar equation for $L(z, u_t, \phi_t)$. Letting $r_f(u)$ be the Fresnel reflectance for an incident angle $\arccos(u)$, we have

$$r(u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) = r_f(u_i) \delta(u_i + u_r) \delta(\phi_i - \phi_r),$$

and

$$t(u_i \rightarrow u_t, \phi_i \rightarrow \phi_t) = (1 - r_f(u_i)) \delta(u_i - u_t) \delta(\phi_i - \phi_t),$$

where u_t is related to u_i through Snell's law. Thus, the incident radiance is partially reflected with $u_r = -u_i$ and $\phi_r = \phi_i$ (the law of reflection), and partially transmitted.

2.8.2 Boundaries in Plane-parallel Media

We have seen that in plane parallel media, when the phase function depends only on the scattering angle, it is useful to decompose the radiance in a Fourier series (Eq. (2.68)). In that case we found that each Fourier component of the radiance individually satisfied an RTE. We might expect similar simplification if the reflectance and transmittance functions are subjected to Fourier analysis; however, simplification is not obtained unless the reflectance and transmittance functions display a considerable amount of symmetry. It is beyond the scope of this work to examine the various symmetries that are possible, so we will consider a specific, and useful, example.

Place the lower boundary of a 1-d slab at τ_b . Then the reflected radiance is

$$L(\tau_b, u_r, \phi_r) = \int_0^{2\pi} d\phi_i \int_0^1 r(\tau_b, u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) L(\tau_b, u_i, \phi_i) du_i, \quad u_i > 0, u_r < 0. \quad (2.90)$$

We will examine the case where the reflectance depends on azimuth through $|\phi_i - \phi_r|$, i.e.,

$$r(u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) = r(u_i \rightarrow u_r, |\phi_i - \phi_r|).$$

Such symmetry is extant when dealing with reflection from the surface of a water body roughened by the wind (when the wind direction is ignored). In this case r is an *even* function of $\phi_i - \phi_r$ so the Fourier expansion becomes particularly simple:

$$r(u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) = r_0(u_i \rightarrow u_r) + 2 \sum_{l=1}^{\infty} r_l(u_i \rightarrow u_r) \cos l(\phi_i - \phi_r),$$

with

$$r_l(u_i \rightarrow u_r) = \frac{1}{2\pi} \int_0^{2\pi} r(u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) \cos l(\phi_i - \phi_r) d\phi_i. \quad (2.91)$$

Note that the integration above could equally well be over ϕ_r . Writing the Fourier series for the radiance incident on the boundary,

$$L(\tau_b, u, \phi) = L_0(u) + 2 \sum_{m=1}^{\infty} L_m^{(c)}(\tau_b, u) \cos m(\phi - \phi_1) + 2 \sum_{m=1}^{\infty} L_m^{(s)}(\tau_b, u) \sin m(\phi - \phi_1), \quad (2.92)$$

and inserting it into Eq. (2.90), we have (after again dropping the τ_b for economy)

$$\begin{aligned} L(u_r, \phi_r) &= \int_0^1 du_i \int_0^{2\pi} d\phi_i \left[r_0(u_i \rightarrow u_r) L_0^{(c)}(u_i) \right. \\ &\quad + 4 \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} r_l(u_i \rightarrow u_r) L_m^{(c)}(u) \cos l(\phi_r - \phi_i) \cos m(\phi_i - \phi_1) \\ &\quad \left. + 4 \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} r_l(u_i \rightarrow u_r) L_m^{(s)}(u) \cos l(\phi_r - \phi_i) \sin m(\phi_i - \phi_1) \right] \\ &= 2\pi \int_0^1 du_i \left[r_0(u_i \rightarrow u_r) L_0^{(c)}(u_i) \right. \\ &\quad + 2 \sum_{m=1}^{\infty} r_m(u_i \rightarrow u_r) L_m^{(c)}(u) \cos m(\phi_r - \phi_1) \\ &\quad \left. + 2 \sum_{m=1}^{\infty} r_m(u_i \rightarrow u_r) L_m^{(s)}(u) \sin m(\phi_r - \phi_1) \right], \end{aligned}$$

or

$$\begin{aligned} L_m^{(c)}(u_r) &= 2\pi \int_0^1 r_m(u_i \rightarrow u_r) L_m^{(c)}(u_i) du_i, \\ L_m^{(s)}(u_r) &= 2\pi \int_0^1 r_m(u_i \rightarrow u_r) L_m^{(s)}(u_i) du_i, \end{aligned}$$

where $u_r < 0$. Thus, in this case, *any given Fourier component of the radiance satisfies an RTE and bottom boundary condition separate from all of the other components*. This is an enormous simplification.

A trivial example of a surface having these properties is the lambertian reflector, for which

$$L(u_r) = \mathcal{A} \frac{E_d}{\pi},$$

where E_d is the downward irradiance onto the surface

$$E_d = \int_0^{2\pi} d\phi_i \int_0^1 u_i L(u_i, \phi_i) du_i.$$

Thus,

$$L(u_r) = 2\mathcal{A} \int_0^1 u_i L_0^{(c)}(u_i) du_i,$$

and the reflectance is given by

$$r_m^{(c)} = \frac{\mathcal{A} u_i}{\pi} \delta_{m0}.$$

So for a lambertian lower boundary, $L_m^{(c)}(u_r) = 0$ and $L_m^{(s)}(u_r) = 0$ if $m \neq 0$.

For a final example of very simple boundary reflectances, we return to the flat Fresnel-reflecting surface. From Eq. (2.91) and the properties of the δ -function, $r_m(u_i \rightarrow u_r) = r_f(u_i)/2\pi$, so

$$L_m^{(c,s)}(u_r) = r_f(u_i) L_m^{(c,s)}(u_i), \quad u_r = -u_i,$$

with a similar equation for the transmitted radiance. In this case, *every Fourier component is reflected in the same manner at the interface.*

2.8.3 A cautionary note

One must be careful not to confuse the “ r ” in the equation

$$L(\tau_b, u_r, \phi_r) = \int_0^{2\pi} d\phi_i \int_0^1 r(\tau_b, u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) L(\tau_b, u_i, \phi_i) du_i, \quad u_i > 0, \quad u_r < 0.$$

with the \mathbf{r} 's in the matrix operator and invariant imbedding methods, although we will see below they are related. Consider for example the situation in which

$$r(u_i \rightarrow u_r, \phi_i \rightarrow \phi_r) = r(u_i \rightarrow u_r, |\phi_i - \phi_r|)$$

discussed above. For this, the boundary condition at the surface was

$$L_m^{(c,s)}(u_r) = 2\pi \int_0^1 r_m(u_i \rightarrow u_r) L_m^{(c,s)}(u_i) du_i.$$

Now, use Gaussian quadratures to discretize the integrals as before,

$$L_m^{(c,s)}(u_j) = 2\pi \sum a_i r_m(u_i \rightarrow u_j) L_m^{(c,s)}(u_i),$$

where $u_i > 0$ and $u_j < 0$. Writing for $L_m^{(c)}$ or $L_m^{(s)}$

$$\mathbf{L}_m^+ = \begin{pmatrix} L(u_n) \\ L(u_{n-1}) \\ \vdots \\ L(u_1) \end{pmatrix}, \quad u_i > 0, \quad \text{and} \quad \mathbf{L}_m^- = \begin{pmatrix} L(u_{-n}) \\ L(u_{-n+1}) \\ \vdots \\ L(u_{-1}) \end{pmatrix}, \quad u_{-j} < 0,$$

we can put the equation in matrix form

$$\mathbf{L}_m^- = \mathbf{r}_s \mathbf{L}_m^+,$$

and \mathbf{r}_s plays a role similar to \mathbf{r}_{bc} in Eq. (2.79). Writing out the matrix shows that the elements of \mathbf{r}_s are

$$[\mathbf{r}_s]_{ij} = a_i r_m(u_i \rightarrow u_{-j}),$$

and the \mathbf{r}_s matrix contains the a_i 's as well as the r 's. If the surface is lambertian, $r_m(u_j \rightarrow u_i) = \mathcal{A} u_i \delta_{m0} / \pi$ and

$$[\mathbf{r}_s]_{ij} = \frac{\mathcal{A}}{\pi} a_i u_i \delta_{m0}.$$

As we have seen before, only the Fourier order with $m = 0$ is reflected from a lambertian surface.

2.8.4 The Bi-directional Reflectance Distribution Function — BRDF

For completeness, we define an alternative to the reflection function $r(\hat{\xi}_i \rightarrow \hat{\xi}_r)$ called the bi-directional reflectance distribution function or the BRDF. This is often used in describing reflection from solid surfaces. It is defined in such a manner that it is constant for a lambertian surface, i.e.,

$$BRDF(\hat{\xi}_i \rightarrow \hat{\xi}_r) \triangleq \frac{r(\hat{\xi}_i \rightarrow \hat{\xi}_r)}{|\hat{\xi}_i \bullet \hat{n}_S|}$$

Also, unlike r ,

$$BRDF(\hat{\xi}_i \rightarrow \hat{\xi}_r) = BRDF(\hat{\xi}_r \rightarrow \hat{\xi}_i),$$

a property known as reciprocity.

In later chapters, we will define the *reflectance* $\rho(\hat{\xi}_0 \rightarrow \hat{\xi}_{\text{out}})$ of the ocean-atmosphere system to be

$$\rho(\hat{\xi}_0 \rightarrow \hat{\xi}_{\text{out}}) = \frac{\pi L(\hat{\xi}_0 \rightarrow \hat{\xi}_{\text{out}})}{|\hat{\xi}_0 \bullet \hat{n}| F_0},$$

where $\hat{\xi}_{\text{out}}$ is the direction of the radiance L exiting the top of the atmosphere, $\hat{\xi}_0$ is the direction in which solar radiation propagates into the atmosphere, \hat{n} is the normal to the top of the atmosphere, and F_0 is the solar irradiance on a plane normal to $\hat{\xi}_0$. Comparing this to Eq. (2.86) and using the definition of the BRDF, we see that

$$\rho(\hat{\xi}_0 \rightarrow \hat{\xi}_{\text{out}}) = \pi BRDF(\hat{\xi}_0 \rightarrow \hat{\xi}_{\text{out}}).$$

If the ocean-atmosphere system reflected light as a lambertian reflector of albedo \mathcal{A} ,

$$\rho(\hat{\xi}_{\text{out}}) = \mathcal{A}.$$

2.9 An Important Integral — Gershun's Equation

Notice that in Eq. (2.61) the u in the derivative term can be placed inside the derivative, i.e.,

$$u \frac{\partial}{\partial z} L(z, u, \phi, \lambda) = \frac{\partial}{\partial z} [u L(z, u, \phi, \lambda)].$$

If we effect this, multiply Eq. (2.61) by $du d\phi$, integrate over all u and ϕ (i.e., over solid angle), and use the form of the irradiances in Eq. (2.63) we find

$$\frac{\partial}{\partial z} [E_d(z, \lambda) - E_u(z, \lambda)] = -a(z, \lambda) E_0(z, \lambda) + \int_{-1}^1 du \int_0^{2\pi} Q(z, u, \phi, \lambda) d\phi, \quad (2.93)$$

where we have used the fact that, after integrating over u and ϕ ,

$$\begin{aligned} & \int_{-1}^1 du \int_0^{2\pi} d\phi \int_{-1}^1 du' \int_0^{2\pi} d\phi' [\beta(z, u' \rightarrow u, \phi' \rightarrow \phi, \lambda) L(z, u', \phi', \lambda)] \\ &= b(z, \lambda) \int_{-1}^1 du' \int_0^{2\pi} d\phi' L(z, u', \phi', \lambda) = b(z, \lambda) E_0(z, \lambda), \end{aligned}$$

and

$$a(z, \lambda) = c(z, \lambda) - b(z, \lambda).$$

Equation (2.93) is usually referred to as *Gershun's law*. It provides a means of using irradiances to deduce $a(z, \lambda)$, if Q is known.

We can interpret Gershun's law physically in the following manner. Consider a slab of medium between z and $z + \Delta z$. Gershun's law states

$$\begin{aligned} E_d(z + \Delta z, \lambda) - E_d(z, \lambda) - E_u(z + \Delta z, \lambda) + E_u(z, \lambda) &= -a(z, \lambda) E_0(z, \lambda) \Delta z \\ &+ \int_{\text{All } \hat{\xi}} Q(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}) \Delta z. \end{aligned} \quad (2.94)$$

If we ignore the Q term for the moment, then

$$\frac{E_d(z + \Delta z, \lambda) - E_d(z, \lambda) - E_u(z + \Delta z, \lambda) + E_u(z, \lambda)}{E_0(z, \lambda) \Delta z} = \frac{1}{E_0(z, \lambda)} \frac{\Delta E_V(z, \lambda)}{\Delta z} = -a(z, \lambda), \quad (2.95)$$

where E_V is the vector irradiance, and the measurement of E_d , E_u , and E_0 , provides a way to determine the absorption coefficient within the slab. Noting the relationship between E_0 and the electromagnetic *spectral* energy density, i.e., $E_0 = c\ell \langle u_{em}^s \rangle$, we see that in the case of a non-scattering medium Gershun's law is just a restatement of Poynting's theorem. Thus, Gershun's law is the radiative transfer statement of the conservation of energy.

What is the effect of the Q term? Equation (2.94) can be rewritten

$$E_{\text{in}}(\lambda) - E_{\text{out}}(\lambda) = a(z, \lambda) E_0(z, \lambda) \Delta z - \int_{\text{All } \hat{\xi}} Q(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}) \Delta z,$$

where $E_{\text{in}}(\lambda) = E_d(z, \lambda) + E_u(z + \Delta z, \lambda)$ and $E_{\text{out}}(\lambda) = E_d(z + \Delta z, \lambda) + E_u(z, \lambda)$ are the irradiances into and out of the slab respectively. How do we interpret the last term? Noting that Q is the intensity density of sources, i.e., $Q(z, \hat{\xi}) = dJ(z, \hat{\xi})/dV$, we have

$$\int_{\text{All } \hat{\xi}} Q(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}) \Delta z = \int_{\text{All } \hat{\xi}} \frac{dP(z, \hat{\xi})}{A_s},$$

where A_s is the lateral area of the volume being examined, i.e., $dV = A_s dz$, and $P(z, \hat{\xi})$ is the spectral radiant power propagating in the direction $\hat{\xi}$ within the slab. The integrand is just the exiting boundary irradiance due to P exiting in the direction $\hat{\xi}$, so the later integral is just an out-going irradiance, i.e., $E_{\text{out}}^Q(\lambda)$. Simplifying this yields

$$E_{\text{in}}(\lambda) - E_{\text{out}}(\lambda) + E_{\text{out}}^Q(\lambda) = a(z, \lambda) E_0(z, \lambda) \Delta z.$$

Now, since $E_{\text{out}}(\lambda)$ already contains $E_{\text{out}}^Q(\lambda)$, because it was derived from the actual irradiances (or irradiances), when we add it to the left hand side, it cancels the $-E_{\text{out}}^Q(\lambda)$ in $-E_{\text{out}}(\lambda)$. Thus, when there are internal sources one needs to determine $E_{\text{out}}^Q(\lambda)$ in order to estimate a from Gershun's law, or to put it another way; when internal sources are present, but ignored, Gershun's law in the form of Eq. (2.95) will always yield a value of $a(z, \lambda)$ that is too small.

It is pedagogically useful to examine Eq. (2.94) using the photon concept, i.e., that radiant power through a surface is proportional to the rate at which photons stream through the surface. The terms in this equation contain radiant power, which is $(hc_\ell/\lambda)(dn/dt)$, energy density, which is $(hc_\ell/\lambda)(dn/dV)$ or intensity density (Q), which is $(hc_\ell/\lambda)(dn_Q(\hat{\xi})/dV d\Omega)$, where the n 's are photon numbers. Thus, rather than working with radiometric units, we can work with photon numbers — the (hc_ℓ/λ) factor occurs in each term. Ignoring Q for the moment, if we consider a horizontal area A_S of the slab, then we can relate the irradiances on the left-hand-side to the rates at which photons enter and leave the slab, and the scalar irradiance on the right-hand-side to the photon density within the slab, i.e.,

$$\frac{1}{A_S} \left[\frac{dn_{\text{in}}}{dt} - \frac{dn_{\text{out}}}{dt} \right] = a(z, \lambda) \left[c_\ell \frac{dn_{\text{inside}}}{dV} \right] \Delta z,$$

or

$$dn_{\text{in}} - dn_{\text{out}} = a(z, \lambda) c_\ell dt \left[A_S \Delta z \frac{dn_{\text{inside}}}{dV} \right].$$

Now, the quantity in the square brackets above is the instantaneous number of photons in the slab. During a time dt these photons will move a distance $dl = c dt$, so the number of these photons absorbed in Δz in the time interval dt is just

$$dn_{\text{absorbed}} = a(z, \lambda) c_\ell dt \left[A_S \Delta z \frac{dn_{\text{inside}}}{dV} \right].$$

Thus, as photons stream through the slab, they are absorbed at a rate

$$\left(\frac{dn}{dt} \right)_{\text{absorbed}} = a(z, \lambda) c_\ell \left[A_S \Delta z \frac{dn_{\text{inside}}}{dV} \right],$$

and in terms of photons, Gershun's law applied to a thin slab states

$$\left(\frac{dn}{dt} \right)_{\text{in}} - \left(\frac{dn}{dt} \right)_{\text{out}} = \left(\frac{dn}{dt} \right)_{\text{absorbed}},$$

i.e., it is a mathematical statement of the conservation of photon numbers (energy). It is readily shown that when the internal source term $\int Q d\Omega$ is included, it leads to a term $-(dn/dt)_{\text{emitted}}$ on the right-hand-side above,⁴² so the full Gershun equation results in

$$\left(\frac{dn}{dt} \right)_{\text{in}} - \left(\frac{dn}{dt} \right)_{\text{out}} = \left(\frac{dn}{dt} \right)_{\text{absorbed}} - \left(\frac{dn}{dt} \right)_{\text{emitted}}.$$

As one expects, the conservation of energy is not restricted to slab geometry, Gershun's law can be derived for arbitrary geometry as well.⁴³

2.10 Summary and Concluding Remarks

In this chapter we have presented the tools needed to understand the propagation of radiation in an absorbing, scattering and refracting medium. We started by defining the

⁴²This is easy to show:

$$\int_{\text{All } \hat{\xi}} Q(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}) \rightarrow \int_{\text{All } \hat{\xi}} \frac{dn_Q(z, \hat{\xi})/dt}{dV d\Omega} d\Omega \Delta z = \frac{1}{A_s} \int_{\text{All } \hat{\xi}} \frac{dn_Q(z, \hat{\xi})}{dt} = \frac{1}{A_s} \left(\frac{dn}{dt} \right)_{\text{emitted}}.$$

⁴³Starting with Eq. (2.33) including the variable m in Eq. (2.5.4), rewriting $\hat{\xi} \bullet \nabla(L/m^2)$ as $\nabla \bullet (\hat{\xi} L/m^2)$, integrating over all solid angles $d\Omega(\hat{\xi})$ and a volume of V of the medium, and applying the divergence theorem yields,

$$\oint_S \hat{n} \bullet \hat{\xi} \frac{L(\vec{r}, \hat{\xi}, \lambda)}{m^2(\vec{r})} dS = \int_V a(\vec{r}, \lambda) \frac{E_0(\vec{r}, \lambda)}{m^2(\vec{r})} dV + \int_V \int_{\text{All } \hat{\xi}} \frac{Q(\vec{r}, \hat{\xi}, \lambda)}{m^2(\vec{r})} d\Omega(\hat{\xi}) dV,$$

where S is the surface bounding V , \hat{n} is the outward normal to S , and the last term in Eq. (2.33) has been replaced by $Q(\vec{r}, \hat{\xi}, \lambda)$. These terms have interpretations identical to the corresponding terms in Eq. (2.93).

radiance, which is of fundamental importance, as it is the quantity measured by remote optical sensors. The propagation of radiance is governed by the radiative transfer equation. After defining the inherent optical properties of the medium — the absorption coefficient and the volume scattering functions (elastic and inelastic) — the radiative transfer equation was developed in a phenomenological manner. Unfortunately, analytical solutions of this equation are possible only in the simplest of cases, e.g., no scattering; however several solution methods were presented that are amenable to numerical approximations, which can be made as accurate as desired. Two methods that were used extensively in preparing this work, the successive order of scattering and Monte Carlo, were developed in considerable detail. Three others that have also been used extensively in environmental optics were sketched with enough detail to provide the reader a starting point for further study if desired. In Appendix 1 an analytical first-order solution, that leads to the single-scattering solution, is derived. As it is a good approximation to radiative transfer in the atmosphere (actually the starting point for the atmospheric correction algorithm) and is the basis for an excellent approximation to radiative transfer in the ocean — the quasi-single scattering approximation — the single-scattering solution is used extensively throughout the rest of this work and should be completely understood before proceeding further. In Appendix 2 we provide detailed derivations of various Monte Carlo estimators of solutions to the RTE with and without consideration of the polarization properties of the radiance.

2.11 Appendix 1: First-Order Solutions for 1-d Media

In this appendix we present the first-order solutions to the radiative transfer equation in a 1-d plane-parallel medium. By “first-order” we mean solutions utilizing only the first two terms in Eq. (2.49) and the first two equations in Eq. (2.50). They contain ω_0 only in powers of 0 and 1. These solutions are reasonably accurate in slab geometry when the optical thickness of the slab is $\ll 1$. We reserve the terms “single-scattering solution” or “single-scattering approximation” for the first-order solution in which we *explicitly* approximate exponentials such as $\exp[-\tau/u]$ by $1 - \tau/u$, which implies only *one* interaction with the medium. The cases chosen for examination are ones that will be useful in later chapters. We also introduce the reader to the additional complexity that must be faced when taking into account the polarization state of the radiation.

2.11.1 First-Order Scattering Solution in Slab Geometry

Consider a homogeneous slab of optical thickness τ_1 illuminated from the top by an incident radiance $L(0, u, \phi)$, $u > 0$. There is no illumination on the bottom, so $L(\tau_1, u, \phi) = 0$ for

$u < 0$. The solution provided by the successive order of scattering method follows from

$$\begin{aligned} u \frac{\partial}{\partial \tau} L^{(0)}(\tau, u, \phi) + L^{(0)}(\tau, u, \phi) &= 0 \\ u \frac{\partial}{\partial \tau} L^{(1)}(\tau, u, \phi) + L^{(1)}(\tau, u, \phi) &= \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) L^{(0)}(\tau, u', \phi') \\ &\vdots, \end{aligned}$$

where

$$\begin{aligned} L^{(0)}(0, u, \phi) &= L(0, u, \phi) & u > 0 \\ L^{(0)}(\tau_1, u, \phi) &= 0 & u < 0 \end{aligned}$$

and

$$\left. \begin{aligned} L^{(n)}(0, u, \phi) &= 0, & u > 0 \\ L^{(n)}(\tau_1, u, \phi) &= 0, & u < 0 \end{aligned} \right\} n > 0.$$

The solution for $L^{(0)}$ satisfying the boundary conditions is

$$\begin{aligned} L^{(0)}(\tau, u, \phi) &= L(0, u, \phi) \exp \left[-\frac{\tau}{u} \right] & u > 0 \\ L^{(0)}(\tau, u, \phi) &= 0 & u < 0. \end{aligned}$$

Thus,

$$u \frac{\partial}{\partial \tau} L^{(1)}(\tau, u, \phi) + L^{(1)}(\tau, u, \phi) = \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) L(0, u', \phi') \exp \left[-\frac{\tau}{u} \right]$$

The solution to the homogeneous equation (left-hand-side set to zero) is

$$L_h^{(1)}(\tau, u, \phi) = C_h(u, \phi) \exp \left[-\frac{\tau}{u} \right],$$

and it is easily verified that

$$L_i^{(1)}(\tau, u, \phi) = \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L(0, u', \phi') \exp \left[-\frac{\tau}{u'} \right]$$

satisfies the inhomogeneous equation, so the general solution for $L^{(1)}$ is

$$L^{(1)}(\tau, u, \phi) = L_h^{(1)}(\tau, u, \phi) + L_i^{(1)}(\tau, u, \phi).$$

$C_h(u, \phi)$ is determined from the boundary conditions:

$$\begin{aligned} L^{(1)}(0, u, \phi) &= 0, & u > 0 \\ L^{(1)}(\tau_1, u, \phi) &= 0, & u < 0. \end{aligned}$$

Thus, at the top

$$\begin{aligned} 0 &= L_h^{(1)}(0, u, \phi) + L_i^{(1)}(0, u, \phi) \\ &= C_h(u, \phi) + \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L(0, u', \phi'), \quad u > 0, \end{aligned}$$

so

$$C_h(u, \phi) = -\frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L(0, u', \phi'), \quad u > 0.$$

Likewise, at the bottom

$$\begin{aligned} 0 &= L_h^{(1)}(\tau_1, u, \phi) + L_i^{(1)}(\tau_1, u, \phi) \\ &= C_h(u, \phi) \exp \left[-\frac{\tau_1}{u} \right] \\ &\quad + \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L(0, u', \phi') \exp \left[-\frac{\tau_1}{u'} \right], \quad u < 0, \end{aligned}$$

so

$$\begin{aligned} C_h(u, \phi) &= \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} \\ &\quad \times L(0, u', \phi') \exp \left[-\tau_1 \left(\frac{1}{u'} - \frac{1}{u} \right) \right], \quad u < 0. \end{aligned}$$

Combining, to first order in ω_0 , we have

$$L(\tau, u, \phi) = L^{(0)}(\tau, u, \phi) + \omega_0 L^{(1)}(\tau, u, \phi),$$

or

$$\begin{aligned} L(\tau, u, \phi) &= L(0, u, \phi) \exp \left[-\frac{\tau}{u} \right] \\ &\quad + \frac{\omega_0}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L(0, u', \phi') \\ &\quad \times \left\{ \exp \left[-\frac{\tau}{u'} \right] - \exp \left[-\frac{\tau}{u} \right] \right\}, \quad u > 0, \end{aligned} \tag{2.96}$$

$$\begin{aligned} L(\tau, u, \phi) &= \exp \left[-\frac{\tau}{u} \right] \\ &\quad \times \frac{\omega_0}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L(0, u', \phi') \\ &\quad \times \left\{ \exp \left[-\tau \left(\frac{1}{u'} - \frac{1}{u} \right) \right] - \exp \left[-\tau_1 \left(\frac{1}{u'} - \frac{1}{u} \right) \right] \right\}, \quad u < 0. \end{aligned} \tag{2.97}$$

2.11.2 The Single-Scattering Solution

The first-order solutions given in Eqs. (2.96) and (2.97) are rigorously correct and do provide the first two terms in Eq. (2.49); however, they are not a good approximation to the total radiance when the the rest of the series is omitted. A much better approximation for media with $\tau_1 \ll 1$ is achieved by expanding the exponentials in a series and keeping only the first term, i.e., replacing $\exp[-\tau/u]$ by $1 - \tau/u$. As we shall see, when this replacement is made the resulting radiance will be proportional to terms containing $\omega_0\tau = b$, i.e., *first order* in b . Such terms imply only one scattering. We refer to the formulae after this replacement as the “single-scattering solution.” It will form the basis for much of the development of atmospheric and oceanic radiative transfer in later chapters.

When the above replacement is made, the solutions become

$$L(\tau, u, \phi) = L(0, u, \phi) \left[1 - \frac{\tau}{u} \right] + \frac{\omega_0\tau_1}{4\pi u} \int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) L(0, u', \phi'), \quad u > 0, \quad (2.98)$$

and

$$L(\tau, u, \phi) = \frac{\omega_0(\tau - \tau_1)}{4\pi u} \int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) L(0, u', \phi'), \quad u < 0. \quad (2.99)$$

The radiances exiting the top and bottom of the slab are, respectively,

$$L(0, u, \phi) = -\frac{\omega_0\tau_1}{4\pi u} \int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) L(0, u', \phi'), \quad u < 0, \quad (2.100)$$

and

$$L(\tau_1, u, \phi) = L(0, u, \phi) \left[1 - \frac{\tau_1}{u} \right] + \frac{\omega_0\tau_1}{4\pi u} \int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) L(0, u', \phi'), \quad u > 0. \quad (2.101)$$

If the phase function and the radiances are expanded in Fourier series in azimuth, i.e., as in Eq. (2.65) and (2.68), then these solutions become

$$L_m^{(c,s)}(0, u) = -\frac{\omega_0\tau_1}{2u} \int_0^1 du' p_m(u' \rightarrow u) L_m^{(c,s)}(0, u'), \quad u < 0,$$

and

$$L_m^{(c,s)}(\tau_1, u) = L_m^{(c,s)}(0, u) \left[1 - \frac{\tau_1}{u} \right] + \frac{\omega_0\tau_1}{2u} \int_0^1 du' p_m(u' \rightarrow u) L_m^{(c,s)}(0, u'), \quad u > 0,$$

where $L_0^{(s)}(\tau, u) = 0$.

Two cases of considerable interest are that of a beam source, e.g., the Sun, for which $L(0, u, \phi) = F_0 \delta(u - u_0) \delta(\phi - \phi_0)$, and a totally diffuse source for which $L(0, u, \phi) = L_0$, a constant. For a beam source the integration is trivial and we have,

$$L(0, u, \phi) = -\frac{\omega_0 F_0 \tau_1}{4\pi u} P(u_0 \rightarrow u, \phi_0 \rightarrow \phi), \quad u < 0, \quad (2.102)$$

and

$$\begin{aligned} L(\tau_1, u, \phi) &= F_0 \delta(u - u_0) \delta(\phi - \phi_0) \left[1 - \frac{\tau_1}{u} \right] \\ &+ \frac{\omega_0 F_0 \tau_1}{4\pi u} P(u_0 \rightarrow u, \phi_0 \rightarrow \phi), \quad u > 0. \end{aligned} \quad (2.103)$$

The first term in Eq. (2.103) is the attenuated incident beam while the second is the radiance created through scattering. The irradiances reflected and transmitted by the slab are

$$E_u(0) = \frac{\omega_0 F_0 \tau_1}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P(u_0 \rightarrow u', \phi_0 \rightarrow \phi') \quad (2.104)$$

and

$$E_d(\tau_1) = u_0 F_0 \left[1 - \frac{\tau_1}{u_0} \right] + \frac{\omega_0 F_0 \tau_1}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' P(u_0 \rightarrow u', \phi_0 \rightarrow \phi'). \quad (2.105)$$

Now, the integrals in Eqs. (2.104) and (2.105) are proportional to the probabilities of scattering from $\hat{\xi}_0$ into upward and downward directions, respectively:

$$\begin{aligned} P_u(u_0, \phi_0) &\triangleq \frac{1}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P(u_0 \rightarrow u', \phi_0 \rightarrow \phi'), \\ P_d(u_0, \phi_0) &\triangleq \frac{1}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' P(u_0 \rightarrow u', \phi_0 \rightarrow \phi'). \end{aligned} \quad (2.106)$$

With these definitions, we have

$$E_u(0) = \omega_0 F_0 \tau_1 P_u(u_0, \phi_0)$$

and

$$E_d(\tau_1) = u_0 F_0 \left[1 - \frac{\tau_1}{u_0} \right] + \omega_0 F_0 \tau_1 P_d(u_0, \phi_0).$$

Since $P_u(u_0, \phi_0) + P_d(u_0, \phi_0) = 1$, adding these gives

$$E_u(0) + E_d(\tau_1) + (1 - \omega_0) F_0 \tau_1 = u_0 F_0.$$

The first two terms on the left-hand-side provide the flux out of the slab, the third term is the flux absorbed by the slab, and the right-hand-side is the flux into the slab. Thus,

the energy is conserved in this approximation. If the exponentials were not expanded, or if more than the linear terms were retained in the expansion of $\exp(-\tau_1/u)$, the energy would *not* be conserved because the first-order equations are not consistent, i.e., the exponentials contain all powers of τ , which imply all orders of scattering (not just one as is implied in $L^{(1)}$).⁴⁴

In the case with a totally diffuse source, Eq. s(2.98) and (2.99) yield

$$L(\tau, u, \phi) = L_0 \left[1 - \frac{\tau}{u} \right] + \frac{\omega_0 \tau L_0}{4\pi u} \int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi), \quad u > 0, \quad (2.107)$$

and

$$L(\tau, u, \phi) = \frac{\omega_0(\tau - \tau_1)L_0}{4\pi u} \int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi), \quad u < 0. \quad (2.108)$$

The integrals in these expressions can be simplified if the phase function depends only on the angle Θ between $\hat{\xi}$ and $\hat{\xi}'$. In that case,

$$P(u' \rightarrow u, \phi' \rightarrow \phi) = P(u \rightarrow u', \phi \rightarrow \phi')$$

so

$$\int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) = \int_0^1 du' \int_0^{2\pi} d\phi' P(u \rightarrow u', \phi \rightarrow \phi').$$

If $u > 0$, this integral represents the scattering from downward propagating radiation (u) into the downward directions ($0 < u' < 1$), so

$$\int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) = 4\pi P_d(u, \phi).$$

Similarly, if $u < 0$, the integral represents scattering by upward propagating radiation into downward directions, which is the same as downward propagating radiation scattered into upward directions. Thus, for $u < 0$,

$$\int_0^1 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) = 4\pi P_u(u, \phi).$$

The radiances then become

$$L(\tau, u, \phi) = L_0 \left[1 - \frac{\tau}{u} \left(1 - \omega_0 P_d(u, \phi) \right) \right], \quad u > 0,$$

$$L(\tau, u, \phi) = L_0 \frac{\omega_0(\tau - \tau_1)}{u} P_u(u, \phi), \quad u < 0,$$

⁴⁴Of course, if all of the terms in Eq. (2.49) were retained, energy *would* be conserved.

and

$$\begin{aligned} L(\tau_1, u, \phi) &= L_0 \left[1 - \frac{\tau_1}{u} \left(1 - \omega_0 P_d(u, \phi) \right) \right], & u > 0, \\ L(0, u, \phi) &= -L_0 \frac{\omega_0 \tau_1}{u} P_u(u, \phi), & u < 0. \end{aligned}$$

Interestingly, the fraction of incident diffuse radiance transmitted with a totally diffuse source is equal to the fraction of incident irradiance transmitted with a beam source, i.e.,

$$\frac{E_d(\tau_1)}{u_0 F_0} = \left[1 - \frac{\tau_1}{u_0} \left(1 - \omega_0 P_d(u_0, \phi_0) \right) \right] = \frac{L(\tau_1, u_0, \phi_0)}{L_0}. \quad (2.109)$$

This is no accident — we shall see later that it is demanded by the reciprocity principle.

2.11.3 The Single-Scattering Solution Including Polarization

To include polarization and find the Stokes vector \mathbf{I} in a slab illuminated by a beam of unpolarized light, e.g., from the Sun, we must solve

$$u \frac{\partial \mathbf{I}(z, u, \phi)}{\partial z} + \mathbf{c} \mathbf{I}(\tau, u, \phi) = \int_{-1}^1 du' \int_0^{2\pi} d\phi' \boldsymbol{\beta}(u' \rightarrow u, \phi' \rightarrow \phi) \mathbf{I}(\tau, u', \phi').$$

In general, the 4×4 extinction matrix \mathbf{c} has no nonzero elements and the solution of this equation is extremely difficult. However, if the scattering is due to atoms (and/or molecules), i.e., Rayleigh scattering or particles that have high enough symmetry, e.g., spheres, then \mathbf{c} is a multiple of the unit matrix, i.e.,⁴⁵

$$\mathbf{c} = c \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In this case the radiative transfer equation simplifies to

$$u \frac{\partial \mathbf{I}(\tau, u, \phi)}{\partial \tau} + \mathbf{I}(\tau, u, \phi) = \frac{\omega_0}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \mathbf{Z}(u' \rightarrow u, \phi' \rightarrow \phi) \mathbf{I}(\tau, u', \phi'),$$

where $\tau = cz$, $\omega_0 = b/c$, and $\mathbf{Z}(u' \rightarrow u, \phi' \rightarrow \phi)$ is the Mueller *phase* matrix for scattering from $u' \rightarrow u$ and $\phi' \rightarrow \phi$: $\mathbf{Z} = 4\pi\boldsymbol{\beta}/b$. (Note that we have replaced \mathbf{I}_L (used earlier) by \mathbf{I}

⁴⁵In general, if the medium contains particles and their mirror images in equal numbers, and the particles are in random orientation, \mathbf{c} is a multiple of the unit matrix. An example of a situation in which \mathbf{c} is not of this form is extinction by long, thin, absorbing cylinders that have a preferred alignment in space.

to make the notation simpler.) The solution must satisfy the boundary conditions

$$\mathbf{I}(0, u, \phi) = \delta(u - u_0)\delta(\phi - \phi_0) \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad u > 0, \quad \text{and} \quad \mathbf{I}(\tau_1, u, \phi) = 0, \quad u < 0.$$

The successive order of scattering method is also applicable to this equation without any changes except replacing L by \mathbf{I} and P by \mathbf{Z} . A computation identical to that provided when polarization is ignored leads to the results for $\tau_1 \ll 1$:

$$\begin{aligned} \mathbf{I}(0, u, \phi) &= -\frac{\omega_0 \tau_1}{4\pi u} \mathbf{Z}(u_0 \rightarrow u, \phi_0 \rightarrow \phi) \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad u < 0; \\ \mathbf{I}(\tau_1, u, \phi) &= \delta(u - u_0)\delta(\phi - \phi_0) \left[1 - \frac{\tau_1}{u}\right] \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ &\quad + \frac{\omega_0 \tau_1}{4\pi u} \mathbf{Z}(u_0 \rightarrow u, \phi_0 \rightarrow \phi) \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad u > 0. \end{aligned}$$

The interpretation of these equations requires considerable care. Conventionally, the reference plane for describing \mathbf{I} in a slab is the plane formed by \hat{e}_z and $\hat{\xi}$ — the vector $\hat{\ell}$ (also called \hat{e}_ℓ) being in this plane and \hat{r} (also called \hat{e}_r) perpendicular to this plane, such that $\hat{r} \times \hat{\ell} = \hat{\xi}$. Thus, in general the incident and scattered Stokes vectors are *not* referenced to the same plane. In addition, the volume scattering function takes its simplest form when the Stokes vectors are referenced to the scattering plane — the plane formed by $\hat{\xi}$ and $\hat{\xi}_0$. Thus, in describing the scattering process there are three reference planes involved: that formed by \hat{e}_z and $\hat{\xi}_0$, that formed by $\hat{\xi}_0$ and $\hat{\xi}$, and that formed by $\hat{\xi}$ and \hat{e}_z . Figure 2.18 shows the scattering process projected onto the surface of a sphere centered on the scattering event, and Figure 2.17 shows the spherical triangle projected on a plane tangent to the sphere to indicate clearly the reference directions for describing the polarizations of the incident and scattered radiation. Let $\mathbf{P}(\Theta)$ be the Mueller scattering phase matrix

$$\mathbf{P}(\Theta) = 4\pi \frac{\beta(\Theta)}{b},$$

e.g., for scattering by a collection of atoms (*Rayleigh scattering*),

$$\mathbf{P}(\Theta) = \frac{3}{4} \begin{pmatrix} (1 + \cos^2 \Theta) & -\sin^2 \Theta & 0 & 0 \\ -\sin^2 \Theta & (1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & 2 \cos \Theta & 0 \\ 0 & 0 & 0 & 2 \cos \Theta \end{pmatrix}. \quad (2.110)$$

In this form of $\mathbf{P}(\Theta)$ the incident and scattered Stokes vectors are referenced to the scattering plane, i.e., the vectors \hat{r}_{sp} and $\hat{\ell}_{sp}$ in Figure 2.17.

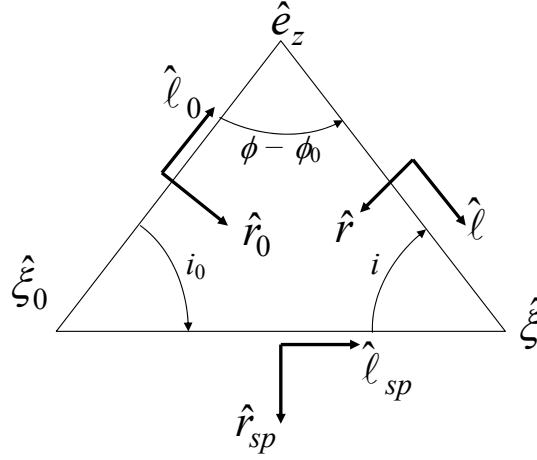


Figure 2.17: This is a planar projection of the spherical triangle that describes the scattering process. Shown are the directions of the reference vectors \hat{r}_0 and $\hat{\ell}_0$ for the incident photon ($\hat{\xi}_0$), \hat{r} and $\hat{\ell}$ for the scattered photon ($\hat{\xi}$), and \hat{r}_{sp} and $\hat{\ell}_{sp}$ for the scattering plane. (See Figure 2.18 for further explanation of the directions and angles.)

Since the incident Stokes vector is referenced to \hat{r}_0 and $\hat{\ell}_0$ in the figure, we must transform the reference for $\mathbf{I}(0, u_0, \phi_0) \triangleq \mathbf{I}_0$ to the scattering plane. Figure 2.17 shows that the transformation from \hat{r}_0 to \hat{r}_{sp} requires a *clockwise* rotation of the former through an angle i_0 . This is effected by $\mathbf{R}(-i_0)$, where (Eq. (1.58))

$$\mathbf{R}(\psi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\psi & -\sin 2\psi & 0 \\ 0 & \sin 2\psi & \cos 2\psi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The Stokes vector of the incident radiation referenced to the scattering plane (\mathbf{I}_{sp}) is

$$\mathbf{I}_{sp} = \mathbf{R}(-i_0)\mathbf{I}_0,$$

and the scattered Stokes vector referenced to the scattering plane is

$$\mathbf{P}(\Theta)\mathbf{R}(-i_0)\mathbf{I}_0.$$

But we want the scattered Stokes vector referenced to the plane formed by $\hat{\xi}$ and \hat{e}_z . To get this we must perform a second rotation (\hat{r}_{sp} clockwise through i to \hat{r}), resulting in the final form for the scattered Stokes vector

$$\mathbf{R}(-i)\mathbf{P}(\Theta)\mathbf{R}(-i_0)\mathbf{I}_0 = \mathbf{Z}(u_0 \rightarrow u, \phi_0 \rightarrow \phi)\mathbf{I}_0.$$

This also provides \mathbf{Z} in terms of \mathbf{P} , i.e.,

$$\mathbf{Z}(u_0 \rightarrow u, \phi_0 \rightarrow \phi) = \mathbf{R}(-i)\mathbf{P}(\Theta)\mathbf{R}(-i_0). \quad (2.111)$$

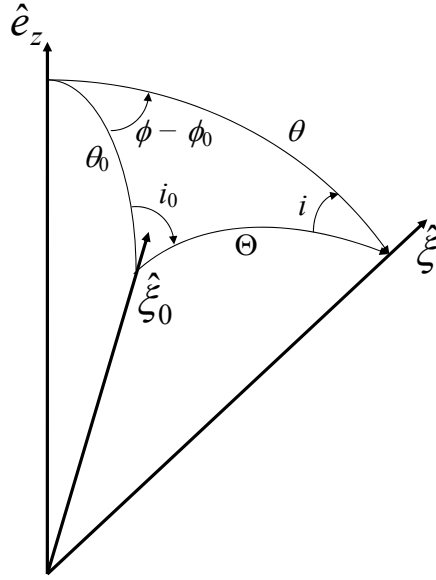


Figure 2.18: Spherical coordinate diagram for scattering from $\hat{\xi}_0$ to $\hat{\xi}$. The scattering angle is Θ , and the angles i_0 and i are the angles between the incident photon's reference plane (plane formed by \hat{e}_z and $\hat{\xi}_0$) and the scattering plane (plane formed by $\hat{\xi}_0$ and $\hat{\xi}$), and the scattered photon's reference plane (plane formed by \hat{e}_z and $\hat{\xi}$) and the scattering plane, respectively.

It remains to determine the angles i_0 and i in terms of the known directions $\hat{\xi}_0$ and $\hat{\xi}$ (or θ_0, ϕ_0, θ , and ϕ). Applying the law of cosines and the law of sines to the spherical triangle in Figure 2.18 provides

$$\begin{aligned}\cos \Theta &= \cos \theta_0 \cos \theta + \sin \theta_0 \sin \theta \cos(\phi - \phi_0), \\ \cos i_0 &= \frac{\cos \theta - \cos \Theta \cos \theta_0}{\sin \Theta \sin \theta_0}, \\ \cos i &= \frac{\cos \theta_0 - \cos \Theta \cos \theta}{\sin \Theta \sin \theta}, \\ \sin i_0 &= \frac{\sin \theta \sin(\phi - \phi_0)}{\sin \Theta}, \\ \sin i &= \frac{\sin \theta_0 \sin(\phi - \phi_0)}{\sin \Theta}.\end{aligned}\tag{2.112}$$

Note, the sign of the i 's is the same as the sign of $\phi - \phi_0$.

Clearly, when polarization is included the resulting equations become very complicated, even in the first-order approximation. In the case examined here, the incident beam is unpolarized, and the results simplify because only the first column of \mathbf{Z} is required (and also the reference plane for the incident beam is irrelevant because the incident irradiance is unpolarized). When the Rayleigh phase matrix describes the scattering, the equations simplify to

$$\begin{aligned}\mathbf{I}(0, u, \phi) &= -\frac{3\omega_0 F_0 \tau_1}{16\pi u} \begin{pmatrix} 1 + \cos^2 \Theta \\ \cos(-2i) \sin^2 \Theta \\ \sin(-2i) \sin^2 \Theta \\ 0 \end{pmatrix}, \quad u < 0; \\ \mathbf{I}(\tau_1, u, \phi) &= \delta(u - u_0) \delta(\phi - \phi_0) \left[1 - \frac{\tau_1}{u} \right] \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ &\quad + \frac{3\omega_0 F_0 \tau_1}{16\pi u} \begin{pmatrix} 1 + \cos^2 \Theta \\ \cos(-2i) \sin^2 \Theta \\ \sin(-2i) \sin^2 \Theta \\ 0 \end{pmatrix}, \quad u > 0.\end{aligned}$$

The degree of polarization of the diffuse light is

$$P = \sqrt{\frac{Q^2 + U^2}{I^2}} = \frac{\sin^2 \Theta}{1 + \cos^2 \Theta},$$

so in first-order the radiation is partially polarized. Comparison of these solutions with Eq. (2.102) and (2.103) for the scalar case shows that the top component of the Stokes

vector \mathbf{I} , i.e., I is identical to the radiance in the scalar approximation (after replacing P by the Rayleigh phase function in the latter). It is not always true that in the first-order approximation the radiance predicted by the scalar (L) and the vector (\mathbf{I}) radiative transfer equations are identical for an incident beam of unpolarized radiation. For example, it does *not* hold if the lower boundary is a Fresnel reflecting surface, e.g., a water surface (cf. Chapter 9, Eq. (9.40)).

The reader should note that Eq. (2.111) holds of any incident direction $\hat{\xi}'$, not just $\hat{\xi}_0$, this means that the radiative transfer equation for a 1-d medium (with \mathbf{c} a multiple of the identity matrix) can always be written in terms of the phase matrix referenced to the scattering plane in the following manner;

$$u \frac{\partial \mathbf{I}(\tau, u, \phi)}{\partial \tau} + \mathbf{I}(\tau, u, \phi) = \frac{\omega_0(\tau)}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \mathbf{R}(-i) \mathbf{P}(\tau, \Theta) \mathbf{R}(-i') \mathbf{I}(\tau, u', \phi'), \quad (2.113)$$

where Θ and i can be determined from Eqs. (2.112) with i_0 , θ_0 , and ϕ_0 , replaced by i' , θ' , and ϕ' , respectively.

2.12 Appendix 2: Monte Carlo from a Multidimensional Integral for Radiance

It is possible to write the solution to any radiative transfer problem in an elastically-scattering medium as a multidimensional integral. In this appendix we derive the integral for the radiance in any direction in a medium given the irradiance incident on the boundaries. For simplicity and for application to systems of interest here, we shall consider the geometry encountered in the ocean-atmosphere system: a medium with a flat surface (or one smoothly varying in shape) illuminated at the boundary by a source of essentially collimated irradiance. After derivation of the multidimensional integral, we will proceed to evaluate it using three techniques: (1) a straightforward, but naive, application of Monte Carlo methods as described in the Mathematical Appendix, (2) a forward Monte Carlo evaluation that, for a horizontally-homogeneous medium and source, is identical to that derived by heuristic means in the text and is important when the IOPs depend only on a single spatial coordinate, e.g., depth in the medium, and (3) a backward Monte Carlo evaluation that is very important for media in which the IOPs depend on three-dimensional position within the medium.

2.12.1 Radiance in a Medium as a Multidimensional Integral

Consider a detector of area ΔA_D accepting radiant power $\Delta^2 P_D(\vec{r}_D, \hat{\xi}_D)$ from the direction $\hat{\xi}_D$ located at the position \vec{r}_D as shown in Figure 2.19. The field of view of the detector is restricted to $\Delta\Omega_D$ by a Gershun tube as shown. Radiant power $\Delta^2 P'(\vec{r}', \hat{\xi}_D)$ is shown

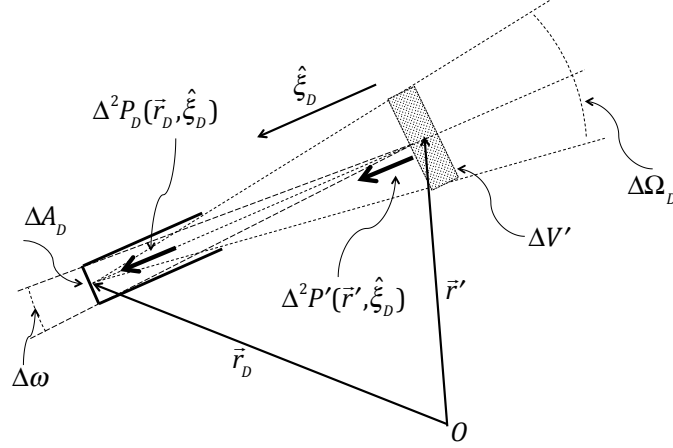


Figure 2.19: Radiation from sources within $\Delta V'$ propagating in the direction $\hat{\xi}_D$ within a medium is detected by a radiometer — a detector of area ΔA_D with its field of view limited by a Gershun tube to $\Delta\Omega_D$.

leaving an element of volume $\Delta V'$, located at \vec{r}' within the detector's field of view, propagating in the $\hat{\xi}_D$ direction. The detector subtends a solid angle $\Delta\omega$ seen by the volume element. Let $\mathcal{I}(\vec{r}', \hat{\xi}_D) \triangleq dI(\vec{r}', \hat{\xi}_D)/dV'$, where $I(\vec{r}', \hat{\xi}_D)$ is the intensity associated with $\Delta^2 P'(\vec{r}', \hat{\xi}_D)$, be the intensity density of sources within $\Delta V'$. Then,

$$\Delta^2 P'(\vec{r}', \hat{\xi}_D) = \mathcal{I}(\vec{r}', \hat{\xi}_D) \Delta\omega \Delta V', \quad \text{where} \quad \Delta\omega = \frac{\Delta A_D}{|\vec{r}' - \vec{r}_D|^2}.$$

The radiance $\Delta L_D(\vec{r}_D, \hat{\xi}_D)$ at the detector due to the sources within $\Delta V'$ is

$$\begin{aligned} \Delta L_D(\vec{r}_D, \hat{\xi}_D) &= \frac{\Delta^2 P_D(\vec{r}_D, \hat{\xi}_D)}{\Delta A_D \Delta\Omega_D} = \frac{\exp[-c|\vec{r}' - \vec{r}_D|]}{\Delta A_D \Delta\Omega_D} \Delta^2 P'(\vec{r}', \hat{\xi}_D) \\ &= \frac{\exp[-c|\vec{r}' - \vec{r}_D|]}{|\vec{r}' - \vec{r}_D|^2 \Delta\Omega_D} \mathcal{I}(\vec{r}', \hat{\xi}_D) \Delta V', \end{aligned}$$

or

$$\Delta L_D(\vec{r}_D, \hat{\xi}_D) \Delta \Omega_D = \frac{\exp[-c|\vec{r}' - \vec{r}_D|]}{|\vec{r}' - \vec{r}_D|^2} \mathcal{I}(\vec{r}', \hat{\xi}_D) \Delta V'.$$

Here, we have assumed that c is constant along the path between \vec{r}' and \vec{r}_D . If it is not, then the argument of the exponential is

$$- \int_{\vec{r}'}^{\vec{r}_D} c d|\vec{r}' - \vec{r}_D|$$

along the straight-line path from \vec{r}' to \vec{r}_D .

Now assume that $\mathcal{I}(\vec{r}', \hat{\xi}_D)$ is due to scattering within $\Delta V'$ from some other direction as shown in Figure 2.20. From the definition of $\beta(\hat{\xi}' \rightarrow \hat{x}_D)$,

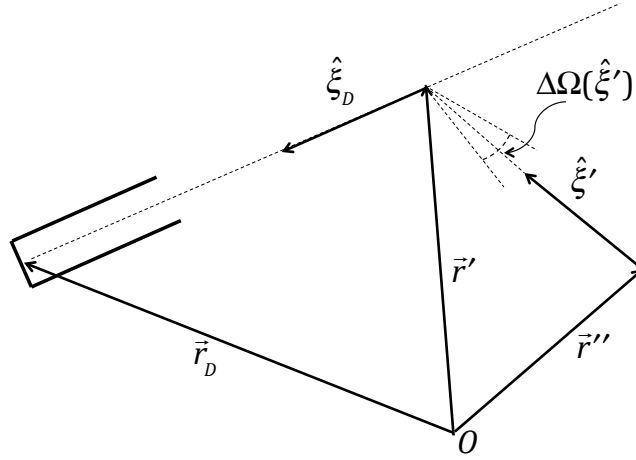


Figure 2.20: The intensity density of sources within $\Delta V'$ in Figure 2.19 results from light scattered into $\Delta V'$ from $\hat{\xi}'$.

$$\beta(\hat{\xi}' \rightarrow \hat{x}) = \frac{\mathcal{I}(\vec{r}', \hat{\xi}_D)}{E(\perp \hat{\xi}')} ,$$

where $E(\perp \hat{\xi}')$ is the source irradiance on a plane perpendicular to $\hat{\xi}'$ which in this case is just $L(\vec{r}', \hat{\xi}') \Delta \Omega(\hat{\xi}')$. Therefore,

$$\mathcal{I}(\vec{r}', \hat{\xi}_D) = \beta(\vec{r}'; \hat{\xi}' \rightarrow \hat{\xi}_D) L(\vec{r}', \hat{\xi}') \Delta \Omega(\hat{\xi}'), \quad (2.114)$$

so

$$\Delta L_D(\vec{r}_D, \hat{\xi}_D) \Delta \Omega_D = \frac{\exp[-c|\vec{r}' - \vec{r}_D|]}{|\vec{r}' - \vec{r}_D|^2} \beta(\vec{r}'; \hat{\xi}' \rightarrow \hat{\xi}_D) L(\vec{r}', \hat{\xi}') \Delta \Omega(\hat{\xi}'), \Delta V'. \quad (2.115)$$

But, in the same manner

$$L(\vec{r}', \hat{\xi}') \Delta \Omega(\hat{\xi}') = \frac{\exp[-c|\vec{r}'' - \vec{r}'|]}{|\vec{r}'' - \vec{r}'|^2} \beta(\vec{r}''; \hat{\xi}'' \rightarrow \hat{\xi}') L(\vec{r}'', \hat{\xi}'') \Delta \Omega(\hat{\xi}''), \Delta V'',$$

⋮

Continuing back to the first collision at which the incident solar beam enters the medium at \vec{r}_0 propagating in the $\hat{\xi}_1$ direction and then is scattered in the $\hat{\xi}_2$ direction, we have (Figure 2.21)

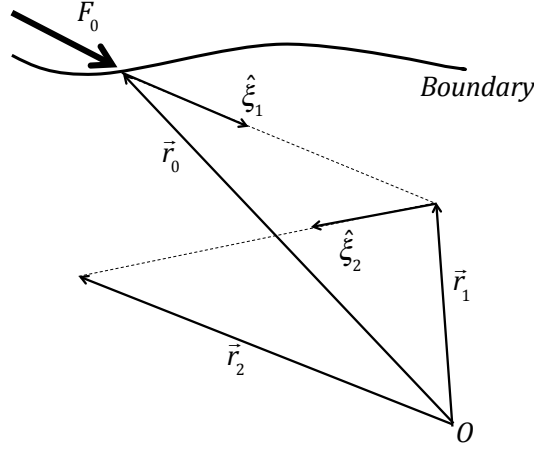


Figure 2.21: Irradiance F_0 in the form of a parallel beam enters the medium traveling in the direction $\hat{\xi}_1$. It is scattered (the first scattering) at \vec{r}_1 into the direction $\hat{\xi}_2$ propagating toward \vec{r}_2 .

$$L(\vec{r}_1, \hat{\xi}_1) \Delta \Omega(\hat{\xi}_1) = F_0(\vec{r}_0) \exp[-c|\vec{r}_1 - \vec{r}_0|],$$

where the incident light is assumed to be in the form of a beam whose irradiance may vary across the beam profile. This yields

$$L(\vec{r}_2, \hat{\xi}_2) \Delta \Omega(\hat{\xi}_2) = \frac{\exp[-c|\vec{r}_2 - \vec{r}_1|]}{|\vec{r}_2 - \vec{r}_1|^2} \beta(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2) F_0(\vec{r}_0) \exp[-c|\vec{r}_1 - \vec{r}_0|] \Delta V_1,$$

Thus, the contribution to the detector from the volume elements $\Delta V_1, \Delta V_2, \dots, \Delta V_N$, where N is the number of collisions, is

$$\begin{aligned} \Delta L_D(\vec{r}_D, \hat{\xi}_D) \Delta \Omega_D &= \frac{\exp[-c|\vec{r}_2 - \vec{r}_1|]}{|\vec{r}_2 - \vec{r}_1|^2} \beta(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2) F_0(\vec{r}_0) \exp[-c|\vec{r}_1 - \vec{r}_0|] \Delta V_1 \\ &\times \frac{\exp[-c|\vec{r}_3 - \vec{r}_2|]}{|\vec{r}_3 - \vec{r}_2|^2} \beta(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3) \Delta V_2 \\ &\vdots \\ &\times \frac{\exp[-c|\vec{r}_N - \vec{r}_{N-1}|]}{|\vec{r}_N - \vec{r}_{N-1}|^2} \beta(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N) \Delta V_{N-1} \\ &\times \frac{\exp[-c|\vec{r}_D - \vec{r}_N|]}{|\vec{r}_D - \vec{r}_N|^2} \beta(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D) \Delta V_N \end{aligned}$$

The final result for the radiance at the detector due to N collisions is achieved by summing over all of the volume elements. However, this is only for N collisions; we also have to sum over all N , i.e., sum over the individual total number of collisions from $N = 1 \rightarrow \infty$. Thus, converting ΔV_i to $dV(\vec{r}_i)$ and the sums over the ΔV 's to $3N$ integrals, we get

$$\begin{aligned} L_D(\vec{r}_D, \hat{\xi}_D) \Delta \Omega_D &= \sum_{N=1}^{\infty} \int \dots \int \frac{\exp[-c|\vec{r}_2 - \vec{r}_1|]}{|\vec{r}_2 - \vec{r}_1|^2} \beta(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2) F_0(\vec{r}_0) \exp[-c|\vec{r}_1 - \vec{r}_0|] dV(\vec{r}_1) \\ &\times \frac{\exp[-c|\vec{r}_3 - \vec{r}_2|]}{|\vec{r}_3 - \vec{r}_2|^2} \beta(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3) dV(\vec{r}_2) \\ &\vdots \\ &\times \frac{\exp[-c|\vec{r}_N - \vec{r}_{N-1}|]}{|\vec{r}_N - \vec{r}_{N-1}|^2} \beta(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N) dV(\vec{r}_{N-1}) \\ &\times \frac{\exp[-c|\vec{r}_D - \vec{r}_N|]}{|\vec{r}_D - \vec{r}_N|^2} \beta(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D) dV(\vec{r}_N), \end{aligned}$$

where the last integral is zero *unless* the point \vec{r}_N is within the field of view of the detector, i.e., within $\Delta \Omega_D$. If the medium is inhomogeneous, then $c \rightarrow c(\vec{r})$, in which case $c|\vec{r}_{i+1} - \vec{r}_i| \rightarrow \tau(i, i+1)$, where

$$\tau(i, i+1) = \int_{\vec{r}_i}^{\vec{r}_{i+1}} c(\vec{r}) dr, \quad (2.116)$$

and \vec{r} is on the straight line between \vec{r}_i and \vec{r}_{i+1} .

Now, we introduce the scattering phase function P through

$$\frac{\beta(\vec{r}; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{b(\vec{r})} = \frac{P(\vec{r}; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi},$$

or

$$\beta(\vec{r}; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1}) = \omega_0(\vec{r})c(\vec{r})\frac{P(\vec{r}; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi}.$$

We note for future reference that $P(\vec{r}; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})/4\pi$ is the *probability density* for scattering from $\hat{\xi}_i \rightarrow \hat{\xi}_{i+1}$ at the point \vec{r} . With these modifications, the radiance at the detector is

$$\begin{aligned} L_D(\vec{r}_D, \hat{\xi}_D)\Delta\Omega_D &= \sum_{N=1}^{\infty} \int \cdots \int \frac{\exp[-\tau(1, 2)]}{|\vec{r}_2 - \vec{r}_1|^2} \omega_0(\vec{r}_1)c(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} F_0(\vec{r}_0) \exp[-\tau(0, 1)] dV(\vec{r}_1) \\ &\quad \times \frac{\exp[-\tau(2, 3)]}{|\vec{r}_3 - \vec{r}_2|^2} \omega_0(\vec{r}_2)c(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} dV(\vec{r}_2) \\ &\quad \vdots \\ &\quad \times \frac{\exp[-\tau(N-1, N)]}{|\vec{r}_N - \vec{r}_{N-1}|^2} \omega_0(\vec{r}_{N-1})c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} dV(\vec{r}_{N-1}) \\ &\quad \times \frac{\exp[-\tau(N, D)]}{|\vec{r}_D - \vec{r}_N|^2} \omega_0(\vec{r}_N)c(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} dV(\vec{r}_N), \end{aligned} \tag{2.117}$$

where the last term contributes only if \vec{r}_N is within the field of view of the detector.

2.12.2 Naive Monte Carlo Evaluation of the Integral

This multi-dimensional integral can be evaluated in a straightforward manner by Monte Carlo techniques. Assume that the radiative transfer process takes place in a huge box of dimension L in all cartesian directions. We simply pick $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$, which provides $\hat{\xi}_2, \hat{\xi}_3, \dots, \hat{\xi}_N$, from a uniform distribution.⁴⁶ The estimator of the integral for this (single)

⁴⁶This is accomplished by choosing x_i, y_i and z_i from uniform distributions on the interval $0 \rightarrow L$.

trial is then

$$\begin{aligned}
Est_1(N) &= F_0(\vec{r}_0) \exp[-\tau(0, 1)] \\
&\times \frac{\exp[-\tau(1, 2)]}{|\vec{r}_2 - \vec{r}_1|^2} \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} \\
&\times \frac{\exp[-\tau(2, 3)]}{|\vec{r}_3 - \vec{r}_2|^2} \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} \\
&\vdots \\
&\times \frac{\exp[-\tau(N-1, N)]}{|\vec{r}_N - \vec{r}_{N-1}|^2} \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} \\
&\times \frac{\exp[-\tau(N, D)]}{|\vec{r}_D - \vec{r}_N|^2} \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi}.
\end{aligned}$$

Compute $Est_1(N)$ for all N ,⁴⁷ and carry this out for \mathcal{N} trials. The Monte Carlo estimate is then

$$L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \sum_{N=1}^{\infty} Est_i(N).$$

This estimate would likely be terrible because all of the singular terms of the integrand, e.g., $|\vec{r}_D - \vec{r}_N|^{-2}$, etc., would cause a large variation in the individual estimates, causing a large variance in the final result.

2.12.3 Forward Monte Carlo Evaluation of the Integral

We can simplify the integrations in Eq. (2.117), and remove some of the variance, by rewriting the integrals centering the origin for integration over $dV(\vec{r}_i)$ on \vec{r}_{i-1} , i.e.,⁴⁸

$$dV(\vec{r}_i) = |\vec{r}_i - \vec{r}_{i-1}|^2 d|\vec{r}_i - \vec{r}_{i-1}| d\Omega(\hat{\xi}_i).$$

This substitution causes many of the singular terms in the denominators in Eq. (2.117) to cancel out, e.g., consider the first two triple integrals. We let $dV(\vec{r}_2) = |\vec{r}_2 - \vec{r}_1|^2 d|\vec{r}_2 - \vec{r}_1| d\Omega(\hat{\xi}_1)$, so the $|\vec{r}_2 - \vec{r}_1|$ in $dV(\vec{r}_2)$ will cancel the $|\vec{r}_2 - \vec{r}_1|^{-1}$ term in the first set of triple

⁴⁷This can be done with a single set $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$ (see Section 2.12.3 for the details).

⁴⁸Note that this is just a generalization of the volume element in a three-dimensional coordinate system centered on the origin, i.e., $dV = r^2 dr \sin \theta d\theta d\phi$, where r is the distance from the origin, θ and ϕ are the polar and azimuth angles, and $\sin \theta d\theta d\phi$ is the element of solid angle $d\Omega$.

$$\begin{aligned}
L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D &= \sum_{N=1}^{\infty} \int F_0(\vec{r}_0) |\hat{\xi}_1 \bullet \hat{n}_S(\vec{r}_0)| dA_S(\vec{r}_0) \\
&\times \int \cdots \int \\
&\times \exp[-\tau(0, 1)] \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} d\Omega(\hat{\xi}_2) d|\vec{r}_1 - \vec{r}_0| \\
&\times \exp[-\tau(1, 2)] \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} d\Omega(\hat{\xi}_3) d|\vec{r}_2 - \vec{r}_1| \\
&\vdots \\
&\times \exp[-\tau(i-1, i)] \omega_0(\vec{r}_i) c(\vec{r}_i) \frac{P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi} d\Omega(\hat{\xi}_{i+1}) d|\vec{r}_i - \vec{r}_{i-1}| \\
&\vdots \\
&\times \exp[-\tau(N-2, N-1)] \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} d\Omega(\hat{\xi}_N) d|\vec{r}_{N-1} - \vec{r}_{N-2}| \\
&\times \frac{\exp[-\tau(N-1, N)]}{|\vec{r}_D - \vec{r}_N|^2} \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} d|\vec{r}_N - \vec{r}_{N-1}| \exp[-\tau(N, D)],
\end{aligned} \tag{2.119}$$

where, again, the point \vec{r}_N must be in the field of view of the detector. This is another form of the multiple dimensional-integral solution to the radiative transfer problem of providing the radiance at a detector within the medium resulting from a parallel beam (propagating in the direction $\hat{\xi}_1$) of irradiance F_0 (on a plane normal to the direction $\hat{\xi}_1$ and variable with position) incident on the boundary of the medium. The various integrals cover the full range of the variables provided by the boundary, i.e., in distance, area, and solid angle.

To proceed with a Monte Carlo evaluation of this multi-dimensional integral, the following observations are important. First, the quantity $P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1}) d\Omega(\hat{\xi}_{i+1})/4\pi$ is the probability of scattering from $\hat{\xi}_i$ to $\hat{\xi}_{i+1}$ within $d\Omega(\hat{\xi}_{i+1})$ at the point \vec{r}_i , and as such, the probability density for scattering into $d\Omega(\hat{\xi}_{i+1})$ is $P/4\pi$. If the medium were homogeneous, $\exp[-\tau(i-1, i)] c(\vec{r}_i) d|\vec{r}_i - \vec{r}_{i-1}| = \exp[-c|\vec{r}_i - \vec{r}_{i-1}|] c d|\vec{r}_i - \vec{r}_{i-1}|$, which is the probability of an interaction within $d|\vec{r}_i - \vec{r}_{i-1}|$. Thus, $c \exp[-\tau(i-1, i)]$ is the probability density for an interaction within $d|\vec{r}_i - \vec{r}_{i-1}|$. If c is not constant, the result is still true, but the appropriate τ is given by Eq. (2.116). We can now evaluate the integrals using Monte Carlo techniques. Consider first the integral

$$\int \int \int \exp[-\tau(i-1, i)] \omega_0(\vec{r}_i) c(\vec{r}_i) \frac{P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi} d\Omega(\hat{\xi}_{i+1}) d|\vec{r}_i - \vec{r}_{i-1}|.$$

Given \vec{r}_{i-1} and $\hat{\xi}_i$, we repeatedly sample $\hat{\xi}_{i+1}$ from the density $P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})/4\pi$, and sample \vec{r}_i from $c(\vec{r}_i) \exp[-\tau(i-1, i)]$. For a given sample, the estimator for the integral is just $\omega_0(\vec{r}_i)$ (see the Mathematical Appendix). Thus, for the set of integrals, if we sample $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$ and $\hat{\xi}_2, \hat{\xi}_3, \dots, \hat{\xi}_N$, in a similar manner, and sample the point on the surface \vec{r}_0 from the appropriate distribution (e.g., the point \vec{r}_0 might be uniformly distributed in area as for solar illumination, or distributed over a small area as in laser illumination), then for the given set of samples the estimator for exactly N interaction events is

$$F_0(\vec{r}_0) | \hat{\xi}_1 \bullet \hat{n}_S(\vec{r}_0) | \omega_0(\vec{r}_1) \omega_0(\vec{r}_2) \cdots \omega_0(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi |\vec{r}_D - \vec{r}_N|^2} \exp[-\tau(N, D)], \quad (2.120)$$

if \vec{r}_N is within the field of view of the detector and zero otherwise. Note, that this is the estimator for a single N .

How does one actually implement this estimation? According to the integral we are evaluating (Eq. (2.119)) we must perform a sum over N . One way to do this is to first pick $N = 1$ then, for say \mathcal{N} trials (or photons), compute \vec{r}_0, \vec{r}_1 and $\hat{\xi}_2$ sampling from the appropriate distributions (see text). If the point \vec{r}_D is within the field of view of the detector, then evaluate the estimator:

$$F_0(\vec{r}_0) | \hat{\xi}_1 \bullet \hat{n}_S(\vec{r}_0) | \omega_0(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_D)}{4\pi |\vec{r}_D - \vec{r}_1|^2} \exp[-\tau(1, D)]. \quad (2.121)$$

Do this \mathcal{N} times, add the values of all of the individual estimators, and divide the sum by \mathcal{N} (the number of trials). The result is an estimation of $L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D$ due to *single* scattering. Then pick $N = 2$ and for \mathcal{N} trials pick $\vec{r}_0, \vec{r}_1, \vec{r}_2, \hat{\xi}_2$, and $\hat{\xi}_3$, sampling from the appropriate distributions. If the point \vec{r}_D is within the field of view of the detector, then evaluate the estimator:

$$F_0(\vec{r}_0) | \hat{\xi}_1 \bullet \hat{n}_S(\vec{r}_0) | \omega_0(\vec{r}_1) \omega_0(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_D)}{4\pi |\vec{r}_D - \vec{r}_2|^2} \exp[-\tau(2, D)], \quad (2.122)$$

sum the values of the individual estimators, and divide by \mathcal{N} , yielding an estimation of $L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D$ due to *double* scattering. Continue this process for $N = 3, \dots, \infty$: however, note that the product of the ω_0 's will cause the estimation of $L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D$ to decrease rapidly with increasing N (unless ω_0 is close to unity). When N is sufficiently large, the contribution of the term with N scatterings will be too small to contribute significantly and the process stops. Finally, sum the estimators for $N = 1$ to the final N to yield the Monte Carlo estimate to $L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D$.

There is however, a more straightforward method: accumulate the estimator sequentially. This proceeds as follows. Carry out the process as above for $N = 1$ with a single trial and

evaluate the single-scattering estimator in Eq. (2.121). Then, rather than starting over with a new photon, continue on the path from \vec{r}_1 by choosing \vec{r}_2 and $\hat{\xi}_3$ from their appropriate distributions. Now evaluate the double-scattering estimator using Eq. (2.122) and add it to the single-scattering estimate just evaluated. Continue the process until $\prod_{i=1}^M \omega_0(\vec{r}_i)$ becomes so small that the M^{th} scattering contributes a negligible amount to the result. At this point stop and start the chain over continuing to accumulate the contribution to $L_D(\vec{r}_D, \hat{\xi}_D) \Delta \Omega_D$ from the previous chain of events. Do this \mathcal{N} times and divide the estimate by \mathcal{N} . The result will be identical to that described above. With both procedures, the uncertainty in the result is proportional to $\mathcal{N}^{-1/2}$, i.e., to decrease the uncertainty by a factor of ten requires increasing \mathcal{N} by a factor of 100. One may of course place many detectors in the medium with the individual detectors being updated at each collision. It is important to note that the singular denominator $|\vec{r}_D - \vec{r}_N|^{-2}$ in the various estimators can become very large. This can cause large variance in the results in the general case where all IOPs depend on \vec{r} — the so-called 3-d case.

What if the medium plus source display horizontal invariance, i.e., $\beta(\vec{r}, \hat{\xi} \rightarrow \hat{\xi}') = \beta(z, \hat{\xi} \rightarrow \hat{\xi}')$, $F_0(\vec{r}_0) = F_0(z_0)$, $\hat{n}_S(\vec{r}_0) = \hat{n}_S(z_0) = \hat{e}_z$, etc.? In this case (called 1-d radiative transfer), $L(\vec{r}, \hat{\xi}) = L(z, \hat{\xi})$, and considerable simplification takes place. The most important is in the computation of the contribution to the detector from the last scattering event. In Figure 2.22 we have redrawn Figure 2.19 to reflect the invariance. In the final result, the radiance over the entire surface at $z = z_D$ must be constant and the radiance at $z = z'$ must be constant. The radiances leaving z' and arriving at z_D are

$$L(z', \hat{\xi}_D) = \frac{\Delta^2 P(z', \hat{\xi}_D)}{|\hat{e}_z \bullet \hat{\xi}_D| \Delta A_H(z') \Delta \omega} \quad \text{and} \quad L(z_D, \hat{\xi}_D) = \frac{\Delta^2 P(z_D, \hat{\xi}_D)}{|\hat{e}_z \bullet \hat{\xi}_D| \Delta A_H(z_D) \Delta \Omega_D}$$

If there were no losses, the numerators would be the same, so if we take $A_H(z) = A_H(z')$ for simplicity, we must have $\Delta \omega = \Delta \Omega_D$. Then in the presence of losses (the actual case)

$$L(z_D, \hat{\xi}_D) = \frac{\Delta^2 P(z', \hat{\xi}_D)}{|\hat{e}_z \bullet \hat{\xi}_D| \Delta A_H(z_D) \Delta \Omega_D} \exp[-c|z_D - z'|/|\hat{e}_z \bullet \hat{\xi}_D|]$$

or

$$\Delta L(z_D, \hat{\xi}_D) \Delta A_H(z_D) = \frac{\mathcal{I}(z', \hat{\xi}_D) \Delta V'}{|\hat{e}_z \bullet \hat{\xi}_D|} \exp[-c|z_D - z'|/|\hat{e}_z \bullet \hat{\xi}_D|].$$

Substituting for $\mathcal{I}(z', \hat{\xi}_D)$ from Eq. (2.114) yields

$$\Delta L(z_D, \hat{\xi}_D) \Delta A_H(z_D) = \frac{\exp[-c|z_D - z'|/|\hat{e}_z \bullet \hat{\xi}_D|]}{|\hat{e}_z \bullet \hat{\xi}_D|} \beta(\vec{r}'; \hat{\xi}' \rightarrow \hat{\xi}_D) L(\vec{r}', \hat{\xi}') \Delta \Omega(\hat{\xi}') \Delta V',$$

and we see that the singular term, $|\vec{r}_D - \vec{r}_N|^{-2}$ in Eq. (2.115) no longer appears. Completing

the chain back to the first scattering as in Eq. (2.118), we get

$$\begin{aligned}
L_D(\vec{r}_D, \hat{\xi}_D) \Delta A_H(z_D) &= F_0 \frac{|\hat{e}_z \bullet \hat{\xi}_1|}{|\hat{e}_z \bullet \hat{\xi}_D|} \sum_{N=1}^{\infty} \int dA_S \\
&\times \int \cdots \int \\
&\times \exp[-\tau(0, 1)] \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} d\Omega(\hat{\xi}_2) d|\vec{r}_1 - \vec{r}_0| \\
&\times \exp[-\tau(1, 2)] \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} d\Omega(\hat{\xi}_3) d|\vec{r}_2 - \vec{r}_1| \\
&\vdots \\
&\times \exp[-\tau(i-1, i)] \omega_0(\vec{r}_i) c(\vec{r}_i) \frac{P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi} d\Omega(\hat{\xi}_{i+1}) d|\vec{r}_i - \vec{r}_{i-1}| \\
&\vdots \\
&\times \exp[-\tau(N-2, N-1)] \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} d\Omega(\hat{\xi}_N) d|\vec{r}_{N-2} - \vec{r}_{N-1}| \\
&\times \exp[-\tau(N-1, N)] \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} d|\vec{r}_N - \vec{r}_{N-1}| \exp[-\tau(N, D)],
\end{aligned} \tag{2.123}$$

Picking $\vec{r}_2, \dots, \vec{r}_N$ and $\hat{\xi}_2, \dots, \hat{\xi}_N$ as before, the estimator for a single set is

$$L_D(\vec{r}_D, \hat{\xi}_D) \Delta A_H(z_D) = F_0 \frac{|\hat{e}_z \bullet \hat{\xi}_1|}{|\hat{e}_z \bullet \hat{\xi}_D|} \int dA_S \sum_{N=1}^{\infty} \left(\prod_{i=1}^N \omega_0(\vec{r}_i) \right) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} \exp[-\tau(N, D)].$$

Note that ΔA_H is completely arbitrary, so we choose it to be the integral of dA_S .⁴⁹ Then the estimator becomes

$$\frac{L_D(\vec{r}_D, \hat{\xi}_D)}{F_0} = \frac{|\hat{e}_z \bullet \hat{\xi}_1|}{|\hat{e}_z \bullet \hat{\xi}_D|} \sum_{N=1}^{\infty} \left(\prod_{i=1}^N \omega_0(\vec{r}_i) \right) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} \exp[-\tau(N, D)], \tag{2.124}$$

We can perform the computation in either of the two ways described for the 3-d case. The simplest is to follow the second one: trace the path of a photon through the medium and at each collision, e.g., the n^{th} , update the estimator of L_D/F_0 for a particular z_D and $\hat{\xi}_D$ by

$$\frac{|\hat{e}_z \bullet \hat{\xi}_1|}{|\hat{e}_z \bullet \hat{\xi}_D|} \left(\prod_{i=1}^n \omega_0(\vec{r}_i) \right) \frac{P(\vec{r}_n; \hat{\xi}_n \rightarrow \hat{\xi}_D)}{4\pi} \exp[-\tau(n, D)],$$

⁴⁹Actually, both $\int dA_s$ and ΔA_H can be taken to be the full (horizontal) plane.

and move on to the next collision.⁵⁰ With the removal of the singularity, the variance of the estimate is significantly lower for the 1-d compare with the 3-d case.

One feature of the forward Monte Carlo that is particularly appealing is the fact that we can solve the problem for the radiance at *an arbitrary number* of internal points at the same time. To effect this, at each collision we simply update the estimate *for each detector*. An unappealing feature (in a 3-d medium) is the fact that the detectors only receive contributions when the point \vec{r}_i is within the field of view of the detector.

2.12.4 Monte Carlo Solution for Irradiances

Is there a way to estimate the various irradiances using Monte Carlo? The answer is yes, and in fact it is quite simple. Consider Eq. (2.123). Suppose we want to compute the following integral

$$X(\vec{r}) = \int_{\Omega_X} f(\vec{r}, \hat{\xi}) L(\vec{r}, \hat{\xi}) d\Omega(\hat{\xi}), \quad (2.125)$$

where Ω_X denotes the range of $\hat{\xi}$ over which the integral is to be evaluated. For example, if $f(\vec{r}, \hat{\xi}) = |\hat{\xi} \bullet \hat{e}_z|$, and Ω_X denotes all directions for which $\hat{\xi} \bullet \hat{e}_z < 0$ (downward), then $X(\vec{r}) = E_d(\vec{r})$. Although the other irradiances E_u , E_0 , and the vector irradiance $E_d - E_u$ for which the $f - \Omega_X$ combinations are, $|\hat{\xi} \bullet \hat{e}_z|$ with $\hat{\xi} \bullet \hat{e}_z < 0$, unity with all $\hat{\xi}$, and $\hat{\xi} \bullet \hat{e}_z$ with all $\hat{\xi}$, respectively, are used extensively in the text, another used in our treatment of Raman scattering is

$$\langle u^2 \rangle_E = \frac{\int_{\text{All } \hat{\xi}} (\hat{\xi} \bullet \hat{e}_z)^2 L(\vec{r}, \hat{\xi}) d\Omega(\hat{\xi})}{E_0(\vec{r})}.$$

Still another integral in which we will have interest in later is

$$\int_{\hat{\xi}' \bullet \hat{e}_z < 0} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}) L(\vec{r}, \hat{\xi}') d\Omega(\hat{\xi}'),$$

which provides the upward scattered radiance in the $\hat{\xi}$ direction at \vec{r} resulting from upward propagating radiance.

To show how to estimate integrals of the Eq. (2.125) type via Monte Carlo, for simplicity we restrict ourselves to the case of a vertically stratified medium for which the radiance is

⁵⁰A variant of the estimator is to pick the photon's fate at each collision. At each interaction the probability of the photon being scattered is ω_0 and the probability of absorption is $1 - \omega_0$. Thus, one could choose a random number ρ from a uniform distribution, $U[0,1]$, and if $\rho > \omega_0$ the photon scatters, otherwise it is absorbed and the chain terminates. Since the action $\omega_0(\vec{r}_i)$ is being chosen from its probability density, it is left out of the estimator. In this case all of the $\prod_{i=1}^N \omega_0(\vec{r})$ factors are absent from the estimator, but the chain terminates when the photon is absorbed.

given by Eq. (2.123). Then the desired integral becomes,

$$\begin{aligned}
& \frac{1}{F_0 |\hat{e}_z \bullet \hat{\xi}_1|} \int_{\Omega_x} f(\vec{r}_D, \hat{\xi}_D) L_D(\vec{r}_D, \hat{\xi}_D) d\Omega(\hat{\xi}_D) \\
&= \int \cdots \int_{\Omega_X} f(\vec{r}_D, \hat{\xi}_D) d\Omega(\hat{\xi}_D) \\
&\times \exp[-\tau(0, 1)] \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} d\Omega(\hat{\xi}_2) d|\vec{r}_1 - \vec{r}_0| \\
&\times \exp[-\tau(1, 2)] \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} d\Omega(\hat{\xi}_3) d|\vec{r}_2 - \vec{r}_1| \\
&\vdots \\
&\times \exp[-\tau(i-1, i)] \omega_0(\vec{r}_i) c(\vec{r}_i) \frac{P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi} d\Omega(\hat{\xi}_{i+1}) d|\vec{r}_i - \vec{r}_{i-1}| \\
&\vdots \\
&\times \exp[-\tau(N-2, N-1)] \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} d\Omega(\hat{\xi}_N) d|\vec{r}_{N-2} - \vec{r}_{N-1}| \\
&\times \exp[-\tau(N-1, N)] \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} d|\vec{r}_N - \vec{r}_{N-1}| \exp[-\tau(N, D)],
\end{aligned}$$

If we sample $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, \dots$ and $\hat{\xi}_2, \hat{\xi}_3, \dots, \hat{\xi}_N, \dots$ in the usual manner, and in addition at each collision *sample $\hat{\xi}_D$ from the probability $P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D) d\Omega(\hat{\xi}_D)/4\pi$* , then the estimate for the right hand side of the above after N collisions is just

$$Est_N = \frac{f(\vec{r}_D, \hat{\xi}_D)}{|\hat{e}_z \bullet \hat{\xi}_D|} \left(\prod_{i=1}^N \omega_0(\vec{r}_i) \right) \exp[-\tau(n, D)], \quad (2.126)$$

if $\hat{\xi}_D$ is within Ω_X , and zero otherwise. Summing Est_N over all the collisions in the chain provides the estimate of the desired integral for this single trial. As before, one can carry this out sequentially for a single trial chain of events, $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, \dots$ and $\hat{\xi}_2, \hat{\xi}_3, \dots, \hat{\xi}_N, \dots$, by updating the irradiance estimator for a given detector sequentially, i.e., after \vec{r}_1 , after \vec{r}_2 , etc., and then averaging over \mathcal{N} trials.

2.12.5 Backward Monte Carlo Evaluation of the Integral

We will now recast Eq. (2.117) in a different manner. Rather than centering the origin of $dV(\vec{r}_i)$ on \vec{r}_{i-1} , we now center it on \vec{r}_{i+1} , so

$$dV(\vec{r}_i) = |\vec{r}_{i+1} - \vec{r}_i|^2 d|\vec{r}_{i+1} - \vec{r}_i| d\Omega(\hat{\xi}_{i+1}).$$

Inserting these into Eq. (2.117), we see that all of the singular terms $|\vec{r}_{i+1} - \vec{r}_i|^{-2}$ cancel out, and we are left with

$$\begin{aligned}
L_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D &= \sum_{N=1}^{\infty} \int \cdots \int F_0(\vec{r}_0) \exp[-\tau(0, 1)] \\
&\quad \times \exp[-\tau(1, 2)] \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{P(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} d\Omega(\hat{\xi}_2), d|\vec{r}_2 - \vec{r}_1| \\
&\quad \times \exp[-\tau(2, 3)] \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{P(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} d\Omega(\hat{\xi}_3) d|\vec{r}_3 - \vec{r}_2| \\
&\quad \vdots \\
&\quad \times \exp[-\tau(N-1, N)] \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} d\Omega(\hat{\xi}_N) d|\vec{r}_N - \vec{r}_{N-1}| \\
&\quad \times \exp[-\tau(N, D)] \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{P(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} d|\vec{r}_D - \vec{r}_N| \Delta\Omega_D.
\end{aligned}$$

Now, canceling the $\Delta\Omega(\hat{\xi}_D)$'s, recognizing that $P(\hat{\xi} \rightarrow \hat{\xi}') = P(-\hat{\xi}' \rightarrow -\hat{\xi})$ and $d\Omega(\hat{\xi}) = d\Omega(-\hat{\xi})$, and rearranging yields

$$\begin{aligned}
L_D(\vec{r}_D, \hat{\xi}_D) &= \sum_{N=1}^{\infty} \int \cdots \int F_0(\vec{r}_0) \exp[-\tau(0, 1)] \\
&\quad \times \exp[-\tau(1, 2)] \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{P(\vec{r}_1; -\hat{\xi}_2 \rightarrow -\hat{\xi}_1)}{4\pi} d\Omega(-\hat{\xi}_2), d|\vec{r}_2 - \vec{r}_1| \\
&\quad \times \exp[-\tau(2, 3)] \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{P(\vec{r}_2; -\hat{\xi}_3 \rightarrow -\hat{\xi}_2)}{4\pi} d\Omega(-\hat{\xi}_3) d|\vec{r}_3 - \vec{r}_2| \\
&\quad \vdots \\
&\quad \times \exp[-\tau(N-1, N)] \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{P(\vec{r}_{N-1}; -\hat{\xi}_N \rightarrow -\hat{\xi}_{N-1})}{4\pi} d\Omega(-\hat{\xi}_N) d|\vec{r}_N - \vec{r}_{N-1}| \\
&\quad \times \exp[-\tau(N, D)] \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{P(\vec{r}_N; -\hat{\xi}_D \rightarrow -\hat{\xi}_N)}{4\pi} d|\vec{r}_D - \vec{r}_N|.
\end{aligned}$$

These integrals are easily seen to represent radiant power exiting the detector at \vec{r}_D and propagating to the point \vec{r}_0 on the surface, i.e., propagating in the *backward* direction. Evaluating the integrals by Monte Carlo, as above, the estimator for a single trial (e.g., a single choice of $\vec{r}_1, \dots, \vec{r}_N$ and $\hat{\xi}_2, \dots, \hat{\xi}_N$ becomes

$$L_D(\vec{r}_D, \hat{\xi}_D) = \sum_{N=1}^{\infty} F_0(\vec{r}_0) \exp[-\tau(0, 1)] \prod_{i=1}^N \omega_0(\vec{r}_i).$$

Note, \vec{r}_0 is determined from \vec{r}_1 and $\hat{\xi}_1$, and $\tau(0, 1)$ from $\vec{r}_0, \vec{r}_1, \hat{\xi}_1$ and $c(\vec{r})$. One can achieve this estimator in the following manner: (1) eject a photon from \vec{r}_D in the direction $-\hat{\xi}_D$.

Determine the interaction point \vec{r}_N and the phase function for scattering toward $-\hat{\xi}_1$. The estimator is

$$Est_1 = F_0(\vec{r}_0)\omega_0(\vec{r}_N)\exp[-\tau(0, N)]\omega_0(\vec{r}_N)P(\vec{r}_1; -\hat{\xi}_D \rightarrow -\hat{\xi}_1).$$

let the photon proceed to the next collision, i.e., choose the new direction $-\hat{\xi}_N$ from the density $P(\vec{r}_N; -\hat{\xi}_N \rightarrow -\hat{\xi}_{N-1})$, etc., providing \vec{r}_{N-1} and $\hat{\xi}_{N-1}$. The estimator for $N = 2$ is then

$$Est_2 = F_0(\vec{r}_0)\exp[-\tau(0, N-1)]\omega_0(\vec{r}_N)\omega_0(\vec{r}_{N-1})P(\vec{r}_{N-1}; -\hat{\xi}_{N-1} \rightarrow -\hat{\xi}_1).$$

Continue this procedure and the estimate for N collisions (one trial set) is just

$$L_D(\vec{r}_D, \hat{\xi}_D) = \sum_{i=1}^N Est_i.$$

The upper limit to N should be ∞ , but the evaluation need not progress beyond the point $\prod_{i=1}^N \omega_0(\vec{r}_i)$ becomes so small to make a significant contribution. As all photons exit the detector in the $-\hat{\xi}_D$ direction, the result is an estimate of $L(\vec{r}_D, \hat{\xi}_D)$ in a vanishingly small range of solid angles, i.e., exactly at $\hat{\xi}_D$ rather than in a non-zero $\Delta\Omega(\hat{\xi}_D)$ for the so-called “forward Monte Carlo” described earlier. Note that unlike the forward Monte Carlo the backward Monte Carlo procedure is no more complex (or inaccurate) for a 3-d medium compared to a 1-d medium. In fact, backward Monte Carlo has been used to assess the shading of radiometers perturbed by 3-dimensional structures in the water, e.g., ships, towers, the instrument itself, etc. There is, however, a tradeoff: in the backward Monte Carlo, the radiance *at only one detector* can be evaluated with the series of trials described above. A difficulty in the backward case arises when sources like lasers, where a narrow beam is incident on the surface, are employed. In this case, the last collision in each step in the chain must be in the field of view of the *source* to contribute by virtue of the $F_0(\vec{r}_0)$ term in the estimator.

2.12.6 Monte Carlo Solution Including Polarization

It is straightforward to derive integrals similar to those in Eq. (2.117), when the full polarization state of the radiance is taken into account. One simply replaces $P(\vec{r}; \hat{\xi} \rightarrow \hat{\xi}')$ by the Mueller scattering phase matrix $\mathbf{Z}(\vec{r}; \hat{\xi} \rightarrow \hat{\xi}')$, noting that since Mueller matrices do not commute, the order of the interactions must be preserved.⁵¹ In addition, F_0 is replaced

⁵¹Recall that $\mathbf{Z}(\vec{r}; \hat{\xi} \rightarrow \hat{\xi}') = \mathbf{R}(-i')\mathbf{P}(\Theta)\mathbf{R}(-i)$, where \mathbf{P} is the scattering phase matrix, $\cos \Theta = \hat{\xi} \bullet \hat{\xi}'$ and the \mathbf{R} 's are rotation matrices (see Section 2.10.3).

by \mathbf{F}_0 , the Stokes vector for the incident irradiance. The final result is

$$\begin{aligned}
\mathbf{I}_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D &= \sum_{N=1}^{\infty} \int \cdots \int \\
&\times \frac{\exp[-\tau(N, D)]}{|\vec{r}_D - \vec{r}_N|^2} \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{\mathbf{Z}(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} dV(\vec{r}_N), \\
&\times \frac{\exp[-\tau(N-1, N)]}{|\vec{r}_N - \vec{r}_{N-1}|^2} \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{\mathbf{Z}(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} dV(\vec{r}_{N-1}) \\
&\vdots \\
&\times \frac{\exp[-\tau(2, 3)]}{|\vec{r}_3 - \vec{r}_2|^2} \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{\mathbf{Z}(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} dV(\vec{r}_2) \\
&\times \frac{\exp[-\tau(1, 2)]}{|\vec{r}_2 - \vec{r}_1|^2} \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{\mathbf{Z}(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} dV(\vec{r}_1) \\
&\times \exp[-\tau(0, 1)] \mathbf{F}_0(\vec{r}_0)
\end{aligned}$$

where the first term (last scattering) contributes only if \vec{r}_N is within the field of view of the detector. This is the counterpart to Eq. (2.117) including polarization. Making replacements as those in deriving Eq. (2.119), we have

$$\begin{aligned}
\mathbf{I}_D(\vec{r}_D, \hat{\xi}_D) \Delta\Omega_D &= \sum_{N=1}^{\infty} \int \cdots \int \\
&\times \frac{\exp[-\tau(N-1, N)]}{|\vec{r}_D - \vec{r}_N|^2} \omega_0(\vec{r}_N) c(\vec{r}_N) \frac{\mathbf{Z}(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} d|\vec{r}_N - \vec{r}_{N-1}| \exp[-\tau(N, D)], \\
&\times \exp[-\tau(N-2, N-1)] \omega_0(\vec{r}_{N-1}) c(\vec{r}_{N-1}) \frac{\mathbf{Z}(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{4\pi} d\Omega(\hat{\xi}_N) d|\vec{r}_{N-1} - \vec{r}_{N-2}| \\
&\vdots \\
&\times \exp[-\tau(i-1, i)] \omega_0(\vec{r}_i) c(\vec{r}_i) \frac{\mathbf{Z}(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi} d\Omega(\hat{\xi}_{i+1}) d|\vec{r}_i - \vec{r}_{i-1}| \\
&\vdots \\
&\times \exp[-\tau(1, 2)] \omega_0(\vec{r}_2) c(\vec{r}_2) \frac{\mathbf{Z}(\vec{r}_2; \hat{\xi}_2 \rightarrow \hat{\xi}_3)}{4\pi} d\Omega(\hat{\xi}_3) d|\vec{r}_2 - \vec{r}_1| \\
&\times \exp[-\tau(0, 1)] \omega_0(\vec{r}_1) c(\vec{r}_1) \frac{\mathbf{Z}(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{4\pi} d\Omega(\hat{\xi}_2) d|\vec{r}_1 - \vec{r}_0| \\
&\times \int \mathbf{F}_0(\vec{r}_0) |\hat{\xi}_1 \bullet \hat{n}_S(\vec{r}_0)| dA_S(\vec{r}_0).
\end{aligned}$$

The Monte Carlo Estimate of this multiple integral is developed in the same manner as that for Eq. (2.119), with one exception. For the

$$\int \int \int \exp[-\tau(i-1, i)] \omega_0(\vec{r}_i) c(\vec{r}_i) \frac{\mathbf{Z}(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{4\pi} d\Omega(\hat{\xi}_{i+1}) d|\vec{r}_i - \vec{r}_{i-1}|$$

term we cannot sample $\hat{\xi}_{i+1}$ from $\mathbf{Z}(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})$ because we cannot find the rotation matrices without knowing $\hat{\xi}_{i+1}$. Thus, we have to sample from an alternate distribution and then remove the bias that is introduced by the improper sampling. The obvious distribution to sample from is the *scalar* phase function $P(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})/4\pi = P_{11}(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})/4\pi$. With \vec{r}_i chosen in the usual manner, the estimator for this particular integral is then

$$\omega_0(\vec{r}_i) \frac{\mathbf{Z}(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}{P_{11}(\vec{r}_i; \hat{\xi}_i \rightarrow \hat{\xi}_{i+1})}.$$

With similar changes to all of the integrals, the final estimator for one trial of N scattering events (e.g., a single choice of $\vec{r}_1, \dots, \vec{r}_N$ and $\hat{\xi}_2, \dots, \hat{\xi}_N$) in the forward Monte Carlo approach becomes

$$\begin{aligned} \exp[-\tau(N, D)] & \frac{|\hat{\xi}_1 \bullet \hat{n}_S(\vec{r}_0)|}{|\vec{r}_D - \vec{r}_N|^2} \frac{\omega_0(\vec{r}_N) \mathbf{Z}(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D)}{4\pi} \\ & \times \left[\frac{\omega_0(\vec{r}_{N-1}) \mathbf{Z}(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{P_{11}(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)} \dots \frac{\omega_0(\vec{r}_1) \mathbf{Z}(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{P_{11}(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)} \right] \mathbf{F}_0(\vec{r}_0). \end{aligned} \quad (2.127)$$

In the case of a system possessing horizontal invariance, this estimator reduces to

$$\begin{aligned} \mathbf{I}_D(\vec{r}_D, \hat{\xi}_D) & = \frac{|\hat{e}_z \bullet \hat{\xi}_1|}{|\hat{e}_z \bullet \hat{\xi}_D|} \sum_{N=1}^{\infty} \left(\prod_{i=1}^N \omega_0(\vec{r}_i) \right) \mathbf{Z}(\vec{r}_N; \hat{\xi}_N \rightarrow \hat{\xi}_D) \exp[-\tau(N, D)] \\ & \times \frac{\mathbf{Z}(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)}{P_{11}(\vec{r}_{N-1}; \hat{\xi}_{N-1} \rightarrow \hat{\xi}_N)} \dots \frac{\mathbf{Z}(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)}{P_{11}(\vec{r}_1; \hat{\xi}_1 \rightarrow \hat{\xi}_2)} \mathbf{F}_0. \end{aligned} \quad (2.128)$$

Equations (2.127) and (2.128) are the counterparts to Eqs. (2.120) and (2.124), respectively, when polarization of the radiance is considered. Note that the number of matrix multiplications required to evaluate Eq. (2.128) is $N + 1$ for each photon trajectory, and recalling that when \mathbf{Z} is replaced by the Mueller phase matrix and 2 rotation matrices, the total number of matrix multiplications becomes $3N + 1$. Further development of the Monte Carlo method when polarization is included follows that in the scalar case in a straightforward manner.

2.12.7 Connection to the Monte Carlo as Described in Section 2.6.2

It is not apparent that our derivation of the Monte Carlo estimates developed here directly from radiative transfer theory produces the same radiances, irradiances, etc., as that developed heuristically in Section 2.6.2. Actually, it will yield identical results, and likely with less variance for a given \mathcal{N} . We show this now using the example of estimating irradiance.

As mentioned earlier, to estimate the downward irradiance $E_d(\vec{r})$ we simply let $f(\vec{r}_D, \hat{\xi}_D) = |\hat{e}_z \cdot \hat{\xi}_D|$ in Eq. (2.126), so the estimate for E_d from the N^{th} collision is just

$$E_{d_N}(\vec{r}_D) = F_0 \left(\prod_{i=1}^N \omega_0(\vec{r}_i) \right) \exp[-\tau(n, D)]. \quad (2.129)$$

This at first sight seems to differ from Section 2.6.2, where it is stated that to estimate the contribution to the downward irradiance from each photon, one collects the total weight $\left(\prod_{i=1}^N \omega_0(\vec{r}_i) \right)$ of *the photon only if it propagates across the plane of a given optical depth*, i.e., there is no contribution unless the photon actually passes the depth in question. Actually, both are correct, the method described earlier is an equivalent, but cruder, estimate than that given in Eq. (2.129). To understand this, consider the simplest of examples: that of the Monte Carlo estimation of the irradiance of a beam in a 1-d medium with no scattering. We know that the probability of an absorption event for a photon in $d\ell$, at a distance ℓ after entering the medium, is given by

$$d\mathcal{P}_\ell(d\ell) = a d\ell \exp(-a\ell)$$

and, given a set of random numbers ρ , uniformly distributed on $[0,1]$, the path length to an absorption event (the i^{th}) is

$$\ell_i = -\frac{1}{a} \ln(1 - \rho_i).$$

Thus, to compute the irradiance due to the beam propagating a distance D into the medium one could inject \mathcal{N} photons and compute the number n that have $\ell_i > D$. Then the irradiance at D is just n/\mathcal{N} . Since we are sampling from the correct distribution of ℓ , $n/\mathcal{N} \rightarrow \exp(-aD)$ as $\mathcal{N} \rightarrow \infty$. Alternatively, we could compute the irradiance another way. When the photon first interacts with the medium, calculate the probability that it will travel a distance D without being absorbed (i.e., $\exp(-aD)$). Now as the estimator, use the probability of survival, i.e., the probability that the photon will reach D . This is $\exp(-aD)$ and a perfect estimation is obtained with a *single* photon! Execute more trials and all photons contribute the same estimate. This example is particularly simple, but it does show how one can use a simple estimator (i.e., one or zero for yes or no in a single trial, and then average over trials) or one can use the probability that an event will actually

happen in each trial (here, $\exp(-aD)$). Note that the variances of the two estimates are vastly different. The standard deviation (square root of the variance) of the irradiance in the first is $\propto 1/\sqrt{N}$, and in the second it is zero!

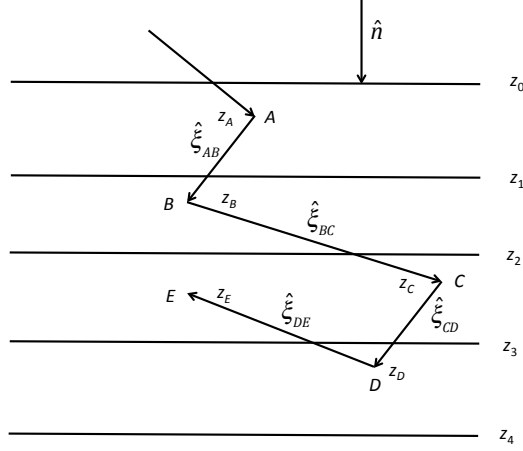


Figure 2.23: Example of a photon path (a single chain of events) in a homogeneous absorbing-scattering medium. The path continues until for N interactions, after which ω_0^{N+1} becomes too small to contribute significantly to any detector.

Let us apply these two methods for computing the irradiance in a homogeneous medium illuminated by a beam of irradiance F_0 at the surface. Choose a photon path, i.e., sample the points z_A, z_B, \dots, z_E and $\hat{\xi}_{AB}, \hat{\xi}_{BC}, \dots, \hat{\xi}_{DE}$ in Figure 2.23 from their appropriate distributions. The goal is to estimate the irradiances at z_1, z_2, \dots . If we use the method in Eq. (2.129), the contribution to the irradiance E_z at z_1, z_2, \dots , from the interaction at A would be

$$\begin{aligned} E_d(z_1) &= F_0 \omega_0 \exp[-c(z_1 - z_A)/|\hat{\xi}_{AB} \bullet \hat{n}|], \\ E_d(z_2) &= F_0 \omega_0 \exp[-c(z_2 - z_A)/|\hat{\xi}_{AB} \bullet \hat{n}|], \\ &\vdots \end{aligned}$$

with each scattering event contributing to several detectors. Similarly, the interaction at

B contributes

$$\begin{aligned} E_d(z_2) &= F_0 \omega_0^2 \exp[-c(z_2 - z_B)/|\hat{\xi}_{BC} \bullet \hat{n}|] \\ E_d(z_3) &= F_0 \omega_0^2 \exp[-c(z_3 - z_B)/|\hat{\xi}_{BC} \bullet \hat{n}|] \\ &\vdots \end{aligned}$$

etc. Continuing this for N events until ω_0^{N+1} is too small to contribute significantly (it was assumed that $N = 4$ in the figure), we have the total contribution to the irradiances from a single chain. Do this \mathcal{N} times, add all the contributions, and divide the result by \mathcal{N} , and we get the Monte Carlo estimate for the irradiances.

Alternatively, applying the procedure in Section 2.6.2 to the chain of events in Figure 2.23, the contributions to irradiances would be

$$\begin{aligned} E_d(z_1) &= F_0 \omega_0, \\ E_d(z_2) &= F_0 \omega_0^2, \\ E_d(z_3) &= F_0 \omega_0^3, \\ E_d(z_4) &= 0, \\ E_u(z_3) &= F_0 \omega_0^4. \end{aligned}$$

As in the earlier example with no scattering, both methods provide the correct irradiances; however, as in the $b = 0$ case, the variance in the first procedure will be lower than that in the second.

2.13 Appendix 3: Thermal Radiation

In Section 2.2 we described various methods of detecting electromagnetic radiation, but we have yet to consider sources of electromagnetic radiation other than the oscillating dipole. Now we examine one of the most important sources of electromagnetic radiation — the black body. This source provides a standard for radiometric calibration of radiometers, as it links radiance directly to the fundamental constants of physics. All objects at a temperature above absolute zero emit radiation, *thermal radiation*, having some of the characteristics of a black body. For example, the radiation incident on the atmosphere from the Sun is close in its spectral radiance to that of a black body at a temperature of about 6000 K.

First we describe in detail the black body. Then we develop the radiometric characteristics of thermal radiation and its radiative transfer. Finally, in the context of radiative

transfer, we briefly consider the problem of estimating the temperature of the Earth (actually, the oceans) from measurements of the radiance exiting the top of the atmosphere in the thermal infrared region of the spectrum ($\lambda > 3\text{-}5\ \mu\text{m}$).

2.13.1 Black Body Radiation

Everyday experience tells us that heated objects emit radiation, and if heated to a high enough temperature the radiation is in the visible part of the spectrum. This radiation is usually called *thermal* radiation. Consider the following experiment. A cylindrical tube of copper is connected between two electrodes. When a sufficiently large current is passed through the copper, it will emit red light. Now, the experiment is repeated, but with a small hole drilled through one side of the tube. In this case the cylinder surface emits red light as before, but the light exiting the small hole is both *brighter* and *whiter*, i.e., the spectrum emitted from the hole is shifted to the blue. If the current in the cylinder is increased, the temperature of the walls will increase, and the emitted light from the hole becomes bluer and brighter still. If the copper cylinder is replaced by a graphite cylinder (or a different metal) and the experiment repeated, it will be found that the radiation escaping the hole is independent of the nature of the material, depending only on its absolute temperature. In addition, if cylinders of different shape, e.g., rectangular as opposed to circular cylinders, or even cavities with very irregular shapes, are used in the experiment, the result is that the radiation exiting a small hole is independent of the *shape* as well as the nature of the material. If portions of the cavity walls are painted different colors, the result is still the same. The radiation emerging from a hole in the cavity depends *only* on the temperature of the walls of the cavity. This implies that the radiation within the cavity depends *only* on the temperature of the walls of the cavity. Finally, if the radiance exiting the hole is measured as a function of the angle from the normal to its surface, it is found that the radiance is uniform, i.e., it is the same in all directions. Summarizing these observations: the radiance exiting a hole cut in the side of a cavity that is heated to a temperature T is uniform and has a distinct spectrum (and brightness) that is independent of the shape of the cavity and the composition of the walls — it depends only on T and varies in a systematic manner with T .

Detailed experimental measurements show that when the walls of a cavity are heated to

a temperature T the radiance $B_\lambda(T, \lambda)$ exiting a small hole in the side is⁵²

$$B_\lambda(T, \lambda) = \frac{c_1}{\lambda^5 [\exp(c_2/\lambda T) - 1]}, \quad (2.130)$$

where

$$c_1 = 1.1910 \times 10^4 \text{ Watt } \mu\text{m}^4/\text{cm}^2 \text{ Sr}, \quad c_2 = 1.43879 \times 10^4 \text{ } \mu\text{m K}.$$

Using these numerical values for c_1 and c_2 requires that λ be in μm and T in Kelvins. $B_\lambda(T, \lambda)$ is called the Planck function and the symbol B (rather than L) is used for the radiance emitted from the hole to remind us of the special nature of thermal radiation. In fact, the nature of the radiation is so special it is given a name: *black body radiation*. Why “black body?” Because for the experiment we have described, any radiation entering the hole from the outside would likely be lost in the cavity, i.e., the cavity would absorb all radiation incident upon it — it would be totally black. The hole in the cavity wall is called a black body. The term is often used to refer to the cavity as well.

The Planck function is provided in Figure 2.24 for several values of T . The position of the maximum in the Planck function is given by $dB_\lambda/d\lambda = 0$, and is easily found from the formula to be given by (Wien’s displacement law)⁵³

$$\lambda_{\text{Max}} T = 2.8978 \times 10^3 \text{ } \mu\text{m K},$$

and at the maximum,

$$B_\lambda(T, \lambda_{\text{Max}}) = T^5 \times 4.09532 \times 10^{-16} \text{ W/cm}^2 \mu\text{m K}^5 \text{ Sr}.$$

Since the radiance leaving the hole, i.e., the black body, is uniform, the emitted irradiance (the radiant emittance) leaving the hole is given by

$$E_{B_\lambda} = \pi B_\lambda(T, \lambda),$$

and the total irradiance emitted, i.e., summed over all wavelengths, is

$$E_B = \int_0^\infty E_{B_\lambda}(T, \lambda) d\lambda = \sigma T^4,$$

⁵²The subscript λ on $B_\lambda(T, \lambda)$ is placed as a reminder that in the experimental measurement of the radiance, the radiant power detected within $\Delta\lambda$ is divided by $\Delta\lambda$ to form the spectral radiant power. This is to distinguish it from $B_\nu(T, \nu)$, where the independent variable is frequency ν rather than wavelength λ . Here the radiant power within $\Delta\nu$ is divided by $\Delta\nu$ to form the spectral radiant power that is used to calculate the radiance. Note that $B_\lambda(T, \lambda) |d\lambda| = B_\nu(T, \nu) |d\nu|$.

⁵³If you try to derive Wien’s law from B_λ , you will obtain the transcendental equation $e^x(x-5)+5=0$, where $x = c_2/(\lambda_{\text{Max}}T)$. This must be solved numerically (iteration is the simplest method) to obtain $x \approx 4.965114$.

where

$$\sigma = 5.6697 \times 10^{-12} \text{ W/cm}^2 \text{ K}^4.$$

From the nature of the black body radiation exiting the hole in the experiments described above, we are led to the conclusion that the radiation *inside* the cavity, must have the same properties as the radiation that escapes. The radiance is the same in all directions and at all points within the cavity, and it is independent of the shape of the cavity and the nature of the walls. Heating the walls of the cavity excites its atoms, which then radiate energy. This radiation bounces around in the cavity and some will be reabsorbed by the walls. At thermal equilibrium the temperature of the walls is constant, so any absorbed radiation must be re-emitted. The rates at which energy enters and leaves the walls must be the same.

One of the seminal events leading to the development of quantum theory was a theoretical description of the nature of black body radiation. Briefly, since the nature of the walls is irrelevant, it was assumed that they are totally reflecting, and since the shape is irrelevant it was assumed to be cubical. Then Maxwell's equations were solved to find the electromagnetic waves (modes) that can exist in the cavity. Planck assumed that these fields are maintained by oscillators in the walls, and by postulating that the energy of the oscillators was *quantized*, and that radiation was emitted and absorbed by these oscillators only when they changed from one *quantum state* to another. With these assumptions he was able to derive the form of Eq. (2.130) and explain the values of c_1 , c_2 , and σ in terms of fundamental constants:

$$c_1 = 2hc^2, \quad c_2 = \frac{hc}{k_B}, \quad \text{and} \quad \sigma = \frac{2\pi^5 k_B^4}{15h^3 c^3},$$

where k_B is the *Boltzmann* constant ($1.381 \times 10^{-23} \text{ J}\cdot\text{K}^{-1}$). In fact, there is no need to consider the walls perfectly reflecting because, as discussed above, any radiation that the walls absorb must be reemitted or the temperature of the walls will change. The only requirement is that the energy flux (irradiance) toward a wall must equal the energy flux away from a wall. It is irrelevant how the flux away from a wall is partitioned between reflection and emission. (In fact, since $E = \pi B$ it is also true that the radiance propagating toward the wall equals that propagating away from the wall, i.e., $L(\vec{r}, \hat{\xi}, \lambda) = L(\vec{r}, -\hat{\xi}, \lambda)$.) Thus, we may assume for purposes of this discussion that the walls are totally reflecting, partially reflecting and partially absorbing, or totally absorbing. The walls were assumed by Planck to be totally reflecting for convenience and to simplify the enumeration of the possible propagation modes within the cavity. Modern derivations of these equations use photon statistics (photons are spin 1 particles and are described by *Bose-Einstein* statistics) to populate the various available modes, arriving at the same result as Planck.

Given that Eq. (2.130) describes the distribution of radiation within a cavity of any

shape and wall composition, one can determine the thermal emission from any surface using simple arguments based on thermodynamics. Consider two black bodies insulated from the outside world and exchanging thermal radiation as shown in Figure 2.25 (a). (Remember the hole drawn in the side of each chamber is the black body.) The cooler one will warm due to the excess radiation from the warmer one, which will cool. Eventually they will come to the same temperature T_e . Then, the radiation exiting the left black body (LBB) and entering the right black body (RBB) will be identical in every way to that leaving the RBB and entering the LBB. Now, place an object, at the same temperature (T_e), inside the RBB as shown in Figure 2.25 (b). If the radiation within, and therefore exiting RBB, changes in any way from that before the object was inserted, then one black body will cool and the other heat, establishing a temperature difference where one did not exist before (without the required external work), in violation of the second law of thermodynamics. Thus, once equilibrium at T_e is established, the inserted body (also at T_e) must be essentially *invisible* within the cavity, i.e., the radiation is exactly as it would be at a temperature T_e *without* the inserted body.

How does this actually happen? It happens because all objects at a temperature $T > 0$ emit electromagnetic radiation, not just black bodies, e.g., the walls of the cavity. Consider the processes at the surface of the inserted object. Figure 2.26 (a) shows a schematic of the radiation propagating in the horizontal direction at a given point in the *absence* of the object. There are equal amounts of radiance propagating to the left and to the right (and all other directions as well). In Figure 2.26 (b) the object has been inserted. The radiance propagating to the right is unchanged, but now that propagating to the left consists of two parts: “r” that has been reflected from the surface of the object, and “e” that has been emitted by the object. Together the reflected and emitted parts form radiance equivalent to that propagating to the left in Figure 2.26 (a). If the object is removed from the black body and placed in a vacuum, it will continue to emit the radiation shown in the figure as long as the temperature remains at T_e . (Of course, as it loses energy through radiation it will cool.) Thus, any object heated to a temperature greater than absolute zero, will emit radiation, and therefore be a potential source of electromagnetic radiation. A source of particular interest to us is the Sun. It emits radiation with a spectral shape (and magnitude) characteristic of a black body at a temperature of approximately 6000 K (Figure 2.24).

To quantify these ideas, assume that the surface depicted in Figure 2.26 has a reflectance $r(\hat{\xi}' \rightarrow \hat{\xi})$ so that the reflected radiance $L_r(\hat{\xi})$ is

$$L_r(\hat{\xi}, \lambda) = \int_{\text{All } \hat{\xi}' \bullet \hat{n} < 0} r(\hat{\xi}' \rightarrow \hat{\xi}) L_i(\hat{\xi}', \lambda) d\Omega(\hat{\xi}'),$$

where \hat{n} is the normal to the surface (on the same side as the incident radiation) and $L_i(\hat{\xi}', \lambda)$ is the incident radiance. Now, since the surface is inside the black body, $L_i(\hat{\xi}, \lambda) =$

$B_\lambda(T, \lambda)$, so the reflected radiance is

$$L_r(\hat{\xi}, \lambda) = B_\lambda(T, \lambda) \int_{\text{All } \hat{\xi}' \bullet \hat{n} < 0} r(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}').$$

In order that object be invisible we must have

$$L_e(\hat{\xi}, \lambda) + L_r(\hat{\xi}, \lambda) = B_\lambda(T, \lambda),$$

or

$$L_e(\hat{\xi}, \lambda) = B_\lambda(T, \lambda) - L_r(\hat{\xi}, \lambda) = \left[1 - \int_{\text{All } \hat{\xi}' \bullet \hat{n} < 0} r(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}') \right] B_\lambda(T, \lambda).$$

Defining the quantity in the square brackets to be $\epsilon(\hat{\xi}, \lambda)$, we have

$$L_e(\hat{\xi}, \lambda) = \epsilon(\hat{\xi}, \lambda) B_\lambda(T, \lambda),$$

where

$$\epsilon(\hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}' \bullet \hat{n} < 0} r(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}') = 1. \quad (2.131)$$

The quantity $\epsilon(\hat{\xi}, \lambda)$ is called the directional spectral emissivity of the surface. We note that the smallest value of ϵ is zero, i.e., $\epsilon(\hat{\xi}, \lambda) \geq 0$ because $L_e(\hat{\xi}, \lambda) \geq 0$. Therefore, the *maximum* value the integral is unity. Conversely, the minimum value of $r(\hat{\xi}' \rightarrow \hat{\xi})$ is zero, so the minimum value of the integral is zero, and the *maximum* value of $\epsilon(\hat{\xi}, \lambda)$ is unity. Thus, $0 \leq \epsilon(\hat{\xi}, \lambda) \leq 1$. Of course, if the object itself is a black body, then $\epsilon(\hat{\xi}, \lambda) = 1$. Equation (2.131) takes a particularly simple form for a flat Fresnel reflecting surface, for which $r(\hat{\xi}' \rightarrow \hat{\xi})$ is a Dirac delta function, and

$$\epsilon(\theta, \phi, \lambda) + r_f(\theta, \phi, \lambda) = 1, \quad (2.132)$$

where θ and ϕ provide the direction of the reflected-emitted radiance, $\theta = \cos^{-1}(|\hat{\xi}' \bullet \hat{n}|) = \cos^{-1}(|\hat{\xi} \bullet \hat{n}|)$, and r_f is the Fresnel reflectivity. This equation allows a simple computation of the emissivity of a specularly reflecting surface.

The emitted spectral *irradiance* from a surface is

$$\begin{aligned} E(\lambda) &= \int_{\text{All } \hat{\xi} \bullet \hat{n} < 0} |\hat{\xi}' \bullet \hat{n}| L_e(\hat{\xi}, \lambda) d\Omega(\hat{\xi}) = B_\lambda(\lambda, T) \int_{\text{All } \hat{\xi} \bullet \hat{n} < 0} |\hat{\xi}' \bullet \hat{n}| \epsilon(\hat{\xi}, \lambda) d\Omega(\hat{\xi}) \\ &\triangleq \epsilon(\lambda) \pi B_\lambda(\lambda, T), \end{aligned}$$

where

$$\epsilon(\lambda) \triangleq \frac{1}{\pi} \int_{\text{All } \hat{\xi} \bullet \hat{n} < 0} |\hat{\xi}' \bullet \hat{n}| \epsilon(\hat{\xi}, \lambda) d\Omega(\hat{\xi}),$$

and the total irradiance emitted from a surface is

$$E = \int_0^\infty E(\lambda) d\lambda = \int_0^\infty \epsilon(\lambda) \pi B_\lambda(\lambda, T) d\lambda \triangleq \epsilon \sigma T^4.$$

Noting that

$$\sigma T^4 = \int_0^\infty \pi B_\lambda(\lambda, T) d\lambda,$$

we have

$$\epsilon = \frac{\int_0^\infty \epsilon(\lambda) B_\lambda(\lambda, T) d\lambda}{\int_0^\infty B_\lambda(\lambda, T) d\lambda}.$$

The quantity ϵ is called the emissivity of the surface, while $\epsilon(\lambda)$ is called the spectral emissivity.

2.13.2 Radiative Transfer of Thermal Radiation

From the above discussion, we know that matter at a temperature above absolute zero must emit radiation (thermal radiation), and that the emitted radiance can be written as the black body radiance modulated by the directional spectral emissivity $\epsilon(\hat{\xi}, \lambda)$. In the radiative transfer equation,

$$\frac{dL(\vec{r}, \hat{\xi})}{dl} = -c(\vec{r})L(\vec{r}, \hat{\xi}) + \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}) L(\vec{r}, \hat{\xi}') d\Omega(\hat{\xi}') + Q(\vec{r}, \hat{\xi}),$$

the thermal emission is part of $Q(\vec{r}, \hat{\xi})$. But, how do we actually relate Q to the emitted radiation? Consider adding a purely absorbing (i.e., non-scattering) layer to one wall of the cavity of a black body. For the radiance within the layer we have

$$\frac{dL(\vec{r}, \hat{\xi})}{dl} = -a(\vec{r})L(\vec{r}, \hat{\xi}) + Q(\vec{r}, \hat{\xi}).$$

Since we know that the radiation within (and exiting) the cavity is independent of the thickness of the layer, $dL(\vec{r}, \hat{\xi})/dl = 0$ within the layer (and everywhere in the cavity). In addition, within the layer, $L(\vec{r}, \hat{\xi}) = B_\lambda(\lambda, T)$. Therefore $Q(\vec{r}, \hat{\xi}) = a(\vec{r})B_\lambda(\lambda, T)$, and in general the radiative transfer equation becomes

$$\begin{aligned} \frac{dL(\vec{r}, \hat{\xi}, \lambda)}{dl} = & -c(\vec{r}, \lambda)L(\vec{r}, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(\vec{r}, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') \\ & + a(\vec{r})B_\lambda(\lambda, T) + Q_{\text{NT}}(\vec{r}, \hat{\xi}, \lambda), \end{aligned}$$

where $Q_{\text{NT}}(\vec{r}, \hat{\xi}, \lambda)$ is the contribution to Q from non-thermal emission, e.g., fluorescence.

In geophysical optics, thermal emission is only important for $\lambda \gtrsim 3 \mu\text{m}$ (Figure 2.24). It is the basis of the estimation of sea surface temperature measurements by Earth-orbiting infrared imagers. Let's see how this works. We will simplify the problem as much as possible by (1) assuming that scattering is negligible compared to absorption and (2) assuming that the atmosphere is homogeneous and isothermal with a temperature T_A . If the water surface has a temperature T_S , then the radiative transfer problem (in a plane parallel ocean-atmosphere system) that we need to solve for the spectral radiance $L(z, u, \lambda)$ is

$$u \frac{dL(z, u, \lambda)}{dz} = -aL(z, u, \lambda) + aB_\lambda(\lambda, T_A),$$

where as in our normal convention $z = 0$ is at the top of the atmosphere and increases with depth *into* the atmosphere. Since we will be interested in the *upward* radiance, it will be convenient for us to replace u by $-\mu$, so upward radiance has $\mu > 0$. The boundary conditions are that the downward radiance at $z = 0$ vanishes, and the upward radiance at $z = z_B$ (the bottom of the atmosphere) is $\epsilon(\mu, \lambda)B_\lambda(\lambda, T_S)$. A typical temperature for the ocean surface is ~ 300 K, so the measurement of $L(z_B, \mu, \lambda)$ will be easiest (in the absence of absorption) near $\lambda = 10 \mu\text{m}$ (Figure 2.24). Therefore, to estimate $\epsilon(\mu, \lambda)$ we use a nominal wavelength of $10 \mu\text{m}$, for which the refractive index of water is approximately $1.22 - 0.051i$.⁵⁴ This index along with the Fresnel equations (Eqs. (1.39) and (1.40)) shows that $r_f \approx 0.01$ at normal incidence $\mu = 1$, so, using Eq. 2.132, we can take $\epsilon(\mu, \lambda) \approx 0.99 \approx 1$ for μ close to unity,⁵⁵ i.e., at $z = z_B$ we take $L(z_B, \mu, \lambda) = B_\lambda(\lambda, T_S)$. This problem is easy to solve for the upward radiance,⁵⁶ and the result is

$$L(0, \mu, \lambda) = [B_\lambda(\lambda, T_S) - B_\lambda(\lambda, T_A)] \exp(-az_B/\mu) + B_\lambda(\lambda, T_A). \quad (2.133)$$

Thus, for a spectral region with negligible absorption ($az_B \ll 1$), the exponential is approximately unity and $L(0, \mu, \lambda) \approx B_\lambda(\lambda, T_S)$, enabling the determination of T_S using the Planck formula (Eq. (2.130)). In contrast, if $az_B \gg 1$, $L(0, \mu, \lambda) \approx B_\lambda(\lambda, T_A)$ and the water surface is totally obscured by the atmosphere. So, in absorption-free regions of the spectrum, one observes $B_\lambda(\lambda, T_S)$, while in regions of strong atmospheric absorption one measures $B_\lambda(\lambda, T_A)$. Typically the atmosphere is colder than the surface, so in the emission spectrum for the ocean atmosphere-system, the absorption bands in the atmosphere appear as a decrease in radiance. In contrast, over polar regions (ice covered) it is possible for $B_\lambda(\lambda, T_A) > B_\lambda(\lambda, T_S)$. In this case, the absorption bands in the atmosphere appear as an *increase* in radiance, i.e., they appear as emission bands on the background of radiance from the surface.

⁵⁴Recall the absorption coefficient is given by $a = 4\pi m_i/\lambda$, so at $\lambda = 10 \mu\text{m}$, $m_i = 0.051$ leads to an absorption coefficient of approximately $64,000 \text{ m}^{-1}$! This means using $B_\lambda(\lambda, T_S)$ to determine T_S provides the temperature of a layer of surface water of the order of $30 \mu\text{m}$ in thickness. This is called the “skin” temperature, and is *not* the same as the bulk temperature, e.g., that measured with an immersed thermometer (the difference can be a few tenths K).

⁵⁵Note that r_f is a slowly varying function of the incidence angle (Figure 1.5).

⁵⁶See Chapter 14, the Mathematical Appendix.

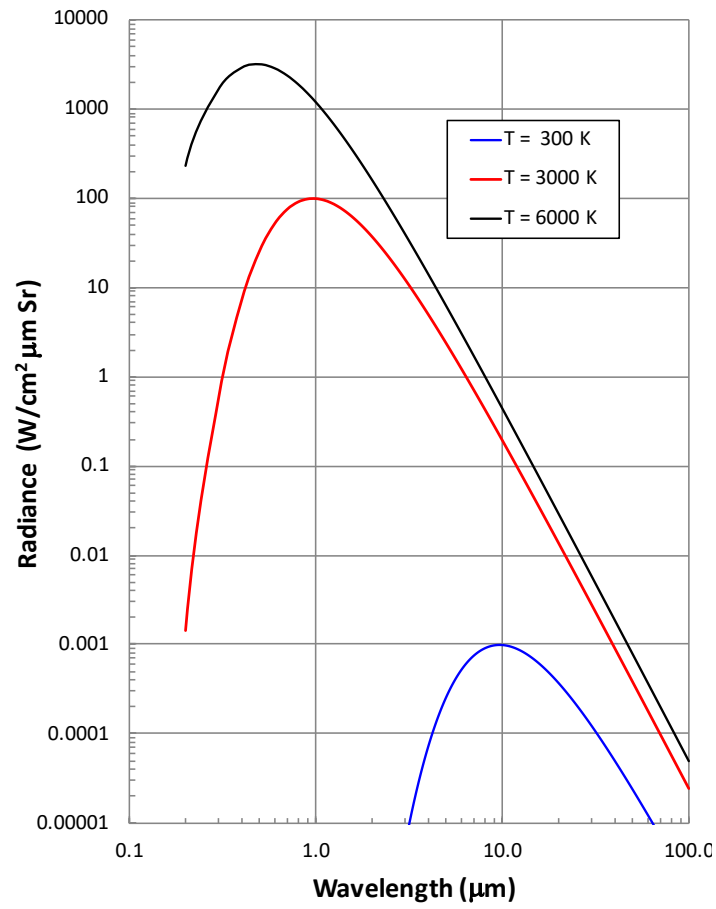


Figure 2.24: Planck function for a 300°K (lower), 3000°K (middle) and 6000°K (upper) black body. The Earth and Sun radiate to space at $T \approx 300$ and 6000 K, respectively..

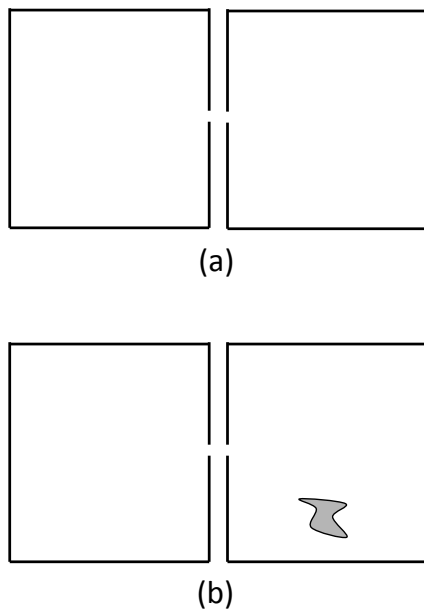


Figure 2.25: Two black bodies (isolated from the environment) exchange radiation until the both come to the same equilibrium temperature T_e (a). An object, also at T_e , is placed in the black body on the right (b).

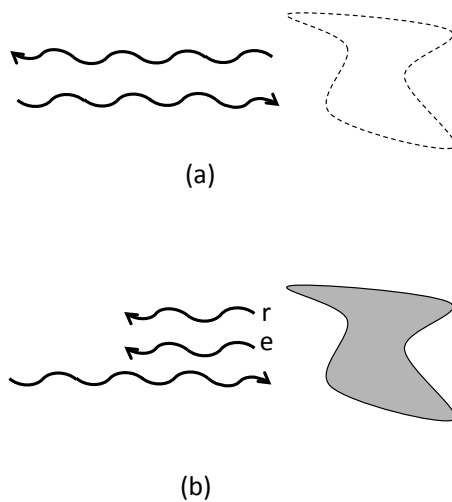


Figure 2.26: A schematic of radiation moving along the $\pm x$ axis in the absence (a) and the presence (b) of the object within the black body of Figure 2.25.

2.14 Bibliographic Notes

The subjects discussed in this chapter are developed in great detail in the books by Preisendorfer (*Radiative Transfer on Discrete Spaces* [1965] and *Hydrologic Optics* [1976] and in the book *Light and Water* by Mobley [1994]. The phenomenological derivation development of the radiative transfer equation is our own. Derivation of the RTE from electromagnetic theory, as well as criticism of the phenomenological derivation is presented in Mishchenko [2013], Mishchenko [2014], Mishchenko [2008], and Mishchenko et al. [2016]. A proof of the uniqueness theorem and the reciprocity principle is given by Case [1957] and in the form presented here by Yang and Gordon [1997].

Numerous methods of solving the RTE along with examples and comparisons of their accuracies are presented in *Radiative Transfer in Scattering and Absorbing Atmospheres: Standard Computational Procedures* by Lenoble [1986]. Excellent surveys are presented by Hansen and Travis [1974] and Irvine [1976]. The book *Multiple Light Scattering, Tables, Formulas, and Approximations* [van de Hulst, 1980] covers the successive order of scattering in great detail as well as the approach to the asymptotic regime in a semi-infinite medium. Monte Carlo methods are discussed by various authors, the most complete being *Monte Carlo Principles and Neutron Transport Problems* by Spanier and Gelbard [1969]. Mobley [1994] provides a good introduction to the application of such methods in natural waters. The discrete ordinate methods are presented in detail in *Radiative Transfer* by Chandrasekhar [1950] and modern computer techniques employing them are presented by Stamnes et al. [1988]. The matrix operator method has a long history and goes by various names, doubling, adding, etc. An account of the history and a very readable development of the method (followed here) is given by Plass et al. [1973]. Invariant imbedding forms the basis of Chandrasekhar's methods and is developed in detail in *Radiative Transfer* as well as in Preisendorfer [1965], Preisendorfer [1976a], Preisendorfer [1976b], Preisendorfer [1976c], Preisendorfer [1976d], Preisendorfer [1976e], Preisendorfer [1976f], and Mobley [1994]. It is the solution technique used in the well-known HydroLight code (www.hydrolight.info). The development of invariant imbedding we presented is our own. A technique that we have not described is the "singular eigenvalue" method [Case and Zweifel, 1967] [McCormick, 1996], which is appealing because it connects radiative transfer to the more traditional approaches in mathematical physics. Still other techniques can be found in Lenoble [1986].

Chapter 3

Scattering Theory

In the previous chapter, we found that the propagation of radiance in a medium is governed by the radiative transfer equation. The parameters in the radiative transfer equation that determine the solution are the inherent optical properties — the absorption coefficient and the volume scattering functions (elastic and inelastic). In the environment the scattering is due to the interaction of electromagnetic radiation with atoms, molecules, and liquid or solid particles with sizes that range from a fraction of the wavelength to many multiples of the wavelength of the electromagnetic radiation. In this chapter we review the theory of light scattering by gases, liquids, and homogeneous particles. One goal is to acquaint the reader with exact and approximate methods of determining the scattering and absorption properties of homogeneous particles. A secondary goal is to show how their optical properties relate to their physical properties.

We begin by reviewing the basic scattering problem by using scattering by individual atoms in a gas as a guide. Rayleigh scattering by gases and liquids is discussed in some detail and followed by the scattering and absorption of a homogeneous sphere (Mie theory). Next, we examine several approximations to scattering and absorption that are applicable to specific size ranges and particle compositions. Finally, we consider scattering by particles with irregular shapes. In the appendices we derive the *optical theorem*, which is used to determine the extinction properties of scatterers, and describe *Raman scattering*.

3.1 Introduction

Consider a plane wave propagating in the $\vec{\kappa}_0$ direction, with electric field given by

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \exp[i(\vec{\kappa}_0 \bullet \vec{r} - \omega t)], \quad (3.1)$$

where $\vec{E}_0 = \hat{e}_{0r}E_{0r} + \hat{e}_{0\ell}E_{0\ell}$ and $\hat{e}_{0r} \times \hat{e}_{0\ell} = \hat{\kappa}_0$. We found in Chapter 1, that when this wave is incident on an atom of polarizability α (located at the origin of coordinates, i.e., $r = 0$), the scattered fields at a distance r from the atom, perpendicular (E_r) and parallel (E_ℓ) to the plane of scattering, are related to the incident field amplitudes through¹

$$\begin{pmatrix} E_r \\ E_\ell \end{pmatrix} = \frac{\exp[i(\kappa r - \omega t)]}{-i\kappa r} \begin{pmatrix} A_1(\Theta) & 0 \\ 0 & A_2(\Theta) \end{pmatrix} \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix}, \quad (3.2)$$

where Θ is the scattering angle, $\kappa = 2\pi/\lambda$, λ is the wavelength *in the medium*,² and

$$A_1(\Theta) = \frac{-i\kappa^3\alpha}{4\pi\epsilon_0} \quad \text{and} \quad A_2(\Theta) = \frac{-i\kappa^3\alpha}{4\pi\epsilon_0} \cos \Theta.$$

Since α has dimensions of ϵ_0 times volume, A_1 and A_2 are dimensionless. The A 's are called scattering amplitudes, and completely describe the scattered field in terms of the incident field.

Recall that for any electromagnetic wave the time-averaged Poynting vector (Eq. (1.11)), expressed in complex notation is

$$\langle \vec{S} \rangle = \frac{\hat{\kappa}}{2\mu_0 c} \Re(\vec{E} \bullet \vec{E}^*).$$

In this form the time-average Poynting vector of the scattered wave (measured *without* polarizing filters) is

$$|\langle \vec{S} \rangle_{\text{Scatt}}| = \frac{1}{\kappa^2 r^2} \left(\frac{A_1 A_1^* E_{0r} E_{0r}^* + A_2 A_2^* E_{0\ell} E_{0\ell}^*}{2\mu_0 c} \right).$$

If the incident field is unpolarized, $E_{0r} E_{0r}^* = E_{0\ell} E_{0\ell}^* = E_0 E_0^*/2$, i.e., half of the irradiance is polarized perpendicular and half parallel to the scattering plane. Then,

$$|\langle \vec{S} \rangle_{\text{Scatt}}| = \frac{1}{\kappa^2 r^2} \left(\frac{A_1 A_1^* + A_2 A_2^*}{2} \right) |\langle \vec{S} \rangle_{\text{Inc}}|.$$

¹The scattering plane is the plane defined by the incident and scattered directions, $\hat{\kappa}_0$ and $\hat{\kappa}$, respectively. The scattered field is $\vec{E}_r = \hat{e}_r E_r + \hat{e}_\ell E_\ell$, where the unit vectors are perpendicular (r) and parallel (ℓ) to the scattering plane, $\hat{e}_r \times \hat{e}_\ell$ is in the direction of propagation of the scattered wave. The unit vectors \hat{e}_{0r} and $\hat{e}_{0\ell}$ are also perpendicular and parallel to the scattering plane as well with $\hat{e}_{0r} \times \hat{e}_{0\ell}$ in the direction of propagation of the incident field.

²Throughout this chapter $\vec{\kappa}$ is the propagation vector *in the medium* in which the scattering atom, molecule, or physical particle resides.

Following the procedure leading to Eq. (2.24), we have

$$\beta(\Theta) = \beta_{11}(\Theta) = \frac{n}{\kappa^2} \left[\frac{A_1(\Theta)A_1^*(\Theta) + A_2(\Theta)A_2^*(\Theta)}{2} \right], \quad (3.3)$$

where n is the number density of scatterers. Inserting the A 's for Rayleigh scattering reproduces the volume scattering function provided in Eq. (2.25).

By analogy to Rayleigh scattering, the scattered field from *any* object, e.g., a small particle, can be related to the incident field amplitudes through

$$\begin{pmatrix} E_r \\ E_\ell \end{pmatrix} = \frac{\exp[i(\kappa r - \omega t)]}{-i\kappa r} \begin{pmatrix} A_1(\Theta, \Phi) & A_4(\Theta, \Phi) \\ A_3(\Theta, \Phi) & A_2(\Theta, \Phi) \end{pmatrix} \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix}, \quad (3.4)$$

where now the scattering amplitude matrix in general depends on both scattering angles (see Figure 2.14).

Given Eq. (3.4) and the definition of the incident and scattered Stokes vectors \mathbf{I}_0 and \mathbf{I} , we can find the Mueller matrix that takes one into the other, i.e.,

$$\mathbf{I} = \mathbf{M}(\Theta, \Phi) \mathbf{I}_0.$$

This is effected using the method described in Section 1.10.8:

$$M_{ji} = \frac{(-1)^{i-j}}{2\kappa^2 r^2} \text{Tr} \left[\boldsymbol{\sigma}_j \mathbf{A} \boldsymbol{\sigma}_i \mathbf{A}^\dagger \right], \quad (3.5)$$

e.g.,

$$\begin{aligned} M_{11} &= \frac{1}{2\kappa^2 r^2} \text{Tr} \left[\boldsymbol{\sigma}_1 \mathbf{A} \boldsymbol{\sigma}_1 \mathbf{A}^\dagger \right] = \frac{1}{2\kappa^2 r^2} \text{Tr} \left[\mathbf{A} \mathbf{A}^\dagger \right] \\ &= \frac{1}{2\kappa^2 r^2} \text{Tr} \left[\begin{pmatrix} A_1 & A_4 \\ A_3 & A_2 \end{pmatrix} \begin{pmatrix} A_1^* & A_3^* \\ A_4^* & A_2^* \end{pmatrix} \right] \\ &= \frac{1}{2\kappa^2 r^2} [A_1 A_1^* + A_2 A_2^* + A_3 A_3^* + A_4 A_4^*] \\ &\triangleq \frac{1}{\kappa^2 r^2} M'_{11} \end{aligned}$$

In a similar manner we find

$$\mathbf{M}(\Theta, \Phi) = \frac{1}{\kappa^2 r^2} \begin{pmatrix} M'_{11}(\Theta, \Phi) & M'_{12}(\Theta, \Phi) & M'_{13}(\Theta, \Phi) & M'_{14}(\Theta, \Phi) \\ M'_{21}(\Theta, \Phi) & M'_{22}(\Theta, \Phi) & M'_{23}(\Theta, \Phi) & M'_{24}(\Theta, \Phi) \\ M'_{31}(\Theta, \Phi) & M'_{32}(\Theta, \Phi) & M'_{33}(\Theta, \Phi) & M'_{34}(\Theta, \Phi) \\ M'_{41}(\Theta, \Phi) & M'_{42}(\Theta, \Phi) & M'_{43}(\Theta, \Phi) & M'_{44}(\Theta, \Phi) \end{pmatrix},$$

where,

$$\begin{aligned}
M'_{11}(\Theta, \Phi) &= \frac{1}{2}(A_1(\Theta, \Phi)A_1^*(\Theta, \Phi) + A_2(\Theta, \Phi)A_2^*(\Theta, \Phi) + A_3(\Theta, \Phi)A_3^*(\Theta, \Phi) + A_4(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{12}(\Theta, \Phi) &= \frac{1}{2}(A_2(\Theta, \Phi)A_2^*(\Theta, \Phi) - A_1(\Theta, \Phi)A_1^*(\Theta, \Phi) + A_4(\Theta, \Phi)A_4^*(\Theta, \Phi) - A_3(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{13}(\Theta, \Phi) &= \Re(A_2(\Theta, \Phi)A_3^*(\Theta, \Phi) + A_1(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{14}(\Theta, \Phi) &= \Im(A_2(\Theta, \Phi)A_3^*(\Theta, \Phi) - A_1(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{21}(\Theta, \Phi) &= \frac{1}{2}(A_2(\Theta, \Phi)A_2^*(\Theta, \Phi) - A_1(\Theta, \Phi)A_1^*(\Theta, \Phi) - A_4(\Theta, \Phi)A_4^*(\Theta, \Phi) + A_3(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{22}(\Theta, \Phi) &= \frac{1}{2}(A_2(\Theta, \Phi)A_2^*(\Theta, \Phi) + A_1(\Theta, \Phi)A_1^*(\Theta, \Phi) - A_4(\Theta, \Phi)A_4^*(\Theta, \Phi) - A_3(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{23}(\Theta, \Phi) &= \Re(A_2(\Theta, \Phi)A_3^*(\Theta, \Phi) - A_1(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{24}(\Theta, \Phi) &= \Im(A_2(\Theta, \Phi)A_3^*(\Theta, \Phi) + A_1(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{31}(\Theta, \Phi) &= \Re(A_2(\Theta, \Phi)A_4^*(\Theta, \Phi) + A_1(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{32}(\Theta, \Phi) &= \Re(A_2(\Theta, \Phi)A_4^*(\Theta, \Phi) - A_1(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{33}(\Theta, \Phi) &= \Re(A_1(\Theta, \Phi)A_2^*(\Theta, \Phi) + A_3(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{34}(\Theta, \Phi) &= \Im(A_2(\Theta, \Phi)A_1^*(\Theta, \Phi) + A_4(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{41}(\Theta, \Phi) &= \Im(A_4(\Theta, \Phi)A_2^*(\Theta, \Phi) + A_1(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{42}(\Theta, \Phi) &= \Im(A_4(\Theta, \Phi)A_2^*(\Theta, \Phi) - A_1(\Theta, \Phi)A_3^*(\Theta, \Phi)) \\
M'_{43}(\Theta, \Phi) &= \Im(A_1(\Theta, \Phi)A_2^*(\Theta, \Phi) - A_3(\Theta, \Phi)A_4^*(\Theta, \Phi)) \\
M'_{44}(\Theta, \Phi) &= \Re(A_1(\Theta, \Phi)A_2^*(\Theta, \Phi) - A_3(\Theta, \Phi)A_4^*(\Theta, \Phi)).
\end{aligned}$$

For scattering from an object with spherical symmetry, $A_3 = A_4 = 0$, i.e.,

$$\begin{pmatrix} E_r \\ E_\ell \end{pmatrix} = \frac{\exp[(i\kappa r - \omega t)]}{-i\kappa r} \begin{pmatrix} A_1(\Theta) & 0 \\ 0 & A_2(\Theta) \end{pmatrix} \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix}, \quad (3.6)$$

and the Mueller matrix reduces in form to

$$\mathbf{M}(\Theta) = \frac{1}{\kappa^2 r^2} \begin{pmatrix} M'_{11}(\Theta) & M'_{12}(\Theta) & 0 & 0 \\ M'_{12}(\Theta) & M'_{11}(\Theta) & 0 & 0 \\ 0 & 0 & M'_{33}(\Theta) & -M'_{34}(\Theta) \\ 0 & 0 & M'_{34}(\Theta) & M'_{33}(\Theta) \end{pmatrix},$$

where,

$$\begin{aligned}
M'_{11}(\Theta) &= \frac{1}{2}(A_1(\Theta)A_1^*(\Theta) + A_2(\Theta)A_2^*(\Theta)) \\
M'_{12}(\Theta) &= \frac{1}{2}(A_1(\Theta)A_1^*(\Theta) - A_2(\Theta)A_2^*(\Theta)) \\
M'_{33}(\Theta) &= \Re(A_1(\Theta)A_2^*(\Theta)) \\
M'_{34}(\Theta) &= \Im(A_1(\Theta)A_2^*(\Theta)).
\end{aligned} \quad (3.7)$$

We provide formulas for $A_1(\Theta)$ and $A_2(\Theta)$ for a homogeneous sphere in the following section.

Again, following the procedure leading to Eq. (2.24),, i.e., $\beta = nr^2 \mathbf{M}$,

$$\beta(\Theta) = \frac{n}{\kappa^2} \begin{pmatrix} M'_{11}(\Theta, \Phi) & M'_{12}(\Theta, \Phi) & M'_{13}(\Theta, \Phi) & M'_{14}(\Theta, \Phi) \\ M'_{21}(\Theta, \Phi) & M'_{22}(\Theta, \Phi) & M'_{23}(\Theta, \Phi) & M'_{24}(\Theta, \Phi) \\ M'_{31}(\Theta, \Phi) & M'_{32}(\Theta, \Phi) & M'_{33}(\Theta, \Phi) & M'_{34}(\Theta, \Phi) \\ M'_{41}(\Theta, \Phi) & M'_{42}(\Theta, \Phi) & M'_{43}(\Theta, \Phi) & M'_{44}(\Theta, \Phi) \end{pmatrix}. \quad (3.8)$$

The scattering coefficient b depends on the polarization of the incident irradiance. For unpolarized irradiance,

$$b = \frac{n}{\kappa^2} \int M'_{11}(\Theta, \Phi) d\Omega, \quad (3.9)$$

where the integration is over the full range of solid angle.³ The scattering cross section $\sigma_b \triangleq b/n$ is

$$\sigma_b = \frac{1}{\kappa^2} \int M'_{11}(\Theta, \Phi) d\Omega. \quad (3.10)$$

Particles will in general absorb as well as scatter radiation. The absorption cross section σ_a , when added to the scattering cross section σ_b , yields the total cross section (the *attenuation* or *extinction* cross section), σ_c , i.e.,

$$\sigma_c = \sigma_a + \sigma_b,$$

³For incident irradiance described by the Stokes vector \mathbf{I}_E , the scattering coefficient is

$$b = \frac{n}{\kappa^2 I_E} \int \left[M'_{11}(\Theta, \Phi) I_E + M'_{12}(\Theta, \Phi) Q_E + M'_{13}(\Theta, \Phi) U_E + M'_{14}(\Theta, \Phi) V_E \right] d\Omega.$$

In this equation both \mathbf{M} and \mathbf{I}_E must be referenced to the same coordinate system, but typically the incident Stokes vector is referenced to the laboratory coordinate “Lab,” while \mathbf{M} is referenced to the scattering plane. Thus, we must transform \mathbf{I}_{Lab} from the laboratory reference to the scattering plane. This means replacing \mathbf{I}_E above by $\mathbf{I}_E = \mathbf{R}(\Phi) \mathbf{I}_{\text{Lab}}$. Then,

$$b = \frac{n}{\kappa^2 I} \int \left[M'_{11}(\Theta, \Phi) I + M'_{12}(\Theta, \Phi) (Q \cos(2\Phi) - U \sin(2\Phi)) + M'_{13}(\Theta, \Phi) (U \cos(2\Phi) + Q \sin(2\Phi)) + M'_{14}(\Theta, \Phi) V \right] d\Omega,$$

where I , Q , U , and V are the (*constant*) components of \mathbf{I}_{Lab} . If \mathbf{M} is independent of Φ , when the integral over Φ is carried out all the terms involving Q and U will integrate to zero, so for *linearly* polarized incident irradiance, this reduces to

$$b = \frac{n}{\kappa^2} \int M'_{11}(\Theta) d\Omega,$$

i.e., the same as for *unpolarized* incident irradiance. However, in general b will depend on the state of polarization of the incident irradiance. Henceforth, when we use the term “scattering coefficient,” it will be for *unpolarized* incident irradiance, unless otherwise specified.

or the equivalent in terms of coefficients

$$c = a + b,$$

where $c = n\sigma_c$, etc. In Appendix 1 to this chapter we derive the *optical theorem* that relates σ_c to the scattering amplitudes at $\Theta = 0$. The result (for unpolarized incident radiation) is⁴

$$\sigma_c = \frac{4\pi}{\kappa^2} \Re \left(\frac{A_1(0) + A_2(0)}{2} \right).$$

This provides σ_c from which σ_a can be found after determining σ_b from Eq. (3.10).

3.2 Rayleigh Scattering

In this section we develop the scattering of light by atoms and isotropic molecules (scalar polarizability), anisotropic molecules, and liquids. Such scattering is collectively called Rayleigh scattering, a term that is generically used for scattering by objects that are significantly smaller than the wavelength of the radiation. In Appendix 2 (Section @@.6.2) we explain Raman scattering — inelastic scattering similar in many respects to Rayleigh scattering.

3.2.1 Rayleigh Scattering by Atoms and Isotropic Molecules

Based on our classical model of an atom interacting with electromagnetic radiation, we found that the volume scattering matrix for a dilute gas of n atoms (or isotropic molecules) per unit volume was⁵

$$\beta(\Theta) = n \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2}\sin^2 \Theta & 0 & 0 \\ -\frac{1}{2}\sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix}.$$

The volume scattering function $\beta(\Theta) = \beta_{11}(\Theta)$ is

$$\beta(\Theta) = n \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right] \frac{1}{2}(1 + \cos^2 \Theta),$$

⁴See Appendix 1 for examples of extinction cross sections for various states of polarization of the incident irradiance.

⁵It is important to note that in deriving this result it was assumed that the scattered radiant power from each atom should be summed at the detector rather than the fields, i.e., there is no *coherence* between the scattering from two atoms. Thus, the result applies only to a situation in which the scatterers are in totally random motion, e.g., the gaseous state as opposed to the liquid or solid state.

and the scattering coefficient,

$$\begin{aligned} b &= \int_0^{2\pi} d\Phi \int_0^\pi \beta(\Theta) \sin \Theta d\Theta \\ &= \frac{8\pi}{3} n \left[\frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \right]. \end{aligned}$$

In these equations, $\kappa = 2\pi/\lambda$, α is the polarizability, and Θ is the scattering angle. Since α is an atomic property, we see that the scattering coefficient is proportional to n . Sometimes it is convenient to rewrite this by relating α to the refractive index m of the gaseous medium.⁶ This is effected through (see Chapter 1, Eq. (1.16))

$$n \frac{\alpha}{\epsilon_0} = 3 \left(\frac{m^2 - 1}{m^2 + 2} \right) \quad (3.11)$$

so,

$$\beta(\Theta) = \frac{9\pi^2}{n\lambda^4} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2 \left(\frac{1 + \cos^2 \Theta}{2} \right) \quad \text{and} \quad b = \frac{24\pi^3}{\lambda^4 n} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2. \quad (3.12)$$

If the equation for b is divided by n , the number density of atoms, we find that the scattering cross section per atom $\sigma_b = b/n$ is given by

$$\sigma_b = \frac{24\pi^3}{\lambda^4 n^2} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2. \quad (3.13)$$

Now, both m and n depend on temperature (T) and pressure (P) through the ideal gas law, $P = nk_B T$, where k_B is the Boltzmann constant ($1.38 \times 10^{-23} J/^\circ K$); however, the quantity

$$\frac{1}{n} \left(\frac{m^2 - 1}{m^2 + 2} \right) = \frac{\alpha}{3\epsilon_0} \quad (3.14)$$

cannot depend on T and P because it is a property of a *single* atom. Thus, σ_b is independent of P and T and

$$b = n\sigma_b = \left(\frac{P}{k_B T} \right) \sigma_b \quad (3.15)$$

provides the dependence of the scattering coefficient on temperature and pressure when the gas can be treated as ideal.

⁶The reader should understand that this relationship provides the refractive index m of a medium composed of individual atoms (each having a polarizability α), when the medium is treated as a continuum. Later, we will discuss homogeneous particles of refractive index m_p (which is related to the polarizability α_p of the atoms comprising the particle by a relationship identical to Eq. (3.11)), suspended in a continuous medium of index m_m . The relative index of the particle m_p/m_m is then written \tilde{m} or \tilde{m}_p .

3.2.2 Rayleigh Scattering by Anisotropic Molecules

By volume, 99% of the atmosphere consists of linear molecules (N_2 , O_2 , and CO_2). Such molecules will have a polarizability of the form

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{\perp} & 0 & 0 \\ 0 & \alpha_{\perp} & 0 \\ 0 & 0 & \alpha_{\parallel} \end{pmatrix},$$

where the subscripts \parallel and \perp refer to parallel and perpendicular to the symmetry axis of the molecule. Thus, one should include molecular anisotropy in the derivation of molecular scattering.

Scattering from an atom with anisotropic polarizability was discussed in Chapter 1. The results there can be used to find $\beta_{11}(\Theta) = \beta(\Theta)$ for such an atom as follows. We note that β_{11} describes the scattered Stokes parameter I for an unpolarized incident beam. Such an incident beam is represented by an uncorrelated (*incoherent*) mixture of radiation polarized parallel and perpendicular to the scattering plane with $E_{0r} = E_{0\ell}$. The Poynting vector for such a beam is

$$|\langle \vec{S}_{\text{Inc}} \rangle_{\text{Ave}}| = \frac{1}{2\mu_0 c} [(E_{0r})^2 + (E_{0\ell})^2].$$

The Stokes parameter I scattered by a single atom is proportional to the Poynting vector measured by a radiometer *without* polarizing filters, i.e.,

$$|\langle \vec{S} \rangle_{\text{Ave}}| = |\langle \langle \vec{S}_r \rangle_{\text{Ave}} \rangle| + |\langle \langle \vec{S}_\ell \rangle_{\text{Ave}} \rangle|,$$

where $\langle \langle \vec{S}_r \rangle_{\text{Ave}} \rangle$ and $\langle \langle \vec{S}_\ell \rangle_{\text{Ave}} \rangle$ are given by Eqs. (1.50) and (1.51) in Chapter 1. Therefore,

$$|\langle \vec{S} \rangle_{\text{Ave}}| = \frac{\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2 r^2} \left[\frac{45\alpha^2 + 13\beta^2 + (45\alpha^2 + \beta^2) \cos^2 \Theta}{45\alpha^2} \right] \frac{|\langle \vec{S}_{\text{Inc}} \rangle_{\text{Ave}}|}{2},$$

and following the definition of $\beta(\Theta)$ for a gas of number density n ,

$$nr^2 |\langle \vec{S} \rangle_{\text{Ave}}| = \beta(\Theta) |\langle \vec{S}_{\text{Inc}} \rangle_{\text{Ave}}|,$$

we have⁷

$$\beta_{11}(\Theta) = \beta(\Theta) = \frac{1}{2} \frac{n\kappa^4 \alpha^2}{(4\pi\epsilon_0)^2} \left[\frac{45\alpha^2 + 13\beta^2 + (45\alpha^2 + \beta^2) \cos^2 \Theta}{45\alpha^2} \right].$$

⁷Rather than changing notation for this one page, we simply warn the reader to be cautious about mixing $\beta(\Theta)$ on the left-hand side of the equation below with β^2 on the right-hand side which was defined just above (1.50).

Integrating, the scattering coefficient is

$$b = \frac{8\pi}{3} \frac{n\kappa^4\alpha^2}{(4\pi\epsilon_0)^2} \left[\frac{45\alpha^2 + 10\beta^2}{45\alpha^2} \right].$$

Recalling the depolarization factor at $\Theta = 90^\circ$,

$$\delta = \frac{6\beta^2}{45\alpha^2 + 7\beta^2},$$

it is easy to verify that

$$\begin{aligned} \frac{45\alpha^2 + \beta^2}{45\alpha^2 + 10\beta^2} &= \frac{2(1 - \delta)}{2 + \delta}, \\ \frac{45\alpha^2 + 13\beta^2}{45\alpha^2 + 10\beta^2} &= \frac{2(1 + \delta)}{2 + \delta}, \quad \text{and} \\ \frac{45\alpha^2 + 10\beta^2}{45\alpha^2} &= \frac{6 + 3\delta}{6 - 7\delta}. \end{aligned}$$

Thus,

$$b = \frac{8\pi}{3} \frac{n\kappa^4\alpha^2}{(4\pi\epsilon_0)^2} \left[\frac{6 + 3\delta}{6 - 7\delta} \right], \quad (3.16)$$

and

$$\beta(\Theta) = \frac{3}{16\pi} b \left[\frac{(1 + \delta) + (1 - \delta) \cos^2 \Theta}{1 + \delta/2} \right]. \quad (3.17)$$

The minimum value of δ is zero for $\beta^2 = 0$. The maximum anisotropy would occur when $\alpha_\perp = 0$. In this case $\delta = 0.5$. Thus, $0 \leq \delta \leq \frac{1}{2}$. The depolarization factor δ is straightforward to measure. Illuminate a sample with light passing through a polarizing filter and measure the light scattered at $\Theta = 90^\circ$ (using a detector without polarizing filters) first when the polarizer pass direction is parallel to the scattering plane and second when it is perpendicular to the scattering plane. The ratio of these measurements (first:second) is δ (See Chapter 1, Eqs. (1.50) and (1.51)). The measured value for air is near 0.03, so $(6 + 3\delta)/(6 - 7\delta) \approx 1.05$, showing that the presence of molecular anisotropy *increases* b by $\sim 5\%$.

Earlier, the polarizability was determined by relating α to m , which can be measured with high precision, through the Lorentz-Lorenz law, i.e., Eq. (3.14). Does this still apply when dealing with anisotropic dielectrics? For randomly orientated molecules the answer is yes. In the molecular frame the induced dipole moment is $p_i = \alpha_i E_i$, where $i = x, y$, or z . In the laboratory frame $P_I = n p_I$, where $I = X, Y$, or Z . From Eq. (1.49), averaging over orientations, we have

$$\langle p_I \rangle = \sum_{i=1}^3 \sum_{J=1}^3 \alpha_i \langle a_{iI} a_{iJ} \rangle E_J.$$

It is shown in the Mathematical Appendix that $\langle a_{iI} a_{iJ} \rangle = \delta_{IJ}/3$, so $\langle P_I \rangle = E_I \sum_{i=1}^3 \alpha_i / 3 = \alpha E_I$. Thus, $\langle \vec{p} \rangle = \alpha \vec{E}$, and the orientationally-averaged induced dipole moment is in the same direction as the field. The Lorentz-Lorenz law then follows as in Chapter 1. Thus, for example, the scattering cross section per molecule becomes

$$\sigma_b = \frac{24\pi^3}{\lambda^4 n^2} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2 \left[\frac{6 + 3\delta}{6 - 7\delta} \right]. \quad (3.18)$$

The full phase matrix for molecular scattering can be derived in a manner similar to β_{11} (note, however, that not all of the required results are derived in Chapter 1). The result is

$$\begin{aligned} \mathbf{P}(\Theta) &= \frac{4\pi\beta(\Theta)}{b} \\ &= \frac{3}{2} \left[\frac{1 - \delta}{1 + \delta/2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2}\sin^2 \Theta & 0 & 0 \\ -\frac{1}{2}\sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \\ &\quad + \frac{3}{2} \left[\frac{\delta}{1 + \delta/2} \right] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\cos \Theta \end{pmatrix}. \end{aligned} \quad (3.19)$$

Henceforth, when we used the term “Rayleigh scattering by a gas” (e.g., air), we are referring to Eqs. (3.16) through (3.19) and will use the notation b_r , β_r , σ_r , and \mathbf{P}_r , respectively.

3.2.3 Rayleigh Scattering by Liquids

One might expect that molecular scattering by liquids could be addressed by simply replacing n , the number density of atoms, in Eq. (3.18) by the appropriate number density for liquids, which is much larger than that for gases. However, experimentally it is known that a given number of atoms (or a given mass) scatters more in the gaseous state than in the liquid state, and in the solid state, transparent materials exhibit almost no scattering. The reason is that, while in the gaseous state atoms (molecules) move rapidly and independently, in the liquid state there are strong correlations between the motions of the atoms in small volumes. These correlations of the motions lead to some coherent cancellation of the scattering of radiation from the individual atoms. If we divide the scattering volume into individual cells, because of the correlations and the cancellations, each cell will *appear* to scatter as if there were fewer atoms than are actually present. Thus, in a reformulation

of scattering theory for application to liquids, we will treat each cell as a scattering center that *coherently* scatters light with all other cells in a sample.⁸

In this section we reformulate the molecular scattering theory along the lines suggested above. We shall show that the volume scattering of a medium can be explained on the basis of fluctuations in the number density of atoms within the cells. In addition we show that in the case of gases, the final result is virtually identical to that obtained in the previous section. Following this, the extension to liquids is immediate.

It is natural to divide the medium into cubic cells; however, we will use spherical cells because the scattering by such spheres has already been described.⁹ From Chapter 1 we found that in a vacuum, the scattering by a sphere with radius $R \ll \lambda$ was the same as that from an atom with the atomic polarizability replaced by

$$\alpha = 4\pi\epsilon_0 R^3 \left[\frac{K_e - 1}{K_e + 2} \right],$$

where K_e is the dielectric constant of the sphere. Assuming that $K_e - 1 \triangleq \Delta K \ll 1$, α can be written

$$\alpha = \frac{4\pi\epsilon_0}{3} R^3 \Delta K = \epsilon_0 V_s \Delta K,$$

where V_s is the volume of the sphere. Now, we model a continuous medium as a collection of N such spheres labeled by the index i , with the i^{th} sphere having a polarizability $\alpha_i = \epsilon_0 V_s \Delta K_i$. Then the scattered field at P in Figure 3.1 from the i^{th} sphere (centered at \mathcal{O}_i) is

⁸In a crystal lattice, each atom is essentially fixed in space, and for the scattered field of an individual atom, say atom a , another atom (b) can always be found somewhere in the crystal (assumed infinite) that will produce a scattered field that will exactly cancel the field of atom a in any given direction (different atoms b will be required for different directions), except in the forward scattering direction where the fields add. Forward scattering is simply the propagation of the incident beam.

⁹Within the context of Rayleigh-Gans scattering theory to be described later, one can show that the scattering by a small cube and a small sphere are identical in the limit that the size is $\ll \lambda$.

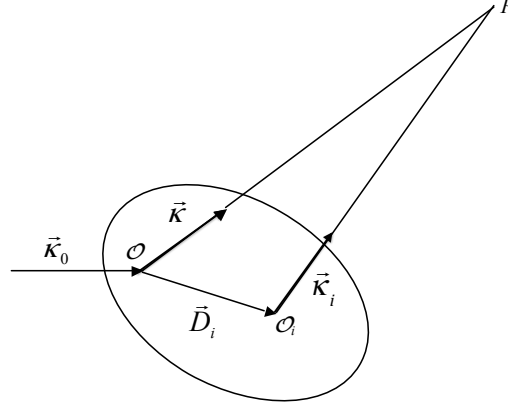


Figure 3.1: Radiation propagating in the direction κ_0 in a medium represented as a collection of spheres is incident on two representative spheres. The scattered field is desired at point P . The vector \vec{r} is directed from \mathcal{O} to P and has length \mathcal{OP} . The vector \vec{r}_i is directed from \mathcal{O}_i to P and has length \mathcal{O}_iP . \vec{D}_i specifies the position of the sphere at \mathcal{O}_i with respect to the sphere at \mathcal{O} . The point \mathcal{O} is taken as the origin of the coordinate system.

$$\begin{pmatrix} \Delta E_r(P, \mathcal{O}_i) \\ \Delta E_\ell(P, \mathcal{O}_i) \end{pmatrix} = \frac{-i\kappa^3}{4\pi\epsilon_0} V_s \Delta K_i \frac{\exp[i(\kappa_i r_i - \omega t)]}{-i\kappa_i r_i} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} E_{0r}(\mathcal{O}_i) \\ E_{0\ell}(\mathcal{O}_i) \end{pmatrix}.$$

To find the total field at P , we need to sum contributions from each sphere. To carry this out, it is useful to write the fields at \mathcal{O}_i on the left-hand-side of this equation in terms of the field at a single point in space, which we take to be the point \mathcal{O} in Figure 3.1. Then, with the origin of coordinates at \mathcal{O} , Eq. (3.1) provides the incident field at \vec{D}_i : $\vec{E}(\vec{D}_i, t) = \vec{E}_0 \exp[i(\vec{\kappa}_0 \cdot \vec{D}_i - \omega t)]$. But the field at \mathcal{O} is just $\vec{E}_0 \exp[-i\omega t]$, so the incident field at \mathcal{O}_i is related to that at \mathcal{O} by

$$\vec{E}_{\text{Inc}}(\mathcal{O}_i) = \vec{E}_{\text{Inc}}(\mathcal{O}) \exp[i\vec{\kappa}_0 \cdot \vec{D}_i],$$

where the vector \vec{D}_i is provided in Figure 3.1. Then the scattered fields are

$$\begin{pmatrix} \Delta E_r(P, \mathcal{O}_i) \\ \Delta E_\ell(P, \mathcal{O}_i) \end{pmatrix} = \frac{-i\kappa^3}{4\pi} V_s \Delta K_i \exp[i\vec{\kappa}_0 \cdot \vec{D}_i] \frac{\exp[i(\kappa_i r_i - \omega t)]}{-i\kappa_i r_i} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} E_{0r}(\mathcal{O}) \\ E_{0\ell}(\mathcal{O}) \end{pmatrix}.$$

Now, $\vec{r} = \vec{r}_i + \vec{D}_i$, so $\kappa_i r_i = \vec{\kappa}_i \cdot \vec{r}_i = \vec{\kappa}_i \cdot (\vec{r} - \vec{D}_i)$, and

$$\exp[i\vec{\kappa}_0 \cdot \vec{D}_i] \exp[i\kappa_i r_i] = \exp[i(\vec{\kappa}_0 - \vec{\kappa}_i) \cdot \vec{D}_i] \exp[i\kappa_i r].$$

Therefore

$$\begin{pmatrix} \Delta E_r(P, \mathcal{O}_i) \\ \Delta E_\ell(P, \mathcal{O}_i) \end{pmatrix} \frac{-i\kappa^3}{4\pi} V_s \Delta K_i \exp[i(\vec{\kappa}_0 - \vec{\kappa}_i) \bullet \vec{D}_i] \frac{\exp[i(\kappa_i r - \omega t)]}{-i\kappa_i r_i} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} E_{0r}(\mathcal{O}) \\ E_{0\ell}(\mathcal{O}) \end{pmatrix}.$$

Noting that $\kappa_0 = \kappa_i = \kappa$, assuming that r is so much larger than $|\vec{D}_i|$ that $\vec{\kappa}_i = \vec{\kappa}$, and replacing r' by r in the denominator,¹⁰ we have

$$\begin{pmatrix} \Delta E_r(P, \mathcal{O}_i) \\ \Delta E_\ell(P, \mathcal{O}_i) \end{pmatrix} \frac{-i\kappa^3}{4\pi} V_s \Delta K_i \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] \frac{\exp[i(\kappa r - \omega t)]}{-i\kappa r} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} E_{0r}(\mathcal{O}) \\ E_{0\ell}(\mathcal{O}) \end{pmatrix}. \quad (3.20)$$

Summing the *fields* over all N spheres, and using the definition of the scattering amplitudes, we have

$$A_1(\Theta) = \frac{-i\kappa^3}{4\pi} V_s \sum_{i=1}^N \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] \Delta K_i$$

$$A_2(\Theta) = \frac{-i\kappa^3}{4\pi} V_s \cos \Theta \sum_{i=1}^N \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] \Delta K_i.$$

To determine the volume scattering function we must compute $A_1 A_1^*$ and $A_2 A_2^*$. However, since A_1 and A_2 are fluctuating (in time) quantities, we actually need their time averages, e.g., $\langle A_1 A_1^* \rangle$. The first is just

$$\langle A_1 A_1^* \rangle = \frac{\kappa^6}{(4\pi)^2} V_s^2 \sum_{i=1}^N \sum_{j=1}^N \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet (\vec{D}_i - \vec{D}_j)] \langle \Delta K_i \Delta K_j \rangle$$

If we assume that the ΔK 's for the various spheres are completely random (uncorrelated), then $\langle \Delta K_i \Delta K_j \rangle = 0$ when $i \neq j$, and

$$\langle A_1 A_1^* \rangle = \frac{\kappa^6}{(4\pi)^2} V_s^2 \sum_{i=1}^N \langle (\Delta K_i)^2 \rangle = \frac{\kappa^6}{(4\pi)^2} V_s^2 N \langle (\Delta K)^2 \rangle,$$

where $\langle (\Delta K)^2 \rangle = N^{-1} \sum_{i=1}^N (\Delta K_i)^2$ is mean of $(\Delta K)^2$ over all the spheres. Writing the sphere volume $V_s = V/N = 1/n$, where V is the total volume of the region under consideration and n is the number density of spheres, we have

$$\frac{\langle A_1 A_1^* + A_2 A_2^* \rangle}{2} = \frac{1}{2} \frac{\kappa^6}{(4\pi)^2 n} V \langle (\Delta K)^2 \rangle (1 + \cos^2 \Theta)$$

¹⁰The reader may wonder why we replace r_i with r some places but not others. When r_i is a multiplicative factor (as in the denominator) the replacement will make only a small difference in the final result, but if it is in the phase, i.e., $\kappa_i r_i$ in $\exp[i\kappa_i r_i]$, a small difference between \vec{r} and \vec{r}_i can make a large difference in $\exp[i\kappa_i r_i]$.

so

$$\beta(\Theta) = \frac{1}{2} \frac{\kappa^4}{(4\pi)^2} V \langle (\Delta K)^2 \rangle (1 + \cos^2 \Theta) = \frac{\pi^2}{2\lambda^4} V \langle (\Delta K)^2 \rangle (1 + \cos^2 \Theta).$$

The problem of determining the mean square fluctuation in the dielectric constant $\langle (\Delta K)^2 \rangle$ is solved by the methods of statistical thermodynamics. If the density ρ and the absolute temperature T are taken to be the independent thermodynamic variables specifying the state of the substance, then

$$\Delta K = \left(\frac{\partial K_e}{\partial \rho} \right)_T \Delta \rho + \left(\frac{\partial K_e}{\partial T} \right)_\rho \Delta T,$$

and since temperature and density fluctuations are independent, $\langle \Delta \rho \Delta T \rangle = 0$, and

$$\langle (\Delta K)^2 \rangle = \left(\frac{\partial K_e}{\partial \rho} \right)_T^2 \langle (\Delta \rho)^2 \rangle + \left(\frac{\partial K_e}{\partial T} \right)_\rho^2 \langle (\Delta T)^2 \rangle.$$

The fluctuations in temperature and density can be shown to be

$$\langle (\Delta T)^2 \rangle = \frac{k_B T^2}{\rho c_v V} \quad \text{and} \quad \langle (\Delta \rho)^2 \rangle = \frac{\rho k_B T}{V} \left(\frac{\partial \rho}{\partial P} \right)_T,$$

where c_v is the specific heat at constant volume. Now, using the fact that the density is the mass of the material divided by V , we can replace the partial derivative in the last expression by

$$\left(\frac{\partial \rho}{\partial P} \right)_T = \rho \beta_T, \quad \text{where} \quad \beta_T \triangleq -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T,$$

is the isothermal compressibility. Then

$$V \langle (\Delta K)^2 \rangle = \left(\frac{\partial K_e}{\partial \rho} \right)_T^2 \rho^2 T \beta_T k_B + \left(\frac{\partial K_e}{\partial T} \right)_\rho^2 \frac{k_B T}{\rho c_v}.$$

The volume scattering function is then determined given the thermodynamic properties of the medium.

It is easy to apply these relationships to an ideal gas. The equation of state is $PV = Nk_B T$, where N is the number of atoms in V . Carrying out the derivative, $\beta_T = 1/P$, so $k_B T \beta_T = V/N = 1/n$. In addition, n and K are related by

$$\frac{\alpha}{3\epsilon_0} = \frac{1}{n} \left(\frac{K-1}{K+2} \right) = \text{Constant},$$

and

$$\left(\frac{\partial K_e}{\partial \rho} \right)_T = \frac{1}{M_a} \left(\frac{\partial K_e}{\partial n} \right)_T,$$

where M_a is the mass of a single atom. Differentiating $\alpha/3\epsilon_0$ the result is

$$\left(\frac{\partial K_e}{\partial \rho}\right)_T^2 = \frac{1}{9\rho}(K-1)^2(K+2)^2.$$

If the density is constant, the dielectric constant is as well, so

$$\left(\frac{\partial K_e}{\partial T}\right)_\rho = 0,$$

and finally

$$\beta(\Theta) = \frac{\pi^2}{\lambda^4} \frac{1}{n} \frac{(K_e - 1)^2 (K_e + 2)^2}{9} \left(\frac{1 + \cos^2 \Theta}{2} \right).$$

Since $K - 1 \ll 1$, this is approximately

$$\beta(\Theta) = \frac{\pi^2}{\lambda^4} \frac{1}{n} (K_e - 1)^2 \left(\frac{1 + \cos^2 \Theta}{2} \right),$$

which is identical to Eq. (3.12) with $m^2 - 1 \ll$ and $m^2 + 2 \approx 3$. Thus, we have been able to re-derive the scattering by molecules in a gas by considering fluctuations in its density.

For a liquid consisting of anisotropic molecules, e.g., water, the factor relating σ_b to molecular anisotropy, i.e., $(6 + 3\delta)/(6 - 7\delta)$, and the substitution

$$1 + \cos^2 \Theta \longrightarrow \frac{1 + \delta + (1 - \delta) \cos^2 \Theta}{1 + \delta/2}$$

can also be deduced in a manner similar to the scattering by an anisotropic molecule by considering fluctuations in molecular orientation within the individual cells. In the anisotropic case, K for an individual cell is replaced by a diagonal matrix (in a reference frame in the cell, but with axes that fluctuate continuously as anisotropic molecules enter and leave it) with elements that differ only slightly. In this case, K in the above equation can be taken to be the mean of the diagonal elements of the matrix.

Although we have assumed in this section that $K_e - 1 \ll 1$, an identical expression is obtained even if this limitation is not imposed (it was here only for convenience), i.e., there really need not be any restriction on K to obtain the final result, and it holds for all K . In particular, for water $K = m^2 \approx 1.77$. Thus, the results of this section can be applied to liquids as well as gases: for either,

$$\beta(\Theta) = \frac{\pi^2}{2\lambda^4} \left[\left(\frac{\partial K_e}{\partial \rho}\right)_T^2 \rho^2 T \beta_T k_B + \left(\frac{\partial K_e}{\partial T}\right)_\rho^2 \frac{k_B T}{\rho c_v} \right] \left(\frac{6 + 3\delta}{6 - 7\delta} \right) \left(\frac{(1 + \delta) + (1 - \delta) \cos^2 \Theta}{1 + \delta/2} \right),$$

and the scattering is determined by the thermodynamic properties.

3.3 Scattering by a Homogeneous Sphere

Most attempts to understand light scattering from particles suspended in the air or in water assume that the particles are homogeneous spheres. In this section we examine the scattering by such particles. We begin by presenting the exact solution for the elements of the scattering matrix $\mathbf{A}(\Theta)$, and then consider some useful approximations.

3.3.1 The Exact Solution (Mie Theory)

The problem of the scattering of a plane electromagnetic wave by a sphere of arbitrary refractive index \tilde{m} , relative to the surrounding medium, and radius R , was first solved by G. Mie in 1908. Deriving the Mie solution from first principles would take us too far afield, so we shall simply summarize the results.

Writing the scattered field in the form of Eq. (3.6), the scattering amplitudes are given by

$$\begin{aligned} A_1(\Theta, \tilde{m}, R) &= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{\ell(\ell+1)} \left[a_{\ell}(x, y) \pi_{\ell}(\cos(\Theta)) + b_{\ell}(x, y) \tau_{\ell}(\cos(\Theta)) \right] \\ A_2(\Theta, \tilde{m}, R) &= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{\ell(\ell+1)} \left[a_{\ell}(x, y) \tau_{\ell}(\cos(\Theta)) + b_{\ell}(x, y) \pi_{\ell}(\cos(\Theta)) \right], \end{aligned} \quad (3.21)$$

where $x = \kappa R$, $y = \tilde{m}x = \tilde{m}\kappa R$, and ℓ is an integer.¹¹ The other quantities in these amplitudes are as follows:

$$\begin{aligned} a_{\ell}(x, y) &= \frac{\psi'_{\ell}(y)\psi_{\ell}(x) - \tilde{m}\psi_{\ell}(y)\psi'_{\ell}(x)}{\psi'_{\ell}(y)\xi_{\ell}(x) - \tilde{m}\psi_{\ell}(y)\xi'_{\ell}(x)} \\ b_{\ell}(x, y) &= \frac{\tilde{m}\psi'_{\ell}(y)\psi_{\ell}(x) - \psi_{\ell}(y)\psi'_{\ell}(x)}{\tilde{m}\psi'_{\ell}(y)\xi_{\ell}(x) - \psi_{\ell}(y)\xi'_{\ell}(x)} \end{aligned} \quad (3.22)$$

where

$$\psi_{\ell}(z) = \sqrt{\frac{\pi z}{2}} J_{\ell+\frac{1}{2}}(z) \quad \text{and} \quad \xi_{\ell}(z) = \sqrt{\frac{\pi z}{2}} H_{\ell+\frac{1}{2}}^{(1)}(z),$$

with J and $H^{(1)}$ representing Bessel and Hankel functions, respectively. The functions ψ_n

¹¹Note that \tilde{m} is the refractive index of the particle divided by the refractive index of the medium in which the particle is suspended — in cases of interest to us, water or air.

and ξ_n satisfy the following recursion relationships:¹²

$$\begin{aligned} g_{\ell-1}(z) + g_{\ell+1}(z) &= \frac{2\ell+1}{z} g_{\ell}(z), \\ \ell g_{\ell-1}(z) - (\ell+1)g_{\ell+1}(z) &= (2\ell+1)g'_{\ell}(z), \end{aligned}$$

where $g = \psi$ or ξ , and $'$ indicates derivative with respect to the argument, i.e., $g'_{\ell}(z) = dg_{\ell}(z)/dz$.

The π 's and τ 's are related to the Legendre polynomials¹³ and satisfy the following relationships:

$$\begin{aligned} \pi_{\ell}(z) &= z \frac{2\ell-1}{\ell-1} \pi_{\ell-1}(z) - \frac{\ell}{\ell-1} \pi_{\ell-2}(z), & \text{with } \pi_0(z) &= 0, & \text{and } \pi_1(z) &= 1, \\ \pi'_{\ell}(z) &= (2\ell-1) \pi_{\ell-1}(z) - \pi'_{\ell-2}(z), & \text{with } \pi'_0(z) &= 0, & \text{and } \pi'_1(z) &= 0, \\ \tau_{\ell}(z) &= z \pi_{\ell}(z) - \sqrt{(1-z^2)} \pi'_{\ell}(z). \end{aligned}$$

These can be used directly to show

$$\pi_{\ell}(-z) = (-1)^{\ell-1} \pi_{\ell}(z) \quad \text{and} \quad \tau_{\ell}(-z) = (-1)^{\ell} \tau_{\ell}(z),$$

as well as

$$\pi_{\ell}(1) = \tau_{\ell}(1) = \frac{\ell(\ell+1)}{2}. \quad (3.23)$$

They also satisfy the integral relationships (with m another integer),

$$\begin{aligned} \int_{-1}^1 [\pi_{\ell}(z) \pi_m(z) + \tau_{\ell}(z) \tau_m(z)] dz &= \left(\frac{2\ell^2(\ell+1)^2}{2\ell+1} \right) \delta_{\ell m}, \\ \int_{-1}^1 [(\pi_{\ell}(z) \pm \tau_{\ell}(z))(\pi_m(z) \pm \tau_m(z))] dz &= 0, \quad \text{if } \ell \neq m, \end{aligned} \quad (3.24)$$

and, since $\pi_{\ell}(-z) \tau_{\ell}(-z) = -\pi_{\ell}(z) \tau_{\ell}(z)$,

$$\int_{-1}^1 \pi_{\ell}(z) \tau_{\ell}(z) dz = 0. \quad (3.25)$$

¹²The Hankel function described here is $H_{\ell+1/2}^{(1)}(z) = J_{\ell+1/2}(z) + iN_{\ell+1/2}(z)$, where N is the Bessel function of the second kind (singular at $z = 0$). For large z , $\xi_{\ell}(z) \propto z^{-1} \exp[iz]$. Also, the recursion relations must be used with care, particularly in numerical computations. We refer the reader to treatises on light scattering for detail.

¹³The relationships are $\pi_{\ell}(z) = dP_{\ell}(z)/dz$, with $P_{\ell}(z) = \frac{1}{2^{\ell} \ell!} \frac{d^{\ell}}{dz^{\ell}} (z^2 - 1)^{\ell}$, where $P_{\ell}(z)$ is the Legendre polynomial of order ℓ .

Direct substitution of Eq. (3.21) into Eq. (3.7) yields

$$\begin{aligned}\beta(\Theta) &= \frac{n}{\kappa^2} M'_{11}(\Theta) = \frac{n}{2\kappa^2} (A_1(\Theta)A_1^*(\Theta) + A_2(\Theta)A_2^*(\Theta)) \\ &= \frac{n}{2\kappa^2} \sum_{\ell} \sum_m \left[\frac{2\ell+1}{\ell(\ell+1)} \right] \left[\frac{2m+1}{m(m+1)} \right] \\ &\quad \times \left[(a_{\ell}a_m^* + b_{\ell}b_m^*)(\pi_{\ell}\pi_m + \tau_{\ell}\tau_m) + 2(a_{\ell}b_m^* + a_{\ell}^*b_m)(\pi_{\ell}\tau_m) \right]\end{aligned}\quad (3.26)$$

The scattering coefficient is

$$b = \int_{\text{All } \Omega} \beta(\Theta) d\Omega = \int_0^{2\pi} d\Phi \int_0^{\pi} \beta(\Theta) \sin \Theta d\Theta = 2\pi \int_{-1}^1 \beta(u) du, \quad (3.27)$$

where $u = \cos \Theta$. Using Eqs. (3.24) and (3.25), we have

$$b = \frac{2\pi n}{\kappa^2} \sum_{\ell} (2\ell+1)(|a_{\ell}|^2 + |b_{\ell}|^2), \quad \text{or} \quad \sigma_b = \frac{2\pi}{\kappa^2} \sum_{\ell} (2\ell+1)(|a_{\ell}|^2 + |b_{\ell}|^2).$$

It is useful to define the *scattering efficiency* $Q_b \triangleq \sigma_b/\pi R^2$, i.e., the ratio of the scattering cross section to the geometrical cross section of the sphere. Thus

$$Q_b = \frac{2}{(\kappa R)^2} \sum_{\ell} (2\ell+1)(|a_{\ell}|^2 + |b_{\ell}|^2).$$

The total attenuation (extinction) cross section is given by the optical theorem (Appendix 1)

$$\sigma_c = \frac{4\pi}{\kappa^2} \Re \left[\frac{A_1(0) + A_2(0)}{2} \right] = \frac{2\pi}{\kappa^2} \sum_{\ell} (2\ell+1) \Re(a_{\ell} + b_{\ell}),$$

and the absorption cross section is $\sigma_a = \sigma_c - \sigma_b$. Defining the absorption and attenuation (extinction) efficiencies, $Q_a \triangleq \sigma_a/\pi R^2$ and $Q_c \triangleq \sigma_c/\pi R^2$, we have

$$Q_c = Q_a + Q_b.$$

The left panel in Figure 3.2 provides an example of $\beta(\Theta)$ for a sphere of radius $R = 5\lambda$ and several values of \tilde{m} . Notable features are the strong oscillations as Θ is varied, the strong scattering in the near-forward direction compared to all others, and the general increase in scattering at large angles as \tilde{m} increases. The right panel provides the degree of polarization P of the scattered light when the incident light is unpolarized. Recall that

$$P \triangleq \frac{\sqrt{Q^2 + U^2 + V^2}}{I} = \sqrt{\frac{M_{12}^2}{M_{11}^2}} = \pm \frac{M_{12}}{M_{11}} = \pm \frac{\beta_{12}}{\beta_{11}}.$$

It is customary to use the $-$ sign in the definition so that for Rayleigh scattering P at 90° is $+1$. Note that for \tilde{m} near unity, i.e., $\tilde{m} - 1 \ll 1$, the degree of polarization is similar to that for Rayleigh scattering ($\sin^2(\Theta)/(1 + \cos^2 \Theta)$).

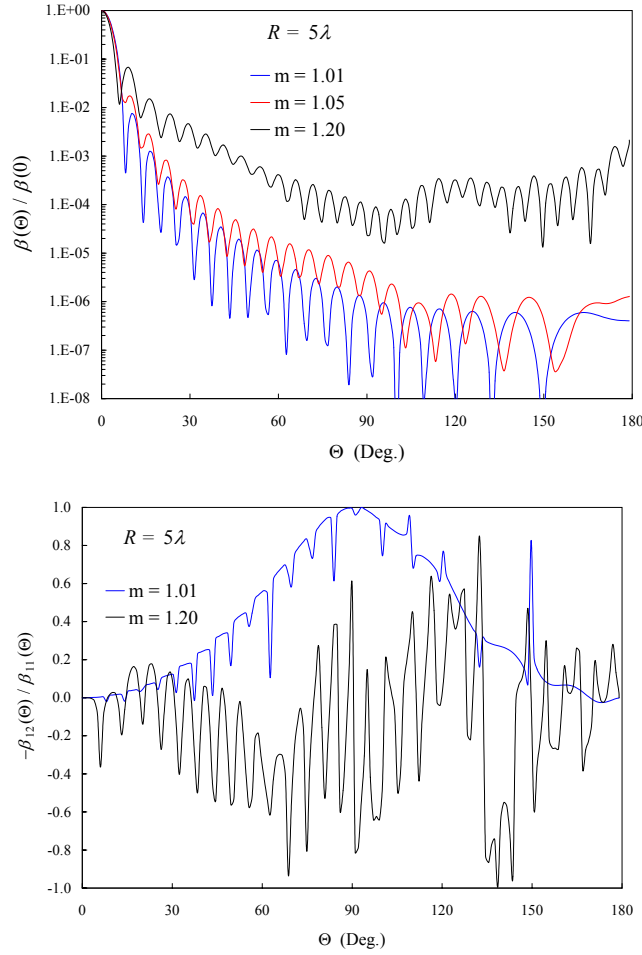


Figure 3.2: Scattering by a sphere with $R = 5\lambda$ for various values of \tilde{m} . Left panel: the volume scattering function normalized to that at $\Theta = 0^\circ$. Right panel: the degree of polarization.

Figure 3.3 provides Q_b as a function of the size of the sphere for the same range of \tilde{m} as in Figure 3.2. In this figure, the size is scaled by the parameter $\rho_r = 2\kappa R(\tilde{m}_r - 1)$, where \tilde{m}_r is the real part of \tilde{m} . In the geometrical optics approximation, ρ_r is the phase shift of the ray that goes through the center of the particle relative to a ray that misses the particle. The scaling is suggested by an approximate theory for large particles discussed later in this chapter. Note that the scattering efficiencies for the various values of \tilde{m} all have a similar shape when plotted in this manner. For the cases plotted in Figure 3.2 (i.e., $R = 5\lambda$), $2\kappa R = 20\pi$, so $\rho_r = 0.628, 3.14$, and 12.56 , for $\tilde{m} = 1.01, 1.05$, and 1.20 ,

respectively. Thus, the scattering cross section for $\tilde{m} = 1.20$ is much larger than that for $\tilde{m} = 1.01$.

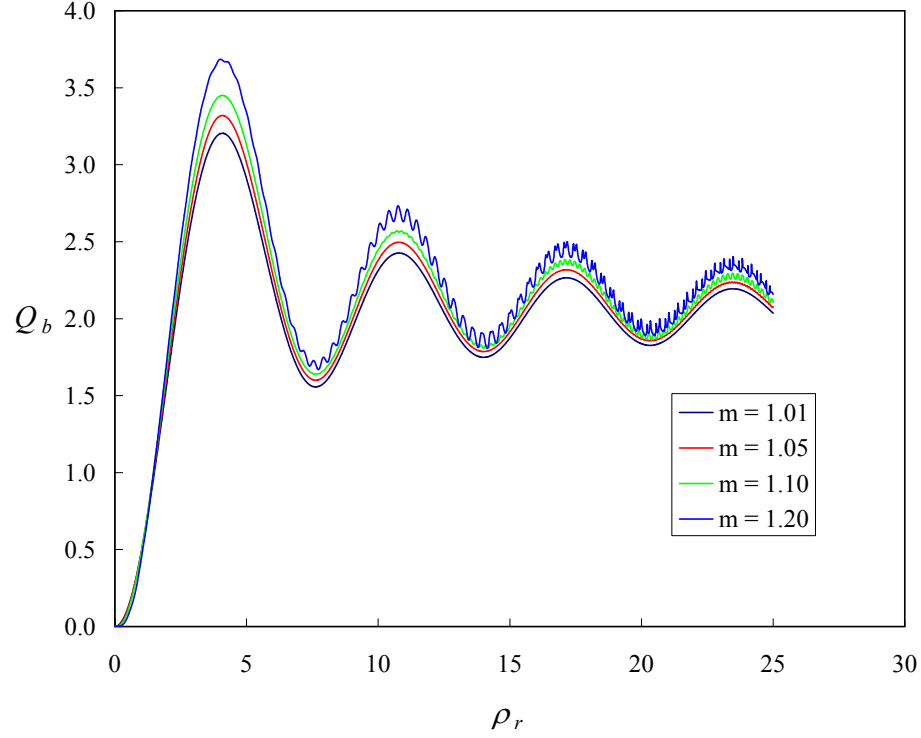


Figure 3.3: The scattering efficiency Q_b , computed by the exact Mie theory, as a function of the parameter $\rho_r = 2\kappa R(\tilde{m}_r - 1)$ for various values of $\tilde{m} = \tilde{m}_r + i\tilde{m}_i$.

Finally, Figure 3.4 provides Q_c and Q_a for $\tilde{m} = 1.05 + i\tilde{m}_i$. The figure shows that as \tilde{m}_i increases (increasing absorption) the oscillations in Q_c in Figure 3.3 are damped, and Q_c rapidly reaches its large-sphere limit of 2. This also shows that the large-sphere limit to Q_a is 1 for $\tilde{m}_i \neq 0$, implying that the large-sphere limit to Q_b for a sphere with *any* absorption at all is unity. In contrast, if there is *no* absorption, the large-sphere limit to Q_b is 2.

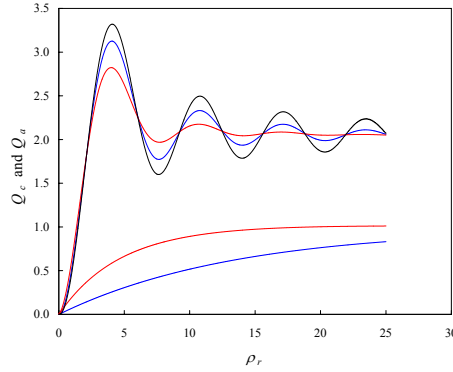


Figure 3.4: Q_c (upper curves) and Q_a (lower curves), computed by the exact Mie theory, as a function of $\rho_r = 2\kappa R(\tilde{m}_r - 1)$ for $\tilde{m} = 1.05 + i\tilde{m}_i$. Black: $\tilde{m}_i = 0$. Blue: $\tilde{m}_i = 0.0025$. Red: $\tilde{m}_i = 0.0075$.

Methods for the exact solution to the problem of scattering by particles of other simple shapes, e.g., coated spheres, spheroids and cylinders, have been developed and are available for numerical computation. In fact, it is even possible to obtain numerical results for particles of arbitrary shape; however, the computational burden is great (see Section @@4.1). Most models of scattering by marine and atmospheric particles are based on Mie theory, i.e., the assumption that the particles are homogeneous spheres. Mie theory serves as a useful guide to understanding how the various parameters influence light scattering.

3.3.2 Spherical Particles Distributed in Size and Composition

Equations (3.26) and (3.27) provide the volume scattering function and the scattering coefficient for a number (n) of spheres per unit volume of radius R and composition reflected in the refractive index \tilde{m} .¹⁴ Consider the i^{th} species of sphere with refractive index \tilde{m}_i . If we have a mixture of different sizes of this species of spheres, each size will contribute independently to scattering properties, e.g., to $\beta_i(\Theta)$, the volume scattering function for the i^{th} species. Let $dN_i(R)$ be the number of particles of the i^{th} species per unit volume with radii between R and $R + dR$. Then for the i^{th} species the volume scattering function is

$$\begin{aligned}\beta_i(\Theta) &= \frac{1}{\kappa^2} \int_0^\infty M'_{11}(\Theta, R, \tilde{m}_i) dN_i(R) \\ &= \frac{1}{\kappa^2} \int_0^\infty M'_{11}(\Theta, R, \tilde{m}_i) n_i(R) dR,\end{aligned}$$

¹⁴In a mixture of several species of particles, each species will scatter independently of the others.

where $n_i(R) \triangleq dN_i(R)/dR$ is the size distribution of the i^{th} species, and the dependence of M'_{11} on particle size and refractive index has been explicitly displayed in the argument list. Then for the entire collection of particles,

$$\beta(\Theta) = \sum_i \beta_i(\Theta).$$

Likewise,

$$c = \sum_i \int_0^\infty \sigma_c(R, \tilde{m}_i) n_i(R) dR = \sum_i \int_0^\infty \pi R^2 Q_c(R, \tilde{m}_i) n_i(R) dR,$$

etc.

3.3.3 Approximations for Small and Large Spheres

Although the Mie theory provides exact equations for the scattering and extinction of light by homogeneous spheres, they are too complex to provide a qualitative understanding of the phenomena involved. With some restriction on size and refractive index, it is possible to arrive at simpler, closed-form, approximate solutions that have been useful in environmental optics and which, in addition, can be applied to non-spherical particles as well. We examine several of these approximate solutions in this section.

3.3.3.1 The Rayleigh-Gans Approximation: $|\tilde{m} - 1| \ll 1$ and $\kappa R |\tilde{m} - 1| \ll 1$

We saw in Chapter 1 that if a homogeneous sphere is sufficiently small it will scatter in a manner identical to Rayleigh scattering with α replaced by $4\pi\epsilon_0 R^3(m_p^2 - 1)/(m_p^2 + 2)$, where m_p is the refractive index of the material comprising the sphere. When the sphere is placed in a medium with refractive index m_m , then one must replace m_p in this formula by $\tilde{m} = m_p/m_m$.¹⁵ The requirement “sufficiently small” means that the sphere must be

¹⁵Clearly, the polarizability of the sphere will not change by placing it in a dielectric medium; however, for the purposes of light scattering, the important consideration is the polarizability of the medium *relative to its surroundings*. If the medium were to have the *same* index as the particle, there would be no scattering (actually there would be forward scattering, which manifests as the propagation of the incident beam *without* scattering). Our simple procedure for relating α to m in Chapter 1 is difficult to extend to a particle in a dielectric medium; however, more advanced methods show that if a sphere is immersed in a medium of dielectric constant K_m , the resulting analogue to Eq. (1.52) is

$$\vec{p} = 4\pi\epsilon_0 R^3 \left[\frac{K_e - K_m}{K_e + 2K_m} \right] \vec{E}_m,$$

where E_m is the field *in* the medium.

small enough that at a given instant of time, the incident electric field (in the medium) is spatially uniform over the particle, i.e., the electric field everywhere in the particle is the same as at the center. What if the particle is too large to satisfy this constraint? Consider such a particle and divide it into a collection of small volumes. If the effective polarizability of the atoms composing the material is α , then

$$\alpha = \frac{3\epsilon_0}{n} \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right),$$

where \tilde{m} is the refractive index of the material composing the particle, relative to that of the medium. A volume ΔV will contain N atoms, where $N = n\Delta V$, so the dipole moment of ΔV will be

$$\Delta \vec{p} = N\alpha \vec{E} = 3\epsilon_0 \Delta V \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \vec{E}, \quad (3.28)$$

where \vec{E} is the total field *inside* the particle.¹⁶ This field is difficult to calculate, but if we assume $m_r \approx 1$ and $m_i \ll 1$, the field inside the particle is approximately the field outside, i.e., \vec{E}_{Inc} . This condition can be succinctly stated as requiring $|\tilde{m} - 1| \ll 1$.¹⁷ In this case the volume will acquire a dipole moment

$$\Delta \vec{p} = 3\epsilon_0 \Delta V \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \vec{E}_{\text{Inc}}. \quad (3.29)$$

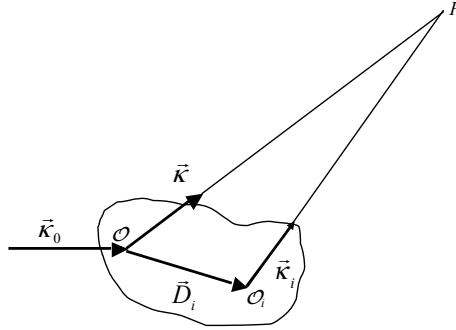


Figure 3.5: Radiation propagating in the direction κ_0 is incident on a particle of arbitrary shape. The scattered field is desired at point P . The vector \vec{r} is directed from \mathcal{O} to P and has length $\mathcal{O}P$. The vector \vec{r}_i is directed from \mathcal{O}_i to P and has length \mathcal{O}_iP . The volume element ΔV_i is located at \mathcal{O}_i , and \vec{D}_i specifies the position of \mathcal{O}_i with respect to \mathcal{O} . The point \mathcal{O} is the origin of the coordinate system and the electric field at \mathcal{O} is taken to be $\vec{E}_0 \exp[-i\omega t]$.

¹⁶Since ΔV is an infinitesimal, this field will always be spatially uniform over ΔV .

¹⁷If we write $\tilde{m}_r = 1 + \delta$, then $|\tilde{m} - 1| = \delta^2 + \tilde{m}_i^2$, so $|\tilde{m} - 1| \ll 1$ implies $\delta \ll 1$ (or $\tilde{m}_r \approx 1$) and $\tilde{m}_i \ll 1$.

Since \vec{E}_{Inc} is oscillatory, ΔV will possess an oscillating dipole moment, and radiate, i.e., scatter the incident radiation. The scattered fields at a point P (Figure 3.5) due to ΔV_i will be¹⁸

$$\begin{pmatrix} \Delta E_r(P, \mathcal{O}_i) \\ \Delta E_\ell(P, \mathcal{O}_i) \end{pmatrix} = \frac{-i\kappa^3}{4\pi\epsilon_0} \Delta V_i \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \frac{\exp[i\kappa_i r_i - \omega t]}{-i\kappa_i r_i} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} E_{0r}(\mathcal{O}_i) \\ E_{0\ell}(\mathcal{O}_i) \end{pmatrix}.$$

To find the total field at P , we need to sum contributions from all parts of the particle. In a manner similar to the derivation of Eq. (3.20) we find

$$\begin{aligned} \begin{pmatrix} \Delta E_r(P, \mathcal{O}_i) \\ \Delta E_\ell(P, \mathcal{O}_i) \end{pmatrix} &= \frac{-i3\kappa^3}{4\pi} \Delta V_i \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] \\ &\quad \times \frac{\exp[i\kappa r - \omega t]}{-i\kappa r} \begin{pmatrix} 1 & 0 \\ 0 & \cos \Theta \end{pmatrix} \begin{pmatrix} E_{0r}(\mathcal{O}) \\ E_{0\ell}(\mathcal{O}) \end{pmatrix}, \end{aligned}$$

the vector \vec{D}_i is provided in Figure 3.5 and r_i in the denominator has been replaced by r . Summing the fields over all ΔV_i , and using the definition of the scattering amplitudes, we have

$$\begin{aligned} A_1(\Theta) &= \frac{-i3\kappa^3}{4\pi} \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \int \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] dV \\ A_2(\Theta) &= \frac{-i3\kappa^3}{4\pi} \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \cos \Theta \int \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] dV \end{aligned}$$

To evaluate the integral, we note that since $\vec{\kappa}$ and $\vec{\kappa}_0$ have the same magnitude, $\vec{\kappa}_0 - \vec{\kappa}$ is in a direction that bisects the angle between the two vectors (Figure 3.6). Volumes between infinitesimally spaced planes that are perpendicular to $\vec{\kappa}_0 - \vec{\kappa}$ all have the same value of $(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i \triangleq \delta$ (see Figure 3.7). Writing $dV = A_\delta(\eta) d\eta$, where $A_\delta(\eta)$ is the area of the surface of constant δ within the particle, we have

$$\int \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] dV = \int \exp[i\delta] A_\delta(\eta) d\eta, \quad (3.30)$$

with the integration covering the volume of the particle. For a homogeneous sphere of radius R centered on \mathcal{O} , $A_\delta(\eta) = \pi(R^2 - \eta^2)$, so

$$\int \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \bullet \vec{D}_i] dV = \int \exp[i\delta] \pi(R^2 - \eta^2) d\eta.$$

It is straightforward to relate δ to η as follows. From Figures 3.6 and 3.7

$$|\vec{\kappa}_0 - \vec{\kappa}| = 2\kappa \cos \left(\frac{\pi - \Theta}{2} \right) = 2\kappa \sin \left(\frac{\Theta}{2} \right).$$

¹⁸To derive this replace α in Footnote 10 of Chapter 1 by $3\epsilon_0 \Delta V \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right)$.

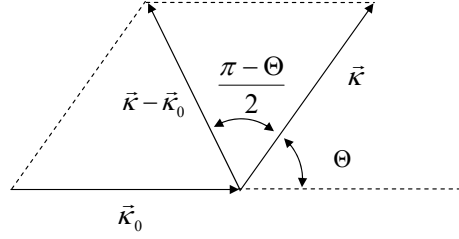


Figure 3.6: Relationship between $\vec{\kappa}_0$, $\vec{\kappa}$ and $\vec{\kappa} - \vec{\kappa}_0$. As usual, Θ specifies the scattering angle.

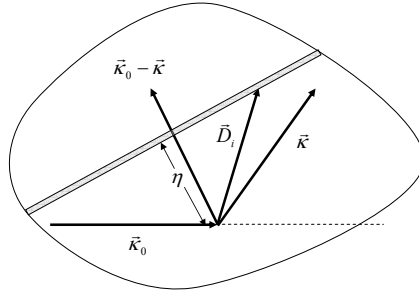


Figure 3.7: A slab of the particle is drawn normal to $\vec{\kappa}_0 - \vec{\kappa}$. Every point on the slab has the same value of δ .

But $\delta = (\vec{\kappa}_0 - \vec{\kappa}) \cdot \vec{D}_i = |\vec{\kappa}_0 - \vec{\kappa}|$ times the projection of \vec{D}_i on the direction of $\vec{\kappa}_0 - \vec{\kappa}$, which is just η . Thus

$$\delta = 2\kappa\eta \sin\left(\frac{\Theta}{2}\right),$$

and

$$\int \exp[i(\vec{\kappa}_0 - \vec{\kappa}) \cdot \vec{D}_i] dV = \int_{-R}^R \pi(R^2 - \eta^2) \exp\left[i2\kappa\eta \sin\left(\frac{\Theta}{2}\right)\right] d\eta = \frac{4\pi R^3}{w^3}(\sin w - w \cos w),$$

where $w = 2\kappa R \sin(\Theta/2)$. This provides the scattering amplitudes

$$\begin{aligned} A_1(\Theta) &= -i\kappa^3 R^3 \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \frac{3}{w^3} (\sin w - w \cos w) \\ A_2(\Theta) &= -i\kappa^3 R^3 \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \frac{3}{w^3} (\sin w - w \cos w) \cos \Theta, \end{aligned} \quad (3.31)$$

and the volume scattering function for unpolarized incident radiation is

$$\beta(\Theta) = n\kappa^4 R^6 \left| \frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right|^2 G(\Theta), \quad (3.32)$$

where

$$G(\Theta) = \left(\frac{3}{w^3} (\sin w - w \cos w) \right)^2 \left(\frac{1 + \cos^2 \Theta}{2} \right).$$

Figure 3.8 shows the variation of $G(\Theta)$ for $\kappa R = 0.1, 0.5, 1$ and 2 . Note that as κR increases, the volume scattering function becomes more concentrated in the forward direction. This is characteristic of scattering by particles that do *not* satisfy the condition $\kappa R \ll 1$. Figure 3.9 provides the difference between the Rayleigh-Gans computation of $\beta(\Theta)/\beta(0)$ and the exact values for $\tilde{m} = 1.01$ and $\kappa R = 2$. The Rayleigh-Gans approximation to this quantity is larger than the exact values. Thus, Rayleigh-Gans theory predicts more backscattering than would actually be observed. The increasing error at larger scattering angles is characteristic of this approximation. However, Figure 3.9 shows that for even relatively large particles ($R/\lambda = 0.318$), i.e., particle diameter greater than $\lambda/2$, the accuracy of the Rayleigh-Gans approximation is quite good for this low value of \tilde{m} .

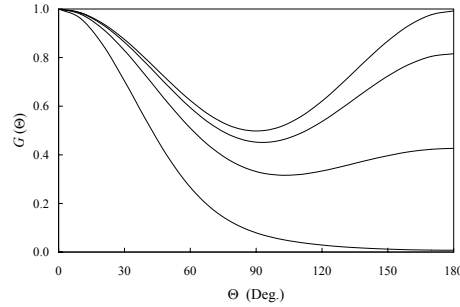


Figure 3.8: The function $G(\Theta)$ computed for several values of κ . The curves from top to bottom are for $\kappa R = 0.1, 0.5, 1$, and 2 , respectively. Note the increasing concentration of scattering in the forward direction as κR increases.

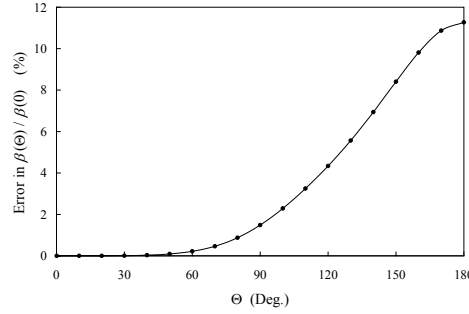


Figure 3.9: Percent error in the Rayleigh-Gans approximation to $\beta(\Theta)/\beta(0)$ in Figure 3.8, with $\kappa R = 2$ and $\tilde{m} = 1.01$. (Rayleigh-Gans is always $>$ exact computation.)

It is important to understand the full nature of the approximations used in obtaining Eqs. (3.31) and (3.32). The *only* approximation we made was to replace the total field \vec{E} in Eq. (3.28) by the incident field \vec{E}_{Inc} in Eq. (3.29). The stated requirement for this was $|\tilde{m} - 1| \ll 1$. However, when a plane wave propagates in a medium¹⁹ of index m_p , i.e., *inside* the particle, it follows

$$\vec{E}_p = \vec{E}_0 \exp[i(m_p \kappa z - \omega t)],$$

rather than

$$\vec{E}_m = \vec{E}_0 \exp[i(m_m \kappa z - \omega t)],$$

in a medium of index m_m . For the field inside the particle, \vec{E}_p , to approximate the field in the medium, \vec{E}_m , over the *whole particle*, we must require

$$\kappa R m_p \approx \kappa R m_w, \text{ or } |\kappa R(m_p - m_w)| \approx 0.$$

Another way to write this is, after division by m_w

$$\kappa R |\tilde{m} - 1| \ll 1.$$

Thus, we require $\kappa R |\tilde{m} - 1| \ll 1$ in *addition to* $|\tilde{m} - 1| \ll 1$.²⁰ It is possible to satisfy this even when the particle's size is comparable, or even larger, than the wavelength, e.g., if $\tilde{m} = 1.01$ for a sphere, the requirement is $R \ll 16\lambda$.

¹⁹The medium in this case being the material of which the particle is composed.

²⁰One can think of the requirement $|\tilde{m} - 1| \ll 1$ as ensuring that there is negligible refraction of the field at the medium-particle interface, and $\kappa R |\tilde{m} - 1| \ll 1$ ensuring that the difference in the phase of the field traversing the particle and the field outside the particle, is small.

Equation (3.32) shows that in the Rayleigh-Gans approximation $\sigma_b \propto R^6$, so $Q_b \propto R^4$ or x^4 . How does Q_a depend on R or x for particles in the Rayleigh-Gans regime? If we apply the optical theorem (Appendix 1, Eq. (3.57)) to the Rayleigh-Gans scattering amplitudes, we can compute the extinction cross section:²¹

$$\begin{aligned}\sigma_c &= \frac{4\pi}{\kappa^2} \Re \left(\frac{A_1(0) + A_2(0)}{2} \right) \\ &= \frac{4\pi}{\kappa^2} \Re \left[-i\kappa^3 R^3 \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \right] \\ &= \frac{4\pi}{\kappa^2} \kappa^3 R^3 \Im \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right).\end{aligned}$$

So,

$$Q_c = 4x \Im \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right).$$

By direct integration of Eq. (3.32), we find that

$$Q_b = x^4 \left| \frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right|^2 \quad \text{and} \quad \sigma_b = \pi \kappa^4 R^6 \left| \frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right|^2,$$

which is $\ll Q_c$. Thus, as x becomes $\ll 1$, $Q_a \rightarrow Q_c$, i.e.,

$$Q_a = 4x \Im \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right) \quad \text{and} \quad \sigma_a = 4\pi \kappa R^3 \Im \left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right).$$

Thus, we see that in the Rayleigh-Gans approximation, σ_a (and a) is proportional to the volume of the particle, while σ_b (and b) is proportional to the *square* of the volume of the particle.²²

An important aspect of the Rayleigh-Gans approximation is that it can be applied to particles of any shape. One need only be able to evaluate the integral in Eq. (3.30). This can always be done numerically. In the case of particles with a large or small aspect ratio, e.g., a long thin cylinder or a thin disk, the requirement $|\tilde{m} - 1|\kappa R \ll 1$ applies only for R equal to the *smallest* dimension. So for a long thin cylinder, R is its radius, while for a thin disk, R is its thickness.

3.3.3.2 The van de Hulst approximation: $|\tilde{m} - 1| \ll 1$ but $\kappa R \gg 1$

Consider a spherical particle with refractive index $\tilde{m} = \tilde{m}_r + i\tilde{m}_i$ (relative to its surroundings) near unity, i.e., $\tilde{m} - 1 \ll 1$. Assume that the particle is large relative to the

²¹In order to evaluate the amplitudes at $\Theta = 0$, we need to use the fact that $\lim_{w \rightarrow 0} [3(\sin w + w \cos w)/w^3] = 1$. Also, for a complex number z , $\Re(-iz) = \Im(z)$.

²²If we expand the quantity $(\tilde{m}^2 - 1)/(\tilde{m}^2 + 2)$ (noting that $|\tilde{m} - 1| \ll 1$), we find $Q_a = 8xm_i/3$.

wavelength, so geometrical optics (ray optics) describes the propagation reasonably well. Because $\tilde{m} \approx 1$, we assume that the individual rays (thin pencils of light) are *not* refracted or reflected at the particle surface. Then as a pencil traverses the sphere, it will undergo a phase shift (and some attenuation if $\tilde{m}_i \neq 0$) relative to adjacent pencils.²³ Consider a plane wave propagating in the $+z$ direction, and a pencil traversing the sphere as shown in Figure 3.10. At $z = z_0$ the spatial part of the wave is²⁴

$$\vec{E}_0 \exp[i\kappa z_0],$$

while at $z = z_1$ the field is

$$\vec{E}_0 \exp[i(\kappa z_0 + 2\kappa R(1 - \cos \theta) + 2\kappa R\tilde{m} \cos \theta)] = \vec{E}_0 \exp[i(\kappa z_1 + 2\kappa R(\tilde{m} - 1) \cos \theta)].$$

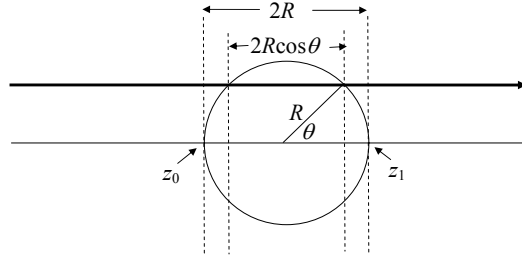


Figure 3.10: A pencil of rays traversing a sphere with $\tilde{m} - 1 \ll 1$.

Thus, if the detector is located at position z , the field there will be

$$\vec{E}_0 \exp[i(\kappa z + 2\kappa R(\tilde{m} - 1) \cos \theta)],$$

whereas in the absence of the particle, the field would be

$$\vec{E}_0 \exp[i\kappa z].$$

The scattered field at the screen \vec{E}_s is the difference between these two:

$$\begin{aligned} \vec{E}_s(x, y) &= \vec{E}_0 \exp[i(\kappa z + 2\kappa R(\tilde{m} - 1) \cos \theta)] - \vec{E}_0 \exp[i\kappa z] \\ &= -\vec{E}_0 \exp[i\kappa z] [1 - \exp[2i\kappa R(\tilde{m} - 1) \cos \theta]] \Delta(x, y), \end{aligned}$$

where $\Delta(x, y) = 1$ for $x \leq R$ and $y \leq R$, simultaneously, and zero otherwise.

²³Unlike the Rayleigh-Gans approximation, here the phase shift between the pencil traversing the particle and the pencil missing it, need not be small. In fact it can be arbitrarily large.

²⁴For this section all of the fields have a time factor $\exp(-i\omega t)$, which we leave out for convenience.

We now employ the arguments leading to the proof of the optical theorem in Appendix 1. There it is shown there that

$$\sigma_c = \frac{w_{\text{Ext}}}{|E_0|^2 / (2\mu_0 c)},$$

with

$$-w_{\text{Ext}} = \frac{1}{2\mu_0 c} \Re \left[\int_S \left[(\vec{E}_s \times \vec{B}_{\text{Inc}}^*) + (\vec{E}_{\text{Inc}} \times \vec{B}_s^*) \right] \cdot \hat{n} dA \right].$$

Here, S is the surface of a large imaginary volume surrounding the particle. The unit vector \hat{n} is the outward normal to the volume. In Appendix 1, this volume was taken to be spherical, but here, because of the geometry, we take it to be rectangular with sides parallel to the cartesian axes. Applying this formula to the present calculation, and using the expression $\vec{B} = \hat{k} \times \vec{E}/c$, valid for electromagnetic waves, we find

$$(\vec{E}_s \times \vec{B}_{\text{Inc}}^*) + (\vec{E}_{\text{Inc}} \times \vec{B}_s^*) = (\vec{E}_s \bullet \vec{E}_{\text{Inc}}^* + \vec{E}_{\text{Inc}} \bullet \vec{E}_s^*) \hat{e}_z.$$

Noting that

$$\begin{aligned} \vec{E}_s \bullet \vec{E}_{\text{Inc}}^* &= -|E_0|^2 [1 - \exp(i\rho \cos \theta)] \Delta(x, y) \\ \vec{E}_{\text{Inc}} \bullet \vec{E}_s^* &= -|E_0|^2 [1 - \exp(-i\rho^* \cos \theta)] \Delta(x, y). \end{aligned}$$

where $\rho = 2\kappa R(\tilde{m} - 1)$, w_{Ext} becomes

$$w_{\text{Ext}} = \frac{|E_0|^2}{2\mu_0 c} \Re \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[[1 - \exp(i\rho \cos \theta)] + [1 - \exp(-i\rho^* \cos \theta)] \right] \Delta(x, y) \bullet \hat{n} dx dy \right].$$

Let $x = \eta \cos \gamma$ and $y = \eta \sin \gamma$, then $dx dy = \eta d\eta d\gamma$, and note from Figure 3.10 that $\eta = R \cos \theta$. This results in

$$w_{\text{Ext}} = \frac{|E_0|^2}{2\mu_0 c} \Re \left[\int_0^{2\pi} \int_0^R \left[[1 - \exp(i\rho \cos \theta)] + [1 - \exp(-i\rho^* \cos \theta)] \right] R \sin \theta R d(\sin \theta) d\gamma \right],$$

but, since $\sin \theta d(\sin \theta) = -\cos \theta d(\cos \theta)$:

$$w_{\text{Ext}} = \frac{|E_0|^2}{2\mu_0 c} 2\pi R^2 \Re \left[\int_0^1 \left[[1 - \exp(i\rho u)] + [1 - \exp(-i\rho^* u)] \right] u du \right],$$

with $u = \cos \theta$. The integral

$$\int_0^1 [1 - \exp(i\rho u)] u du = \left[\frac{1}{2} - \frac{\exp(i\rho)}{i\rho} - \frac{1 - \exp(i\rho)}{(i\rho)^2} \right] \triangleq f(i\rho),$$

so

$$w_{\text{Ext}} = \frac{|E_0|^2}{2\mu_0 c} 2\pi R^2 \Re [f(i\rho) + f(-i\rho^*)],$$

yielding extinction cross section:

$$\sigma_c = 2\pi R^2 \Re [f(i\rho) + f(-i\rho^*)].$$

If $\tilde{m} = \tilde{m}_r + i\tilde{m}_i$ is real, i.e., there is no absorption so $\tilde{m}_i = 0$, then ρ is real, $\rho = \rho_r = 2\kappa R(\tilde{m}_r - 1)$. In this case taking the real part of the above expression is easy and

$$\sigma_c = 2\pi R^2 \left[1 - 2 \frac{\sin \rho_r}{\rho_r} + 2 \frac{(1 - \cos \rho_r)}{\rho_r^2} \right]. \quad (3.33)$$

This approximate cross section for a large spherical particle with relative index near unity is referred to as the *van de Hulst* approximation. If the particle is absorbing, $i\rho$ can be written

$$i\rho = 2i\kappa R[(\tilde{m}_r - 1) + i\tilde{m}_i] = i\rho_r \left[1 + i \frac{\tilde{m}_i}{\tilde{m}_r - 1} \right] \triangleq i\rho_r (1 + i \tan \gamma),$$

where $\tan \gamma = \tilde{m}_i/(\tilde{m}_r - 1)$. Then, after a considerable amount of algebra,

$$\begin{aligned} \sigma_c = 2\pi R^2 \left[1 - 2 \frac{\cos \gamma}{\rho_r} \sin(\rho_r - \gamma) \exp(-\rho_r \tan \gamma) \right. \\ \left. - 2 \frac{\cos^2 \gamma}{\rho_r^2} \cos(\rho_r - 2\gamma) \exp(-\rho_r \tan \gamma) + 2 \frac{\cos^2 \gamma}{\rho_r^2} \cos 2\gamma \right]. \end{aligned} \quad (3.34)$$

The absorption cross section for this model can be determined directly. Consider a beam of cross sectional area A_L and total power P_i . Divide it into pencils of area $dx dy$. Then the power in each pencil is

$$P_i \frac{dx dy}{A_L}$$

In traversing the particle, the field associated with each pencil is diminished by $\exp(-2\kappa R\tilde{m}_i \cos \theta)$, so the power reaching the detector is diminished to

$$dP_D = P_i \frac{dx dy}{A_L} \exp(-4\kappa R\tilde{m}_i \cos \theta),$$

and the loss in power received at the detector (due to the presence of the particle) is

$$dP_D - P_i \frac{dx dy}{A_L} = P_i \frac{dx dy}{A_L} [\exp(-4\kappa R\tilde{m}_i \cos \theta) - 1].$$

Integrating over area,

$$P_D - P_i = \frac{P_i}{A_L} \int \int [\exp(-4\kappa R\tilde{m}_i \cos \theta) - 1] dx dy,$$

where the integral over the right-hand-side is taken over the area of the particle. This integral is identical to the one encountered in the derivation of σ_c above, i.e.,

$$P_D - P_i = \frac{P_i}{A_L} 2\pi R^2 f(-\rho'),$$

where $\rho' \triangleq 4\kappa R \tilde{m}_i = 2\rho_r \tan \gamma$. Since the absorption coefficient is defined by

$$a \triangleq \frac{P_D - P_i}{P_i \ell},$$

we have

$$a = \frac{1}{A_L \ell} 2\pi R^2 f(-\rho')$$

for one particle in $A_L \ell$. For N particles in the sampling volume,

$$a = n 2\pi R^2 f(-\rho') \quad \text{and} \quad \sigma_a = 2\pi R^2 f(-\rho'),$$

where $n = N/A_L \ell$ and σ_a is the absorption cross section. Explicitly,

$$\sigma_a = \pi R^2 \left[1 + 2 \frac{\exp(-\rho')}{\rho'} - 2 \frac{[1 - \exp(-\rho')]}{\rho'^2} \right]. \quad (3.35)$$

Given σ_c and σ_a , the scattering cross section is

$$\sigma_b = \sigma_c - \sigma_a.$$

Figures 3.11 shows the extinction (Q_c), absorption (Q_a), and scattering (Q_b) efficiencies derived from Eqs. (3.34) and (3.35). The extinction efficiency Q_c displays an oscillatory structure that is more pronounced at low absorption. This phenomenon is due to the interference of the various pencils traversing the sphere. As the phase shift of the central ray through the particle is ρ_r , one would expect that the oscillations should be periodic in ρ_r with a period of 2π . Inspection of Figure 3.11 shows that this is indeed the case. The figures also show that the oscillations are increasingly damped as m_i increases. Figure 3.12, which compares the exact result with the van de Hulst approximation, shows the efficacy of the latter when \tilde{m} is near unity. It cannot, however, reproduce the small scale ripples in Q_c that become more apparent as \tilde{m} increases (Figure 3.3).

We note that as $\rho_r \rightarrow \infty$, $Q_c \rightarrow 2$, $Q_a \rightarrow 1$, and $Q_b \rightarrow 1$. Thus, a very large particle attenuates light with a cross section that is *twice* its geometrical cross section. This phenomenon is sometimes called the *extinction paradox*. The explanation is that all rays that actually strike the sphere will be reflected or refracted (and possibly absorbed), i.e., their directions will be changed, and these contribute πR^2 to σ_c . In addition, rays that pass close to the sphere will change direction slightly, i.e., will be *diffracted*, and this diffraction will contribute the other πR^2 to σ_c . We now examine diffraction.

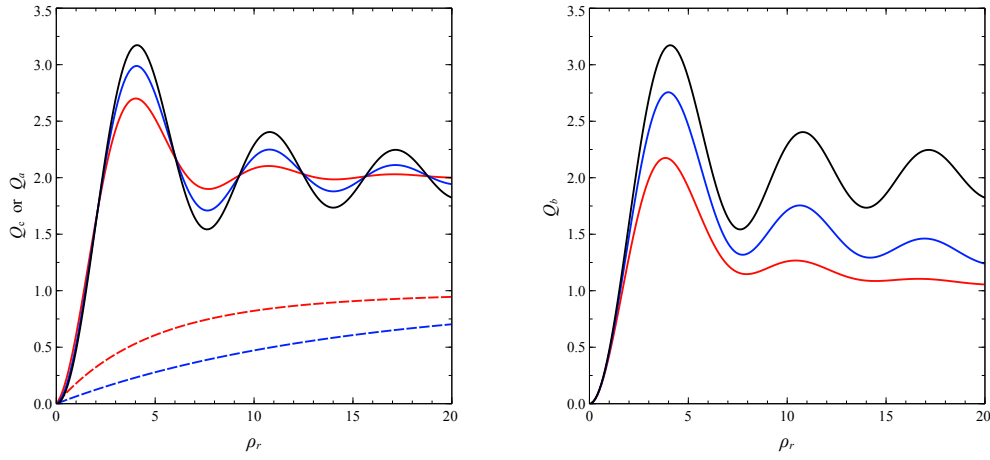


Figure 3.11: Left Panel: The attenuation and absorption efficiencies Q_c and Q_a as a function of $\rho_r = 2\kappa R(\tilde{m}_r - 1)$ computed using Eqs. (3.34) and Eq. (3.35), the van de Hulst approximation, for $m_i/(m_r - 1) = 0$ (black), 0.05 (blue), and 0.15 (red). Solid lines are Q_c and dashed lines are Q_a . Compare this to the exact computations in Figure (3.4) with identical parameters. Right Panel: The scattering efficiency Q_b as a function of $\rho_r = 2\kappa R(\tilde{m}_r - 1)$ computed using $Q_b = Q_c - Q_a$ and the left panel of this figure.

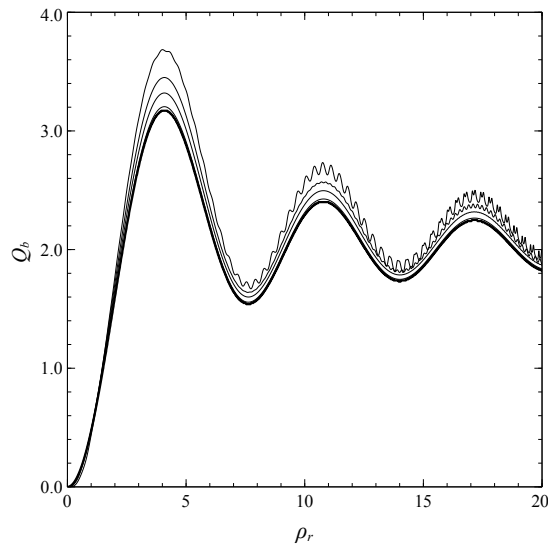


Figure 3.12: The scattering efficiency Q_b for *non-absorbing* spheres computed using the Mie equations (thin lines) from Figure 3.3 compared to the van de Hulst approximation (thick line and bottom curve) from Figure 3.11 (top curve) as a function of ρ_r . From top to bottom for the Mie calculations $\tilde{m}_r = 1.20, 1.10, 1.05$, and 1.01.

3.3.3.3 Diffraction

The Huygens-Fresnel principle states that every point on a wave front (surface of constant phase) is a secondary source of spherical waves, and that the interference of these waves describes the propagation of the wave front. This can be formulated to describe the phenomenon of *diffraction*. Consider a field propagating in the z direction, encountering an opaque (totally absorbing) screen containing a circular aperture (Figure 3.13). The field amplitude in the aperture will be essentially that of the incident field. The Huygens-Fresnel principle says that the field at the point (x, y, z) is²⁵

$$\vec{E}(x, y, z, t) = K \iint \vec{E}_0 \frac{\exp[i(\kappa s - \omega t)]}{s} dx' dy', \quad (3.36)$$

where the integral is over the aperture and K is a proportionality constant.²⁶ Taking the center of the aperture to be at $z' = 0$,

$$\begin{aligned} s &= \sqrt{(x - x')^2 + (y - y')^2 + z^2} \\ &= \sqrt{r + \frac{x'^2 + y'^2}{r^2} - 2\frac{xx' + yy'}{r^2}} \\ &\approx r + \frac{x'^2 + y'^2}{2r} - \frac{xx' + yy'}{r} \end{aligned}$$

From Figure 3.13, $x = r \sin \Theta \cos \phi$ and $y = r \sin \Theta \sin \phi$, so

$$\frac{xx' + yy'}{r} = (x' \cos \phi + y' \sin \phi) \sin \Theta.$$

We assume that the radius of the aperture $R \ll r$, so we can ignore the terms in x'^2 and y'^2 compared to x' and y' . In the integral, $1/s$ is slowly varying compared to $\exp(i\kappa s)$, so we can safely replace $1/s$ by $1/r$ and remove it from the integral. Then

$$\vec{E}(x, y, z, t) = K \frac{\exp[i(\kappa r - \omega t)]}{r} \iint \exp[-i\kappa(x' \cos \phi + y' \sin \phi) \sin \Theta] \vec{E}_0 dx' dy'.$$

²⁵Recall that the field of a spherical wave emanating from a point source (at $r = 0$), has the form $\exp[i(\kappa r - \omega t)]/r$.

²⁶The Huygens-Fresnel principle as formulated here is approximate. Clearly, contrary to Eq. (3.36) we do not expect any radiation backscattered from the aperture. Various attempts have been made to resolve this inconsistency by introducing so called obliquity factors; however, the basic formulation requires inconsistent assumptions. Kirchhoff diffraction theory provides a rigorous formulation, which can be found in most advanced texts on electromagnetic theory. Our interest here concerns the field propagating in directions that deviate only slightly from the direction of propagation of the incident field, and Eq. (3.36) is sufficient for this purpose. Note that if the screen were absent, the plane wave should simply propagate unchanged. To force Eq. (3.36) to yield this result requires that we set $K = -i/\lambda$.

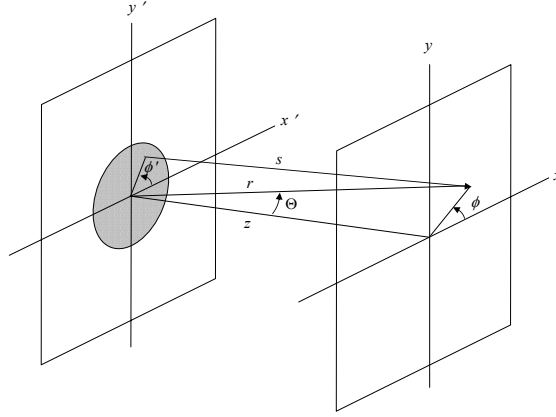


Figure 3.13: Geometry for studying diffraction. A plane wave propagates normal to an opaque screen located at $z' = 0$. A hole of radius R is cut in the screen (shaded area) through which the wave can propagate. The field is computed at a point (x, y) in the plane at z .

To carry out this integral, let $x' = \eta' \cos \phi'$ and $y' = \eta' \sin \phi'$ so $dx' dy' = \eta' d\eta' d\phi'$. Because of the symmetry, the amplitude over the screen should be independent of ϕ , so we choose $\phi = 0$, and

$$\vec{E}(x, y, z, t) = K \frac{\exp[i(\kappa r - \omega t)]}{r} \iint \exp[-i\kappa\eta' \cos \phi' \sin \Theta] \vec{E}_0 \eta' d\eta' d\phi'.$$

The ϕ' integral can be evaluated in terms of Bessel functions:²⁷

$$\begin{aligned} \vec{E}(x, y, z, t) &= 2\pi K \frac{\exp[i(\kappa r - \omega t)]}{r} \int_0^R \eta' J_0(\kappa\eta' \sin \Theta) \vec{E}_0 d\eta' \\ &= 2\pi K \frac{\exp[i(\kappa r - \omega t)]}{r} R^2 \frac{J_1(\kappa R \sin \Theta)}{\kappa R \sin \Theta} \vec{E}_0. \end{aligned}$$

The average Poynting vector (irradiance) on the x - y plane is

$$\langle S(x, y, z) \rangle = \frac{4\pi^2}{2\mu_0 c} \frac{K^2}{r^2} R^4 \left(\frac{J_1(\kappa R \sin \Theta)}{\kappa R \sin \Theta} \right)^2 E_0^2.$$

The function $J_1(x)/x$ is provided in Figure 3.14. Thus, radiant energy does not propagate

²⁷The appropriate relationships that we need are

$$J_n(x) = \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{\ell!(n+\ell)!} \left(\frac{x}{2}\right)^{n+2\ell}, \quad J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp[\pm ix \cos \alpha] d\alpha, \quad \text{and} \quad xJ_0(x) = \frac{d}{dx} [xJ_1(x)].$$

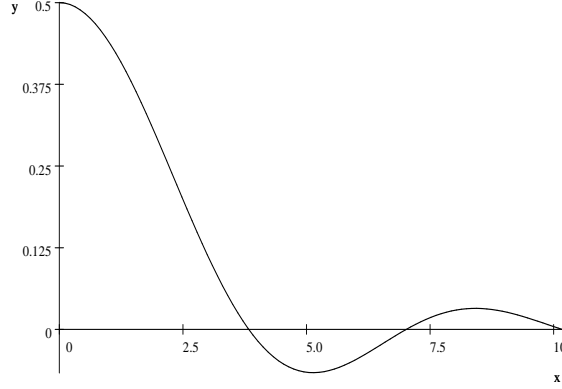


Figure 3.14: Plot of the Bessel function $J_1(x)/x$. Note the location of the first zero at $x \approx 3.8$.

straight through the aperture, but is *diffracted* into various angles Θ with respect to the incident beam direction. Note, however, that $[J_1(x)/x]^2$ becomes negligible for $x \gtrsim 10$, so for large κR the radiation is confined to small Θ . For example, for blue-green light ($\lambda \sim 0.5 \mu m$) and an aperture with $R = 5 \mu m$, most of the radiation would be contained within a cone of half angle $\Theta \approx 3.5^\circ$.

What if the opaque screen with the aperture is replaced by an opaque disk of radius R ? If $E_1(x, y, z, t)$ is the field when the screen and aperture are in place, and $E_2(x, y, z, t)$ is the field when the opaque disk is in place, then when both are in place simultaneously, the field is $E_1(x, y, z, t) + E_2(x, y, z, t)$. But if they are both in place, the field is zero, so $E_1(x, y, z, t) = -E_2(x, y, z, t)$, and the field diffracted (or scattered) by the disk is

$$\vec{E}_{\text{Disk}}(x, y, z, t) = -2\pi K \frac{\exp[i(\kappa r - \omega t)]}{r} R^2 \frac{J_1(\kappa R \sin \Theta)}{\kappa R \sin \Theta} \vec{E}_0.$$

This is called *Babinet's principle*.

It is simple to compute the scattering by a disk in the diffraction approximation. The incident field is $\vec{E}_{\text{Inc}} = \vec{E}_0 \exp[i(\kappa z - \omega t)]$, so

$$\vec{E}_{\text{Disk}}(x, y, z, t) = -\frac{\exp[i(\kappa r - \omega t)]}{i\kappa r} \kappa^2 R^2 \frac{J_1(\kappa R \sin \Theta)}{\kappa R \sin \Theta} \vec{E}_0,$$

where we have used the fact that $K = -i/\lambda$ (See Footnote 19). Noting that the scattered and incident fields are in the same direction, comparison with Eq. (3.4) shows that

$$A_1(\Theta) = A_2(\Theta) = \kappa^2 R^2 \frac{J_1(\kappa R \sin \Theta)}{\kappa R \sin \Theta},$$

yielding

$$\beta(\Theta) = n\kappa^2 R^4 \left[\frac{J_1(\kappa R \sin \Theta)}{\kappa R \sin \Theta} \right]^2$$

as the volume scattering function for a collection of n disks per unit volume. The optical theorem provides the extinction cross section

$$\sigma_c = \frac{\sigma_{cr} + \sigma_{cl}}{2} = \frac{4\pi}{\kappa^2} \Re \left(\frac{A_1(0) + A_2(0)}{2} \right) = 2\pi R^2,$$

where we have used the fact that $J_1(x)/x \rightarrow \frac{1}{2}$ as $x \rightarrow 0$. Clearly, the contribution to the extinction cross section resulting from rays that actually strike the disk is πR^2 , proving that diffraction also contributes πR^2 to the extinction cross section.

3.4 Scattering by Irregularly-Shaped Particles

So far, we have discussed only particles that are very small compared to the wavelength (Rayleigh scattering), tenuous particles that have sizes of at most a few wavelengths, but must have refractive index near unity (Rayleigh-Gans) and possessing simple shapes (spheres, cylinders, etc.), and spherical particles with no restriction on size (Mie theory). However, because of their size, shape, or refractive index, many particle species found nature do not fall within the range of applicability of these methods. Thus, it is of interest to discuss the scattering of light by irregularly-shaped particles arbitrary size and composition. We start by discussing a technique (the discrete-dipole approximation: DDA) that is capable of providing essentially exact computations of scattering for particles of any shape, as long as they are not too much larger than λ . Then, we discuss an approximate method (geometrical-optics/diffraction) of deriving scattering and absorption properties of such particles as long as the size is $\gg \lambda$. Of course, these techniques are also applicable to the scattering by spherical particles.

3.4.1 Discrete-Dipole Approximation (DDA)

The basic idea of the DDA is the following. Imagine the particle for which we want the scattered field to be subdivided into a large number of small subunits. Under the action of an incident plane wave, the i^{th} subunit will acquire a dipole moment $\vec{p}_i = \alpha_i \vec{E}_i$, where α_i is the subunit's polarizability (which we assume here for simplicity is a scalar) and \vec{E}_i is the field of the incident wave *plus* the field due to all of the other subunits.²⁸ Once all of the \vec{p}_i 's have been found, it is a simple matter to compute the field at a large distance from the particle. The key to finding the \vec{p}_i 's is finding \vec{E}_i at each subunit.

²⁸Note the essential difference between this and the Rayleigh-Gans approximation. Here, \vec{E}_i is the *total field* at each dipole, i.e., the incident field and the field due to *all* the other dipoles, while in the Rayleigh-Gans approximation it is just the incident field.

In Chapter 1 we found that for a dipole located at the origin of coordinates with moment given by $\vec{p}(t) = \vec{p}_0 \exp(-i\omega t)$, the electric field at a point specified by \vec{r} is

$$\vec{E}(r, t) = \frac{1}{4\pi\epsilon_0} \left\{ \frac{\kappa^2 \hat{r} \times (\hat{r} \times \vec{p}_0)}{r} + [\vec{p}_0 - 3\hat{r}(\hat{r} \cdot \vec{p}_0)] \left[\frac{i\kappa}{r^2} - \frac{1}{r^3} \right] \right\} \exp[i(\kappa r - \omega t)]. \quad (3.37)$$

Thus the subunit located at \vec{r}_j with dipole moment \vec{p}_j in Figure 3.15 will produce a field at P given by²⁹

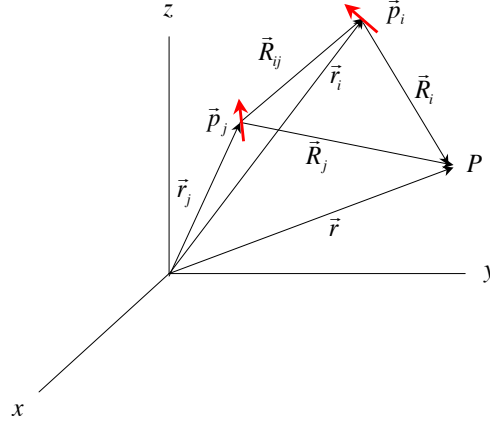


Figure 3.15: This shows two dipoles of moments \vec{p}_i and \vec{p}_j of subunits i and j , respectively. It is desired to find the field at point P when these two dipoles are induced by the field of each other and the field of a plane electromagnetic wave.

$$\vec{E}_j = \frac{1}{4\pi\epsilon_0} \left\{ \frac{\kappa^2 \hat{R}_j \times (\hat{R}_j \times \vec{p}_j)}{R_j} + [\vec{p}_j - 3\hat{R}_j(\hat{R}_j \cdot \vec{p}_j)] \left[\frac{i\kappa}{R_j^2} - \frac{1}{R_j^3} \right] \right\} \exp(i\kappa r), \quad (3.38)$$

and a field at subunit i located at \vec{r}_i given by

$$\vec{E}_{ij} = \frac{1}{4\pi\epsilon_0} \left\{ \frac{\kappa^2 \hat{R}_{ij} \times (\hat{R}_{ij} \times \vec{p}_j)}{R_{ij}} + [\vec{p}_j - 3\hat{R}_{ij}(\hat{R}_{ij} \cdot \vec{p}_j)] \left[\frac{i\kappa}{R_{ij}^2} - \frac{1}{R_{ij}^3} \right] \right\} \exp(i\kappa R_{ij}). \quad (3.39)$$

This latter field will contribute $\alpha_i \vec{E}_{ij}$ to the dipole moment \vec{p}_i of subunit i . It is convenient to rearrange the triple vector product

$$\hat{R} \times (\hat{R} \times \vec{p}) = \vec{p}(\hat{R} \cdot \hat{R}) - \hat{R}(\hat{R} \cdot \vec{p}) = \vec{p} - \hat{R}(\hat{R} \cdot \vec{p}),$$

²⁹All of the fields here will have the same time dependence, i.e., $\exp(-i\omega t)$, so the time factor will be suppressed in what follows.

which in dyadic form³⁰ is

$$\hat{R} \times (\hat{R} \times \vec{p}) = (\mathbf{I} - \hat{R}\hat{R}) \bullet \vec{p}.$$

Thus, noting that $\vec{p}_j = \alpha_j \vec{E}_j$, for a collection of N subunits we can write the electric field at the i^{th} subunit as

$$\begin{aligned} \vec{E}_i(\vec{r}_i) = & \frac{1}{4\pi\epsilon_0} \sum_{\substack{j=1 \\ j \neq i}}^N \left\{ \frac{\kappa^2 (\mathbf{I} - \hat{R}_{ij}\hat{R}_{ij}) \bullet \alpha_j \vec{E}_j}{R_{ij}} \right. \\ & + [\alpha_j \vec{E}_j - 3\hat{R}_{ij}(\hat{R}_{ij} \bullet \alpha_j \vec{E}_j)] \left[\frac{i\kappa}{R_{ij}^2} - \frac{1}{R_{ij}^3} \right] \Bigg\} \exp(i\kappa R_{ij}) \\ & + \vec{E}_0 \exp(i\kappa \bullet \vec{r}_i), \end{aligned} \quad (3.40)$$

where the sum is over all of the subunits except the i^{th} . Note that the right-hand side of Eq. (3.40) is linear in \vec{E}_j , so we can write this in the form

$$\vec{E}_i = \sum_{\substack{j=1 \\ j \neq i}}^N C_{ij} \vec{E}_j + \vec{E}_0 \exp(i\kappa \bullet \vec{r}_i), \quad (3.41)$$

where the $3N \times 3N$ matrix C_{ij} is easily derived from Eq. (3.40). These are a set of $3N$ linear equations, which when solved yield E_j for all j , i.e., the total field at each subunit, from which we can find the induced dipole moment $\vec{p}_j = \alpha_j \vec{E}_j$ for each subunit.³¹ The scattered field (Eq. (3.37)) is

$$\vec{E}_s(\vec{r}) = \frac{\kappa^2}{4\pi\epsilon_0} \sum_{j=1}^N (\mathbf{I} - \hat{R}_j\hat{R}_j) \bullet \vec{p}_j \frac{\exp(i\kappa R_j)}{R_j},$$

and since $\vec{r} = \vec{R}_j + \vec{r}_j$ with $R_j \gg r_j$, we can approximate R_j by $R_j \approx r(1 - \vec{r}_j \bullet \hat{r})$, and \hat{R}_j by \hat{r} so

$$\vec{E}_s(\vec{r}) = \frac{\kappa^2}{4\pi\epsilon_0} \frac{\exp(i\kappa r)}{r} \sum_{j=1}^N (\mathbf{I} - \hat{r}\hat{r}) \bullet \vec{p}_j \exp(-i\kappa \vec{r}_j \bullet \hat{r}). \quad (3.42)$$

As usual the magnetic field in the radiation zone is

$$\vec{B}_s(\vec{r}) = \frac{\vec{\kappa} \times \vec{E}_s(\vec{r})}{c}.$$

³⁰The scalar product between the dyad $\vec{A}\vec{B}$ and the vector \vec{C} is defined by $\vec{A}\vec{B} \bullet \vec{C} = \vec{A}(\vec{B} \bullet \vec{C})$ or $\vec{C} \bullet \vec{A}\vec{B} = (\vec{C} \bullet \vec{A})\vec{B}$, where the quantity in the parenthesis is evaluated first. The unit dyad \mathbf{I} has the property that $\mathbf{I} \bullet \vec{C} = \vec{C} \bullet \mathbf{I} = \vec{C}$.

³¹If we ignore the term with C_{ij} on the right hand side of Eq. (3.41), the result, $E_i = \vec{E}_0 \exp(i\kappa \bullet \vec{r}_i)$, is just the Rayleigh-Gans approximation described earlier.

Since we do not have the scattered field in terms of the incident field, the techniques for finding the various scattering cross sections used earlier are not available here. From Chapter 2, the volume scattering function is defined by

$$\beta(\hat{r}) = \frac{\Delta P(\hat{r})/\Delta\Omega}{e_0\Delta V},$$

where $\Delta P(\hat{r})$ is the radiant power scattered into a solid angle $\Delta\Omega$ around the scattering direction \hat{r} , e_0 is the incident irradiance on a plane normal to the incident beam,³² and ΔV is the (small) volume from which the scattering originates (the intersection of the incident beam and the field of view of the sensor detecting the scattered light). Noting that $\Delta\Omega = \Delta A/r^2$, where ΔA is the area of the detector (oriented normal to the scattered light, i.e., with its surface normal in the direction \hat{r}), we have

$$\frac{\Delta P}{\Delta\Omega} = r^2 \frac{\Delta P}{\Delta A} = r^2 \hat{r} \bullet \langle \vec{S}_s(\vec{r}) \rangle_{\text{Avg}},$$

where $\langle \vec{S}_s(\vec{r}) \rangle_{\text{Avg}}$ is the time-averaged Poynting vector of the scattered light at \vec{r} . The incident irradiance is $e_0 = \hat{\kappa}_0 \bullet \langle \vec{S}_{\text{Inc}} \rangle_{\text{Avg}}$, where $\hat{\kappa}_0$ is the direction of propagation of the incident beam, and $\langle \vec{S}_{\text{Inc}} \rangle_{\text{Avg}}$ is the time-averaged Poynting vector of the incident beam. Therefore,

$$\beta = r^2 \frac{\hat{r} \bullet \langle \vec{S}_s(\vec{r}) \rangle_{\text{Avg}}}{\Delta V \hat{\kappa}_0 \bullet \langle \vec{S}_{\text{Inc}} \rangle_{\text{Avg}}}.$$

This is the volume scattering function for a single particle. The quantity multiplying $1/\Delta V$ is called the differential scattering cross section $d\sigma_b/d\Omega$:³³

$$\frac{d\sigma_b}{d\Omega} = r^2 \frac{\hat{r} \bullet \langle \vec{S}_s(\vec{r}) \rangle_{\text{Avg}}}{\hat{\kappa}_0 \bullet \langle \vec{S}_{\text{Inc}} \rangle_{\text{Avg}}}.$$

Now, for an electromagnetic wave, the time-averaged Poynting vector is given by (Chapter 1)

$$\langle \vec{S} \rangle_{\text{Avg}} = \frac{\hat{\kappa}}{2\mu_0 c} \Re(\vec{E} \bullet \vec{E}^*),$$

so finally,

$$\frac{d\sigma_b(\hat{r})}{d\Omega} = \left(\frac{\kappa^2}{4\pi\epsilon_0} \right)^2 \frac{\left| \sum_{j=1}^N (\mathbf{I} - \hat{r}\hat{r}) \bullet \vec{p}_j \exp(-i\kappa\vec{r}_j \bullet \hat{r}) \right|^2}{|E_0|^2}. \quad (3.43)$$

³²Here we use e_0 , rather than E_0 as the latter is used for the amplitude of the incident electric field.

³³Note that if we have N particles in ΔV then

$$\beta = \frac{N}{\Delta V} \frac{d\sigma_b}{d\Omega} = n \frac{d\sigma_b}{d\Omega}, \quad \text{and with} \quad \sigma_b = \int_{\text{All } \Omega} \frac{d\sigma_b}{d\Omega} d\Omega,$$

we retrieve the relationship $b = n\sigma_b$.

The total extinction cross section can be found using the second form of the optical theorem derived in Appendix 1:

$$\sigma_c = \frac{4\pi}{\kappa} \Im[\vec{E}_0^* \bullet \vec{G}(\hat{e}_z)].$$

In our case here, $\vec{G}(\hat{r})$ is given by

$$\vec{G}(\hat{r}) = \frac{\kappa^2}{4\pi\epsilon_0} \sum_{j=1}^N (\mathbf{I} - \hat{r}\hat{r}) \bullet \vec{p}_j \exp(-i\kappa\vec{r}_j \bullet \hat{r}).$$

Noting that for $\hat{r} = \hat{e}_z$, $\vec{r}_j \bullet \hat{r} = z_j$, the position of the j^{th} subunit along the direction of the incident beam, and that $\vec{E}_0 \bullet \hat{e}_z = 0$,

$$\begin{aligned} \sigma_c &= \frac{4\pi}{\kappa} \Im \left[\frac{\kappa^2}{4\pi\epsilon_0} \exp(-i\kappa z_i) \vec{E}_0^* \bullet \vec{p}_j \right] \\ &= \frac{\kappa}{\epsilon_0} \Im[\vec{E}_{\text{Inc}}^*(\vec{r}_j) \bullet \vec{p}_j], \end{aligned} \quad (3.44)$$

where $\vec{E}_{\text{Inc}}(\vec{r}_j) = \vec{E}_0 \exp(i\kappa z_j)$ is the incident field at subunit j .

Generally, the subunits are chosen to all be of the same size and shape (and so all have the same polarizability) and are arranged in a cubic lattice. The lattice spacing (d – the center-to-center distance between adjacent subunits) must be a small fraction of the wavelength, i.e., $\kappa d \ll 1$. In practice the required value of κd is determined by trial and error. The number of subunits can become very large. Consider cubic particle that is $1 \mu\text{m}$ on a side and let $\lambda = 0.5 \mu\text{m}$. If the subunits are cubes $0.05 \mu\text{m}$ on a side ($\kappa d = 0.628$, not really $\ll 1$), then there are $1/0.05 = 20$ subunits along each edge of the cube, and a total of $N = 20^3 = 8,000$ subunits, which means that Eq. (3.41) is a system of 24,000 simultaneous equations that must be solved to find \vec{E}_j . Furthermore, if one desires to average the cross sections over particle orientation, it is usually accomplished by brute force: computing the associated quantities for each of a large number of orientations and averaging. Clearly, there are an enormous number of computations required to find the cross sections for even a moderately-sized particle.

This brief sketch of the DDA method does not address its full complexity; however, it does provide the basic principles on which it rests. Fortunately, computer codes (freely available) exist for carrying out DDA computations with as many as 10^6 or more subunits.

3.4.2 Geometrical-Optics/Diffraction Approximation

In the limit of the particle size $\gg \lambda$, photons incident on the particle are governed by the laws of geometrical optics (Snell's law and the Fresnel equations). That is, photons travel

in straight lines within a particle of uniform refractive index and refract and reflect from boundaries where the refractive index undergoes a discontinuous change. They also diffract around the boundary of the particle. If the particle is absorbing, some photons vanish as they propagate through the particle.

Consider a homogeneous irregularly-shaped particle in a particular orientation with respect to a beam of photons. Although it may be a trying exercise to follow the propagation of photons through an irregularly shaped particle, it is conceptually simple. When a photon strikes the particle surface the real part of the refractive index is used with Snell's law to determine the transmitted direction inside the particle. The incident and transmitted directions are used with the Fresnel equations to determine whether it is transmitted through, or reflected from, the surface. If reflected it contributes to the scattering at the appropriate scattering angle.³⁴ If transmitted the imaginary part of the refractive index is used to determine whether it is absorbed in the particle or propagates through the particle all the way to the next boundary.³⁵ Upon reaching the boundary, Snell's law and the Fresnel equations are again applied to determine the photon's fate.

To determine the scattering properties, a large imaginary sphere surrounding the particle is divided into sections of solid angle $\Delta\Omega = \sin\Theta\Delta\Theta\Delta\Phi$. The number of photons exiting in the solid angle specified by $(\Delta\Theta, \Delta\Phi)$ around the direction (Θ, Φ) , $\Delta N(\Theta, \Phi)$, is accumulated and used to compute the geometrical-optics scattering phase function:³⁶

$$P_{\text{Geom}}(\Theta, \Phi) = \frac{\Delta N(\Theta)/N}{\sin\Theta\Delta\Theta\Delta\Phi}. \quad (3.45)$$

This phase function is normalized to unity if there is no absorption, i.e.,

$$\int_0^{2\pi} d\Phi \int_0^\pi P_{\text{Geom}}(\Theta, \Phi) \sin\Theta d\Theta = 1,$$

as the ΔN 's must sum to N . In the presence of absorption, the value of this integral is just one minus the fraction of photons that were absorbed. Thus, as the phase function is usually normalized to unity (or sometimes 4π), it is more consistent to rewrite Eq. (3.45) as

$$\omega_{\text{Geom}} P_{\text{Geom}}(\Theta, \phi) = \frac{\Delta N(\Theta)/N}{\sin\Theta\Delta\Theta\Delta\Phi}, \quad (3.46)$$

³⁴This assumes that the particle is convex, i.e., it has no indentations. If it is not convex, then a photon could reflect from an external surface and encounter the external surface of the particle again. In this case the first-reflected photon does *not* contribute to the scattering.

³⁵Recall that the absorption coefficient of the material, a , is related to the imaginary part of the refractive index, m_i , through $a = 4\pi m_i/\lambda_v$, where λ_v is the *vacuum* wavelength of the radiation.

³⁶When the particle is absorbing, it is best to think of photons as having a "weight," which starts out at unity. As the photon propagates through the particle that weight decreases due to absorption, i.e., on propagating through a distance d in the particle the photon's weight decreases by a factor $\exp(-ad)$. What is accumulated as $\Delta N(\Theta, \Phi)$ is not the number of photons exiting, but the sum of their weights.

where ω_{Geom} is the single scattering albedo for the geometrical-optics portion of the scattering. Note that all photons that strike the particle will be either scattered or absorbed, i.e., leave the incident beam. Thus, the geometrical-optics extinction cross section is just the cross sectional area of the particle G , i.e., the area of the particle's shadow on a screen when illuminated by light normal to the screen:

$$\sigma_{c,\text{Geom}} = \sigma_{a,\text{Geom}} + \sigma_{b,\text{Geom}} = G,$$

and $\omega_{\text{Geom}} = \sigma_{b,\text{Geom}}/\sigma_{c,\text{Geom}}$.

The geometrical-optics approximation provides the scattering and absorption effects on photons that actually strike the particle. However, photons that pass close to the particle undergo diffraction, and thus, also contribute to the scattering. The angular distribution of scattering, i.e., the phase function, can be computed using the Huygens-Fresnel principle modified by Babinet's principle (see the discussion of diffraction earlier in this chapter).

As in the case of diffraction by a disk, the diffracted portion of the beam also contributes an amount G to the extinction cross section, i.e., $\sigma_{b,\text{Diff}} = G$. The easiest way to show this is to note that the diffracted fields of the particle and a screen with an aperture having the same shape as the geometric shadow of the particle, differ only by sign (Babinet's principle), so their Poynting vectors are identical. Since every photon that goes through an aperture will be diffracted (i.e., its path is altered), the scattering cross section of the aperture (and the particle) due to diffraction is simply G .

Combining the geometrical-optics contribution and the diffraction contributions, we have the following optical properties

$$\omega_0 P(\Theta, \Phi) = \frac{\omega_{\text{Geom}} P_{\text{Geom}}(\Theta, \Phi) + P_{\text{Diff}}(\Theta, \Phi)}{2} \quad (3.47)$$

$$\begin{aligned} \omega_0 &= \frac{\sigma_{b,\text{Geom}} + \sigma_{b,\text{Diff}}}{\sigma_{c,\text{Geom}} + \sigma_{c,\text{Diff}}} \\ &= \frac{\omega_{\text{Geom}} \sigma_{c,\text{Geom}} + \sigma_{c,\text{Diff}}}{\sigma_{c,\text{Geom}} + \sigma_{c,\text{Diff}}} \\ &= \frac{\omega_{\text{Geom}} G + G}{G + G} \\ &= \frac{\omega_{\text{Geom}} + 1}{2} \end{aligned} \quad (3.48)$$

To gain an appreciation of the potential accuracy of this method, we compare the geometrical-optics/diffraction solution to a problem that can be solved exactly: scattering by a large, non-absorbing, spherical particle. Figure 3.16 provides the geometry. A photon

is incident on the particle from the left and encounters the surface at an angle θ_i with respect to the surface normal. Snell's law is used to determine the refracted angle θ_t , and the Fresnel equations determine the reflected and transmitted Poynting vectors. The reflected part contributes to the scattering at an angle Θ_1 with respect to the incident direction. The transmitted fraction propagates through the particle and encounters the surface at an angle θ_t with respect to the surface normal. Snell's law and the Fresnel equations again determine the directions of the reflected and transmitted Poynting vectors. The transmitted portion contributes to the scattering at an angle Θ_2 with respect to the incident direction, etc. We will use the Monte Carlo approach to examine the fates of a large number of photons uniformly distributed over the geometrical cross sectional area ($G = \pi R^2$) of the particle, and determine the angular distribution of the scattering Eq. (3.45).

Because of the spherical symmetry, the process is simplified by the fact that the photon always remains in the plane formed by the incident direction and the center of the sphere. Note that the incident angle θ_i (Figure 3.16) is given by $\sin \theta_i = h_i/R$. Because a uniform distribution of photons is desired on the area G , we require that

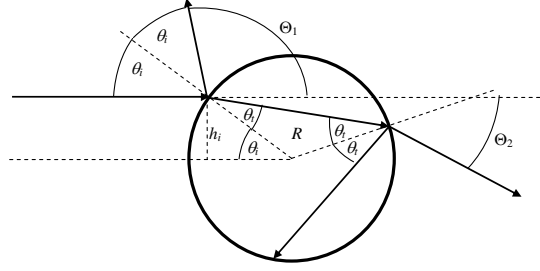


Figure 3.16: Schematic of some interactions of a photon incident on a spherical particle from the left and propagating according to the laws of geometrical optics.

$$\frac{N}{\pi R^2} = \frac{dN(h_i)}{2\pi h_i dh_i},$$

where N is the total number of photons incident, and $dN(h_i)$ is the number with $h_i - dh_i/2 \leq h_i \leq h_i + dh_i/2$. Thus,

$$\frac{dN(h_i)}{N} = 2 \frac{h_i dh_i}{R^2} = d\mathcal{P}(h_i) = p(h_i) dh_i,$$

where $d\mathcal{P}(h_i)$ is the probability that $h_i - dh_i/2 \leq h_i \leq h_i + dh_i/2$, and $p(h_i)$ is the associated probability density. Now, consider a random number ρ that is distributed uniformly on

the interval $0 \rightarrow 1$, i.e., $U[0, 1]$. Then, as in radiative transfer, we require

$$p(\rho) d\rho = d\rho = p(h_i) dh_i = 2 \frac{h_i dh_i}{R^2},$$

so

$$\rho_i = \int_0^{h_i} 2 \frac{h_i dh_i}{R^2} = \frac{h_i^2}{R^2} = \sin^2 \theta_i.$$

So given ρ_i , the incident angle is given by $\theta_i = \sin^{-1}(\sqrt{\rho_i})$. Once θ_i is determined, Snell's law determines θ_t and the photon's fate at the interface (reflection or transmission) is determined by a second random number ρ_f , distributed $U[0, 1]$, with the photon reflected if $\rho_f < r_f(\theta_i)$ and transmitted otherwise.³⁷ The photon continues until it exits the particle, at which point the scattering angle Θ is determined. Following Eq. (3.46), let the number of photons exiting in $\Delta\Theta$ be $\Delta N(\Theta)$, which is independent of Φ because of the spherical symmetry. This quantity is accumulated and used to compute the geometrical-optics scattering phase function ($\omega_{\text{Geom}} = 1$):

$$P_{\text{Geom}}(\Theta) = \frac{\Delta N(\Theta)/N}{2\pi \sin \Theta \Delta\Theta}.$$

The total scattering cross section $\sigma_{b,\text{Geom}}$ is πR^2 .

The phase function for diffraction can be computed by noting that the diffraction by a sphere is identical to that of a totally absorbing disk of the same R , oriented with its normal parallel to the incident beam. This disk diffraction was computed earlier in this chapter with the result that the phase function is proportional to $[xJ_1(x \sin \Theta)/x \sin \Theta]^2$, where $x = \kappa R$. If we normalize the diffraction phase function by requiring that³⁸

$$2\pi \int_0^{\pi/2} P_{\text{Diff}}(\Theta) \sin \Theta d\Theta = 1,$$

we find

$$P_{\text{Diff}}(\Theta) = \frac{x^2}{\pi} \left[\frac{J_1(x \sin \Theta)}{x \sin \Theta} \right]^2.$$

³⁷If we are interested in computing the complete Mueller matrix for scattering, we need to keep track of the polarization properties (the Stokes vector) of each photon. In that case, at this point we would have to use the Mueller matrix for reflection at an interface given in Chapter 1. If, as here, we are interested only in the phase *function* for an unpolarized incident beam, we use the Fresnel reflectance of the interface for unpolarized light:

$$r_f(\theta_i) = \frac{1}{2} \left[\frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} + \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} \right].$$

³⁸Normally the upper limit in the normalization integral should be π rather than $\pi/2$; however, in our derivation of diffraction using Huygens' principle, it is clear that *no* radiation can be diffracted into the backward hemisphere, so for consistency, the upper limit must be taken to exclude the backward hemisphere.

Since $\omega_{\text{Geom}} = 1$, the scattering phase function for a sphere by a homogeneous sphere in the limit $R \gg \lambda$ is (Eq. (3.47))

$$P(\Theta) = \frac{P_{\text{Diff}}(\Theta) + P_{\text{Geom}}(\Theta)}{2},$$

and $\sigma_b = \sigma_c = 2\pi R^2$.

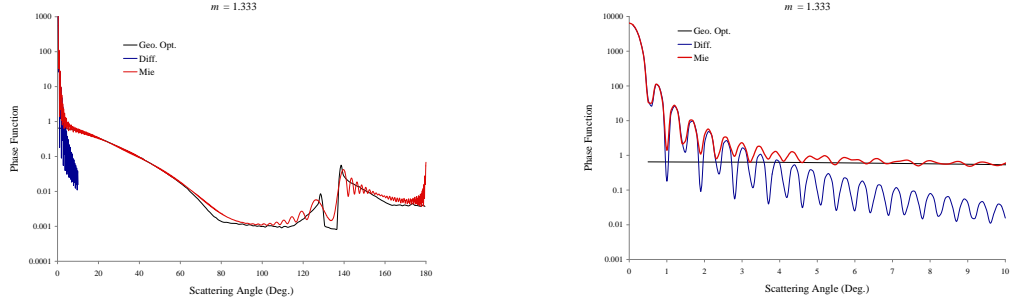


Figure 3.17: Left Panel: Comparison of the scattering phase functions computed using Mie theory (Red) and the partial phase functions in the geometrical optics and the diffraction approximations for $395 \leq \kappa R \leq 405$. The latter are $P_{\text{Geom}}(\Theta)/2$ (black) and $P_{\text{Diff}}(\Theta)/2$ (blue) for diffraction and geometrical-optics approximations respectively. The sum of the black and blue curves should coincide with the red curve if the approximations were exact. Right Panel: Expanded version of the left panel covering the first 10° in Θ .

The efficacy of the geometrical-optics/diffraction approximation in deriving the phase function for a non-absorbing sphere is demonstrated in Figure 3.17. This figure is for a sphere with $\kappa R = 400$ and an index of refraction 1.333.³⁹ The geometrical-optics component of the phase function is independent of the particle size. For the diffraction component and the exact Mie computation, the results have been averaged over a uniform distribution of particle sizes such that $395 \leq \kappa R \leq 405$ to reduce the magnitude of the phase function oscillations. The approximate computations follow the exact computations rather well, especially for $\Theta \lesssim 70^\circ$; however, the geometrical-optics/diffraction approximation is not capable of reproducing the fine scale structure of the exact phase function at larger angles, and fails totally in reproducing the exact results near $\Theta = 180^\circ$. The complicated structure in the geometrical-optics component with maxima near $\Theta = 129^\circ$ and 138° are

³⁹For $\lambda = 500$ nm, the mean value of R is $31.8 \mu\text{m}$, so the sphere is about $60 \mu\text{m}$ in diameter.

the secondary and primary *rainbows*. The broad minimum between the two is known as *Alexander's dark band*.

As κR increases, Mie theory phase function more closely resembles the geometrical-optics/diffraction phase function. Figure 3.18 (left panel) compares geometrical-optics/diffraction with the exact computations for $3950 \leq \kappa R \leq 4050$. The overall accuracy is improved over that in Figure 3.17, especially in the range $0-110^\circ$ and in Alexander's dark band, but neither the oscillations near the rainbow angles nor the scattering near 180° can be reproduced.⁴⁰ Conversely, for smaller particles, the geometrical-optics/diffraction approximation becomes increasingly poorer. For example, Figure 3.18 (right panel) shows the results for $39.5 \leq \kappa R \leq 40.5$ plotted in the same manner as Figure 3.17. Note that the diffraction contribution becomes exactly out of phase with the Mie result on the fourth maximum, and the rainbow structure near $130^\circ-140^\circ$ is totally absent from the Mie computations.

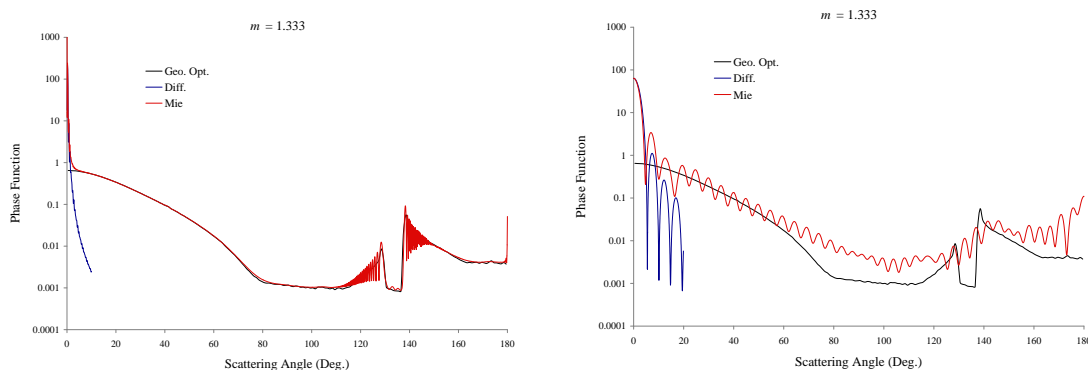


Figure 3.18: Same as Figure 3.17 (left panel) but with $3950 \leq \kappa R \leq 4050$ (left panel) and $39.5 \leq \kappa R \leq 40.5$ (right panel).

3.5 Summary

In this chapter we have summarized the basic theory of light scattering by matter. The results or relationships developed here will be used extensively in Chapters 4 and 5, which discuss the optical properties of the atmosphere and water bodies. We began with scattering by gases consisting of single atoms and/or anisotropic molecules. Then we examined

⁴⁰It should be noted that for a wider distribution of particle sizes the oscillations near the rainbow angles will be damped and the agreement would be much better; however, the disagreement near 180° would still remain.

the theory of scattering by a homogeneous sphere (Mie theory), and showed that the two relevant parameters are the size and the refractive index (which is determined by the material composition). Next we derived a series of approximations aimed at developing simple formulas or concepts that provide a more intuitive understanding than the complex formulas of Mie theory. In addition, a more complete examination of the physics underlying one of these approximations (Rayleigh-Gans) allowed development of the Discrete-Dipole approximation for solving scattering problems that are inaccessible by other means. Finally in appendices we derived the optical theorem relating the scattering amplitude in the forward direction to the extinction coefficient, provided a phenomenological discussion of the mechanism leading to Raman scattering and its relationship to Rayleigh scattering, and reviewed the salient aspects of electromagnetic scattering theory.

3.6 Appendix 1: The Optical Theorem

In this appendix we derive what is referred to as the *optical theorem*. It relates the extinction of a beam of radiation by a single particle to its scattering amplitude matrix

$$\mathbf{A}(\Theta) = \begin{pmatrix} A_1(\Theta) & A_4(\Theta) \\ A_3(\Theta) & A_2(\Theta) \end{pmatrix}$$

evaluated at $\Theta = 0$.

Consider a single particle interacting with a plane electromagnetic wave propagating in the $+z$ direction, i.e., $\hat{\kappa}_{\text{Inc}} = \hat{e}_z$. Let the particle be located at the origin of coordinates. The incident field at the particle is then $\vec{E}_{\text{Inc}} = \vec{E}_0 \exp[-i\omega t]$, and the scattered field the position \vec{r} is⁴¹

$$\vec{E}_s(\vec{r}) = -\frac{\exp[i(\kappa r - \omega t)]}{i\kappa r} \mathbf{A}(\Theta) \bullet \vec{E}_0 = -\frac{\exp[i(\kappa r - \omega t)]}{i\kappa r} \vec{F}(\Theta). \quad (3.49)$$

Thus, the total electric field at the point \vec{r} is

$$\vec{E}(\vec{r}) = \vec{E}_{\text{Inc}}(\vec{r}) + \vec{E}_s(\vec{r}) = \vec{E}_0 \exp[i(\kappa z - \omega t)] - \frac{\exp[i(\kappa r - \omega t)]}{i\kappa r} \vec{F}.$$

⁴¹The notation $\mathbf{A} \bullet \vec{E}_0$ is related to the matrix notation as follows,

$$\vec{F} = \mathbf{A} \bullet \vec{E}_0 = \begin{pmatrix} A_1(\Theta) & A_4(\Theta) \\ A_3(\Theta) & A_2(\Theta) \end{pmatrix} \begin{pmatrix} E_{0r} \\ E_{0\ell} \end{pmatrix} = \begin{pmatrix} A_1(\Theta)E_{0r} + A_4(\Theta)E_{0\ell} \\ A_3(\Theta)E_{0r} + A_2(\Theta)E_{0\ell} \end{pmatrix} = \begin{pmatrix} F_r \\ F_\ell \end{pmatrix}.$$

Also,

$$\vec{E}_i \bullet \vec{F} = \begin{pmatrix} E_{0r} & E_{0\ell} \end{pmatrix} \begin{pmatrix} F_r \\ F_\ell \end{pmatrix} \triangleq \vec{E}_0 \bullet \mathbf{A} \bullet \vec{E}_0 = \widetilde{\vec{E}_0} \mathbf{A} \vec{E}_0,$$

where the last equality is in the matrix notation used in the Mathematical Appendix.

The associated magnetic field is

$$\vec{B}(\vec{r}) = \frac{\hat{e}_z \times \vec{E}_{\text{Inc}}(\vec{r})}{c} + \frac{\hat{r} \times \vec{E}_s(\vec{r})}{c}.$$

Now consider a large sphere around the origin. We want to integrate the time-averaged Poynting vector of the total field over this sphere, that is, we want to compute

$$w = \int_{\text{Sphere}} \langle \vec{S}(\vec{r}) \rangle \bullet \hat{n} dA = \frac{1}{2\mu_0} \Re \left[\int_{\text{Sphere}} (E(\vec{r}) \times B^*(\vec{r})) \bullet \hat{n} dA \right].$$

But, first let's interpret this integral. From Poynting's theorem (Eq. (1.2)), the value of the integral is the net power *exiting* the sphere. Thus, if $w = 0$, the power entering the sphere equals that leaving, while if $w < 0$ more power enters than leaves, i.e., power is absorbed by the particle within the sphere. If we call the power absorbed within the sphere w_a , then

$$w_a = - \int_{\text{Sphere}} \langle \vec{S}(\vec{r}) \rangle \bullet \hat{n} dA = - \frac{1}{2\mu_0} \Re \left[\int_{\text{Sphere}} (\vec{E}(\vec{r}) \times \vec{B}^*(\vec{r})) \bullet \hat{n} dA \right].$$

To evaluate this we need to compute

$$\begin{aligned} \vec{S} &\triangleq (\vec{E}(\vec{r}) \times \vec{B}^*(\vec{r})) = (\vec{E}_{\text{Inc}}(\vec{r}) + \vec{E}_s(\vec{r})) \times (\vec{B}_{\text{Inc}}^*(\vec{r}) + \vec{B}_s^*(\vec{r})) \\ &= (\vec{E}_{\text{Inc}}(\vec{r}) \times \vec{B}_{\text{Inc}}^*(\vec{r})) + (\vec{E}_{\text{Inc}}(\vec{r}) \times \vec{B}_s^*(\vec{r})) + (\vec{E}_s(\vec{r}) \times \vec{B}_{\text{Inc}}^*(\vec{r})) + (\vec{E}_s(\vec{r}) \times \vec{B}_s^*(\vec{r})) \\ &\triangleq \vec{S}_{\text{Inc}} + \vec{S}_{\text{Inc},s} + \vec{S}_{s,\text{Inc}} + \vec{S}_s \end{aligned}$$

Now by direct calculation, or by noting that if there is no absorbing medium surrounding the particle,

$$\int_{\text{Sphere}} \vec{S}_{\text{Inc}} \bullet \hat{n} dA = 0.$$

Also,

$$\Re \int_{\text{Sphere}} \vec{S}_s \bullet \hat{n} dA = w_s,$$

the total power scattered by the particle and exiting the sphere. Writing

$$\Re \int_{\text{Sphere}} \vec{S}_{\text{Inc},s} \bullet \hat{n} dA + \Re \int_{\text{Sphere}} \vec{S}_{s,\text{Inc}} \bullet \hat{n} dA \triangleq -w_{\text{Ext}},$$

we see that

$$w_{\text{Ext}} = w_a + w_s,$$

i.e., w_{Ext} is the total power removed from the incident beam by absorption and by scattering due to the presence of the particle (the subscript “Ext” stands for *extinction*). We now go through the rather laborious process of calculating w_{Ext} .

First,

$$\begin{aligned} c\vec{\mathcal{S}}_{\text{Inc},s} &= \vec{E}_{\text{Inc}} \times (\hat{r} \times \vec{E}_s^*) = \hat{r}(\vec{E}_{\text{Inc}} \bullet \vec{E}_s^*) - \vec{E}_s^*(\hat{r} \bullet \vec{E}_{\text{Inc}}) \\ c\vec{\mathcal{S}}_{s,\text{Inc}} &= \vec{E}_s \times (\hat{e}_z \times \vec{E}_{\text{Inc}}^*) = \hat{e}_z(\vec{E}_{\text{Inc}}^* \bullet \vec{E}_s) - \vec{E}_{\text{Inc}}^*(\hat{e}_z \bullet \vec{E}_s). \end{aligned}$$

Then

$$(c\vec{\mathcal{S}}_{\text{Inc},s} + c\vec{\mathcal{S}}_{s,\text{Inc}}) \bullet \hat{r} \triangleq c\vec{\mathcal{S}}_{\text{Ext}} \bullet \hat{r} = (\vec{E}_{\text{Inc}} \bullet \vec{E}_s^*) + (\hat{r} \bullet \hat{e}_z)(\vec{E}_{\text{Inc}}^* \bullet \vec{E}_s) - (\hat{r} \bullet \vec{E}_{\text{Inc}}^*)(\hat{e}_z \bullet \vec{E}_s),$$

where we have used the fact that $\hat{r} \bullet \vec{E}_s^* = 0$. To proceed and calculate $2\mu_0 w_{\text{Ext}} = -\Re(\mathcal{I})$, where

$$\mathcal{I} = \int_{\text{Sphere}} \vec{\mathcal{S}}_{\text{Ext}} \bullet \hat{r} dA = \int_0^{2\pi} d\phi \int_0^\pi (\vec{\mathcal{S}}_{\text{Ext}} \bullet \hat{r}) r^2 \sin \theta d\theta, \quad (3.50)$$

and θ and ϕ are defined in Figure 1.7, we need to know \vec{E}_{Inc} . Letting $\vec{E}_{\text{Inc}} = \vec{E}_0 \exp[i(\kappa z - \omega t)]$, if we consider the case where $\vec{E}_0 = E_0 \hat{e}_x = E_0(\hat{e}_r \cos \phi + \hat{e}_\ell \sin \phi)$, and note that

$$\hat{r} = \hat{e}_x \cos \phi \sin \theta + \hat{e}_y \sin \phi \sin \theta + \hat{e}_z \cos \theta,$$

we have,

$$\begin{aligned} c\vec{\mathcal{S}}_{\text{Ext}} \bullet \hat{r} &= \exp[i(\kappa z - \omega t)](\vec{E}_0 \bullet \vec{E}_s^*) \\ &\quad + \exp[-i(\kappa z - \omega t)] \cos \theta (\vec{E}_0^* \bullet \vec{E}_s) \\ &\quad - \exp[-i(\kappa z - \omega t)] \cos \phi \sin \theta (\hat{e}_z \bullet \vec{E}_s) E_{\text{Inc}}^*. \end{aligned}$$

When this is inserted into Eq. (3.50), the last term will when integrated over ϕ will vanish,⁴² and so we will ignore the $(\hat{e}_z \bullet \vec{E}_s)$ from now on. Using Eq. (3.49), i.e.,

$$E_s(\vec{r}) = -\frac{\exp[i(\kappa r - \omega t)]}{i\kappa r} \vec{F}(\Theta),$$

where the scattering angle Θ is recognized as the polar angle θ ;

$$\begin{aligned} c\vec{\mathcal{S}}_{\text{Ext}} \bullet \hat{r} &= \frac{\exp[-i\kappa r]}{i\kappa r} \exp[i\kappa z] (\vec{E}_0 \bullet \vec{F}^*(\theta)) \\ &\quad - \frac{\exp[i\kappa r]}{i\kappa r} \exp[-i\kappa z] \cos \theta (\vec{E}_0^* \bullet \vec{F}(\theta)). \end{aligned} \quad (3.51)$$

Inserting this into Eq. (3.50), the ϕ integration is trivial, and we have to evaluate θ integrals of the form

$$\begin{aligned} c\mathcal{I} &= 2\pi \frac{\exp[-i\kappa r]}{i\kappa r} r^2 \int_0^\pi \exp[i\kappa z] (\vec{E}_0 \bullet \vec{F}^*(\theta)) \sin \theta d\theta \\ &\quad - 2\pi \frac{\exp[+i\kappa r]}{i\kappa r} r^2 \int_0^\pi \exp[-i\kappa z] (\vec{E}_0^* \bullet \vec{F}(\theta)) \cos \theta \sin \theta d\theta. \end{aligned}$$

⁴²The same can be said if $\vec{E}_{\text{Inc}} = \hat{e}_y E_0$, as the corresponding term in Eq. (3.50) will contain $\sin \phi$.

Now, note that $z = r \cos \theta$, and change the integration variable from θ to $u = \cos \theta$. The result is

$$\begin{aligned} c\mathcal{I} = & 2\pi \frac{\exp[-i\kappa r]}{i\kappa r} r^2 \int_{-1}^{+1} \exp[i\kappa r u] (\vec{E}_0 \bullet \vec{F}^*(u)) du \\ & - 2\pi \frac{\exp[+i\kappa r]}{i\kappa r} r^2 \int_{-1}^{+1} \exp[-i\kappa r u] (\vec{E}_0^* \bullet \vec{F}(u)) u du. \end{aligned} \quad (3.52)$$

These integrals are of the form

$$\int_{-1}^{+1} \exp[i\kappa r u] g(u) du,$$

and we want to evaluate them over a large sphere, i.e., as $r \rightarrow \infty$. If we integrate the above by parts, we get

$$\begin{aligned} \int_{-1}^{+1} \exp[i\kappa r u] g(u) du = & \frac{1}{i\kappa r} (\exp[i\kappa r] g(1) - \exp[-i\kappa r] g(-1)) \\ & - \frac{1}{i\kappa r} \int_{-1}^{+1} \exp[i\kappa r u] \frac{dg}{du} du. \end{aligned} \quad (3.53)$$

The magnitude of the second term in Eq. (3.53) can be estimated by replacing dg/du , by its maximum value, M , i.e.,

$$\left| \int_{-1}^{+1} \exp[i\kappa r u] \frac{dg}{du} du \right| \leq |M| \left| \int_{-1}^{+1} \exp[i\kappa r u] du \right| = |M| \left| \frac{1}{i\kappa r} \sin \kappa r \right|,$$

Thus, the first term in Eq. (3.53) is proportional to $1/r$ while the second is proportional to $1/r^2$, so as r becomes very large,

$$\int_{-1}^{+1} \exp[i\kappa r u] g(u) du \rightarrow \frac{1}{i\kappa r} (\exp[i\kappa r] g(1) - \exp[-i\kappa r] g(-1)).$$

Using this relationship, as $r \rightarrow \infty$

$$\begin{aligned} c\mathcal{I} \rightarrow & 2\pi \frac{1}{-\kappa^2} \left[(\vec{E}_0 \bullet \vec{F}^*(\theta = 0)) - \exp[-2i\kappa r] (\vec{E}_0 \bullet \vec{F}^*(\theta = \pi)) \right] \\ & - 2\pi \frac{1}{-\kappa^2} \left[(\vec{E}_0^* \bullet \vec{F}(\theta = 0)) + \exp[+2i\kappa r] (\vec{E}_0^* \bullet \vec{F}(\theta = \pi)) \right]. \end{aligned}$$

Recalling that for a complex number $z = a + ib$, $z + z^* = 2a = 2\Re(z)$ and $z - z^* = 2ib = 2i\Im(z)$, we see that in the above equation the terms without the exponential sum to a real number and those with the exponential sum to a complex number, i.e.,

$$c\mathcal{I} \rightarrow 2\pi \frac{1}{-\kappa^2} \left[2\Re(\vec{E}_0^* \bullet \vec{F}(\theta = 0)) + 2i\Im \left(\exp[+2i\kappa r] (\vec{E}_0^* \bullet \vec{F}(\theta = \pi)) \right) \right].$$

Now, $2\mu_0 w_{\text{Ext}} = -\Re(\mathcal{I})$, so

$$w_{\text{Ext}} = \frac{1}{2\mu_0 c} \frac{4\pi}{\kappa^2} \Re(\vec{E}_0^* \bullet \vec{F}(\theta=0)) = \frac{1}{2\mu_0 c} \frac{4\pi}{\kappa^2} \Re(\vec{E}_0^* \bullet \mathbf{A}(0) \bullet \vec{E}_0). \quad (3.54)$$

The last equation is one form of the *optical theorem*.⁴³ How is the optical theorem used? Consider a collimated beam propagating in the \hat{e}_z having cross sectional area \mathcal{A} and passing through a sample of thickness $d\ell$ (along the z direction) containing N identical particles that scatter independently of one another. The radiant power removed from the beam by absorption and scattering would be $Nw_{\text{Ext}} = n\mathcal{A}d\ell w_{\text{Ext}}$, where n is the particle number density. The incident radiant power is $|E_0|^2 \mathcal{A}/(2\mu_0 c)$, so the beam attenuation coefficient (or extinction coefficient) defined as

$$c \triangleq \frac{\text{Power Lost in } d\ell}{\text{Incident Power} \times d\ell}$$

is

$$c = \frac{n\mathcal{A}d\ell w_{\text{Ext}}}{(|E_0|^2/(2\mu_0 c))\mathcal{A} \times d\ell} = \frac{nw_{\text{Ext}}}{(|E_0|^2/(2\mu_0 c))} = n \frac{4\pi}{\kappa^2} \frac{\Re(\vec{E}_0^* \bullet \mathbf{A}(0) \bullet \vec{E}_0)}{|E_0|^2}, \quad (3.55)$$

and finally, the extinction cross section, $\sigma_c = c/n$, is

$$\sigma_c = \frac{4\pi}{\kappa^2} \frac{\Re(\vec{E}_0^* \bullet \mathbf{A}(0) \bullet \vec{E}_0)}{|E_0|^2}. \quad (3.56)$$

This is the more useful form of the theorem and is the one most often referred to as the “optical theorem.” It provides a useful way of deriving the extinction cross section of a single particle in terms of its scattering matrix, $\mathbf{A}(\Theta)$ evaluated at $\Theta = 0$, i.e., in the exact forward scattering direction.

We now provide some examples by computing σ_c for various states of polarization of the incident beam:

$$(a) \quad \vec{E}_0 = \hat{e}_r |\vec{E}_0|, \text{ so } (\vec{E}_0^* \bullet \mathbf{A}(0) \bullet \vec{E}_0) = A_1(0) |\vec{E}_0|^2, \text{ and } \sigma_{cr} = \frac{4\pi}{\kappa^2} \Re(A_1(0)),$$

⁴³An alternative method of carrying out the integrations required in Eq. (3.50) to derive the optical theorem is to use the fact that for large κr a plane wave can be represented by a combination of two spherical waves through

$$\exp(i\kappa z) = \exp(i\vec{\kappa} \bullet \vec{r}) \longrightarrow \frac{2\pi i}{\kappa} \left[\delta(\hat{r} + \hat{\kappa}) \frac{\exp(-i\kappa r)}{r} - \delta(\hat{r} - \hat{\kappa}) \frac{\exp(+i\kappa r)}{r} \right],$$

as $\kappa r \rightarrow \infty$. The presence of the δ -functions then make the integrals trivial. The method employed in the text was used to avoid this expansion, as we will not justify it here.

- (b) $\vec{E}_0 = \hat{e}_\ell |\vec{E}_0|$, so $(\vec{E}_0^* \bullet \mathbf{A}(0) \bullet \vec{E}_0) = A_2(0) |\vec{E}_0|^2$, and $\sigma_{c\ell} = \frac{4\pi}{\kappa^2} \Re(A_2(0))$,
- (c) unpolarized (“up”) incident radiation (equal uncorrelated mixtures of cases (a) and (b)),
- $$\sigma_{up} = \frac{4\pi}{\kappa^2} \Re\left(\frac{A_1(0) + A_2(0)}{2}\right), \quad (3.57)$$
- (d) RCP, $\vec{E}_0 = (\hat{e}_r - i\hat{e}_\ell) |\vec{E}_0| / \sqrt{2}$, so $(\vec{E}_0^* \bullet \mathbf{A}(0) \bullet \vec{E}_0) = \frac{1}{2}(A_1(0) + A_2(0) + iA_3(0) - iA_4(0)) |\vec{E}_0|^2$, and $\sigma_{RCP} = \frac{2\pi}{\kappa^2} \Re(A_1(0) + A_2(0) + iA_3(0) - iA_4(0))$.

Note, the Stokes vectors for the incident radiation in the four cases above are

$$(a) : \begin{pmatrix} I \\ -I \\ 0 \\ 0 \end{pmatrix} \quad (b) : \begin{pmatrix} I \\ I \\ 0 \\ 0 \end{pmatrix} \quad (c) : \begin{pmatrix} I \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (d) : \begin{pmatrix} I \\ 0 \\ 0 \\ I \end{pmatrix}.$$

Clearly, σ_c is a function of the state of polarization of the incident radiation. From the definition of the components of the Stokes vector (Eq. (1.54)), it is a straightforward exercise to show that for any incident vector \mathbf{I}_{Inc} ,

$$\begin{aligned} \sigma_c = \frac{2\pi}{\kappa^2} \frac{1}{I_{\text{Inc}}} \Re \Big[& (A_2(0) + A_1(0)) I_{\text{Inc}} + (A_2(0) - A_1(0)) Q_{\text{Inc}} \\ & + (A_3(0) + A_4(0)) U_{\text{Inc}} + i(A_3(0) - A_4(0)) V_{\text{Inc}} \Big], \end{aligned} \quad (3.58)$$

in agreement with the examples above.

Equation (3.58) provides the extinction of the total power in a beam with Stokes vector \mathbf{I}_{Inc} , i.e., the extinction of I_{Inc} , as a function of the *state of polarization* of the beam; however, it does not tell us how the *state of polarization of the beam* itself is modified by the particle, e.g., how is Q_{Inc} modified by the presence of the particle? To address this question, consider a collimated beam with Stokes vector \mathbf{I}_E . We want the 4×4 extinction matrix \mathbf{c} , such that upon traversing a slab of thickness $d\ell$ containing n identical particles per unit volume and all having the same orientation in space, the change in \mathbf{I}_E is given by

$$\frac{d\mathbf{I}_E}{d\ell} = -\mathbf{c}\mathbf{I}_E.$$

The extinction cross section is then $\sigma_c = \mathbf{c}/n$. We state without proof the resulting σ_c as a function of the scattering amplitudes evaluated at $\Theta = 0$:

$$\sigma_c = \frac{2\pi}{\kappa^2} \begin{pmatrix} +\Re[A_2(0) + A_1(0)] & +\Re[A_2(0) - A_1(0)] & +\Re[A_4(0) + A_3(0)] & +\Im[A_4(0) - A_3(0)] \\ +\Re[A_2(0) - A_1(0)] & +\Re[A_2(0) + A_1(0)] & -\Re[A_4(0) - A_3(0)] & -\Im[A_4(0) + A_3(0)] \\ +\Re[A_4(0) + A_3(0)] & +\Re[A_4(0) - A_3(0)] & -\Re[A_2(0) + A_1(0)] & +\Im[A_2(0) - A_1(0)] \\ +\Im[A_4(0) - A_3(0)] & +\Im[A_4(0) + A_3(0)] & -\Im[A_2(0) - A_1(0)] & +\Re[A_2(0) + A_1(0)] \end{pmatrix}.$$

From this we see that, in general, a collimated beam of unpolarized radiation will become polarized upon passage through a medium containing particles. For most of the particles we discussed in the text, e.g., spheres, $A_1(0) = A_2(0)$ and $A_3(\Theta) = A_4(\Theta) = 0$. In these cases, σ_c is diagonal, and in fact a multiple of the unit matrix, i.e.,

$$\sigma_c = \frac{2\pi}{\kappa^2} \Re[A_2(0) + A_1(0)] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

For cases such as this, all of the components of the Stokes vector are extinguished at the same rate, and the *state of polarization* of the beam is unaltered as it propagates in the medium.

An alternate form of the optical theorem is useful when we are not given the scattered field in terms of the incident field, i.e., we are not explicitly given $\mathbf{A}(\Theta)$. Suppose instead we are given the scattered field as

$$\vec{E}_s(\vec{r}) = \frac{\vec{G}(\hat{r}) \exp(ikr)}{r},$$

with $\vec{G}(\hat{r})$ describing the strength and the angular distribution of \vec{E}_s . This differs from Eq. (3.49) only by a factor $-i\kappa$. When this is taken into account Eq. (3.56) becomes

$$\sigma_c = \frac{4\pi}{\kappa} \Im \left[\vec{E}_0^* \bullet \vec{G}(\hat{e}_z) \right], \quad (3.59)$$

when the incident field is propagating in the \hat{e}_z direction.⁴⁴

3.7 Appendix 2: Raman Scattering

All of the processes that were discussed in text of this chapter fall under the heading *elastic* scattering, i.e., scattering without a change in wavelength. In this appendix we consider an *inelastic* scattering process in which the wavelength of the scattered photon differs from the wavelength of the incident photon.

In Chapter 1, we examined the induced dipole moment in atoms and molecules subjected to an electromagnetic wave. In particular we developed the properties of Rayleigh

⁴⁴This result is easy to derive:

$$\vec{F} = -i\kappa\vec{G}, \text{ so, } 2\Re(\vec{E}_0^* \bullet \vec{F}) = \vec{E}_0^* \bullet \vec{F} + \vec{E}_0 \bullet \vec{F}^* = -i\kappa(\vec{E}_0^* \bullet \vec{G} - \vec{E}_0 \bullet \vec{G}^*) = -i\kappa \, 2i\Im(\vec{E}_0^* \bullet \vec{G}) = 2\kappa\Im(\vec{E}_0^* \bullet \vec{G}).$$

scattering by atoms and molecules in terms of the polarizability tensor α that relates the induced dipole moment \vec{p} to the applied electric field \vec{E} :

$$\vec{p} = \alpha \vec{E}.$$

When \vec{E} is harmonic, the induced \vec{p} is also harmonic (at the same frequency), and the emitted radiation from the oscillating dipole is the (Rayleigh) scattered field. In the development, we assumed that α was a constant tensor when referenced to a coordinate system fixed with respect to the molecule. However, consider a diatomic molecule such as N_2 . The separation between the two nitrogen atoms varies harmonically in time (molecular vibration), and it is reasonable to expect that α will vary with time as well. Thus, in the molecular-based coordinate system (x , y , and z , with z along the internuclear axis)

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} \alpha_{\perp} & 0 & 0 \\ 0 & \alpha_{\perp} & 0 \\ 0 & 0 & \alpha_{\parallel} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}, \quad (3.60)$$

where we expect α_{\perp} and α_{\parallel} to each have an harmonic component of angular frequency ω_v , the vibrational frequency of the nuclei, i.e.,

$$\begin{aligned} \alpha_{\perp} &= \alpha_{\perp}^{(0)} + \alpha_{\perp}^{(1)} \cos \omega_v t, \\ \alpha_{\parallel} &= \alpha_{\parallel}^{(0)} + \alpha_{\parallel}^{(1)} \cos \omega_v t. \end{aligned}$$

Note that we have implicitly assumed that $\alpha^{(0)}$ and $\alpha^{(1)}$ have the same structure, i.e., both are diagonal and have two identical diagonal elements (xx and yy). This structure is forced by the symmetry of the diatomic molecule. Therefore, if \vec{E} is harmonic with a frequency ω_0 , we have for example

$$\begin{aligned} p_x(t) &= \left[\alpha_{\perp}^{(0)} + \alpha_{\perp}^{(1)} \cos \omega_v t \right] E_{0x} \cos \omega_0 t \\ &= \alpha_{\perp} E_{0x} \cos \omega_0 t + \alpha_{\perp}^{(1)} E_{0x} \cos \omega_v t \cos \omega_0 t \\ &= \alpha_{\perp} E_{0x} \cos \omega_0 t + \frac{\alpha_{\perp}^{(1)} E_{0x}}{2} [\cos(\omega_0 + \omega_v)t + \cos(\omega_0 - \omega_v)t], \end{aligned}$$

so p_x oscillates with frequencies ω_0 , $\omega_0 + \omega_v$ and $\omega_0 - \omega_v$, and will emit radiation (scatter) at these frequencies. The radiation emitted at ω_0 is the familiar Rayleigh scattering, while that emitted at $\omega_0 \pm \omega_v$ is called Raman scattering.

This treatment suggests that the scattering cross section and the polarization of Raman scattering should be similar to those of Rayleigh scattering (Eqs. (3.16) and (3.17)), but with the elements of $\alpha^{(1)}$ replacing those of $\alpha^{(0)}$:

$$b_{\text{Ram}}(\omega_0 \pm \omega_v) = \frac{8\pi}{3} \frac{n\kappa^4 \alpha'^2}{(4\pi\epsilon_0)^2} \left[\frac{6 + 3\delta_{\text{Ram}}}{6 - 7\delta_{\text{Ram}}} \right], \quad (3.61)$$

and

$$\beta_{\text{Ram}}(\Theta) = \frac{3}{16\pi} b_{\text{Ram}} \left[\frac{(1 + \delta_{\text{Ram}}) + (1 - \delta_{\text{Ram}}) \cos^2 \Theta}{1 + \delta_{\text{Ram}}/2} \right], \quad (3.62)$$

where

$$\delta_{\text{Ram}} = \frac{6\beta'^2}{45\alpha'^2 + 7\beta'^2},$$

with

$$\begin{aligned} \alpha' &\triangleq \frac{1}{3} \times \frac{1}{2} (\alpha_{\perp}^{(1)} + \alpha_{\perp}^{(1)} + \alpha_{\parallel}^{(1)}) \\ \beta'^2 &\triangleq \frac{1}{2} \times \frac{1}{4} [(\alpha_{\perp}^{(1)} - \alpha_{\perp}^{(1)})^2 + (\alpha_{\perp}^{(1)} - \alpha_{\parallel}^{(1)})^2 + (\alpha_{\parallel}^{(1)} - \alpha_{\perp}^{(1)})^2]. \end{aligned}$$

There is no simple relationship between $\alpha^{(0)}$ and $\alpha^{(1)}$, so there is no relationship between α and α' or β and β' , although generally $\alpha' \ll \alpha$.

The scattering at $\omega_0 - \omega_v$ is called the “Stokes line” and that at $\omega_0 + \omega_v$, the “anti-Stokes” line. The Stokes line occurs when the molecule in a given vibrational energy state absorbs a photon (at ω_0) increasing the molecule’s energy, and then emits a photon (at $\omega_0 - \omega_v$) as the molecule returns to a vibrational state of *higher* energy than the original. In contrast, the anti-Stokes line occurs when the molecule in a given vibrational energy state absorbs a photon (at ω_0) increasing the molecule’s energy, and then emits a photon (at $\omega_0 + \omega_v$) as the molecule returns to a vibrational state of *lower* energy than the original. The relative strengths of the Stokes and anti-Stokes lines then depend on the number density of atoms in the state that initially absorbs the photon, i.e., n in Eq. (3.61) is a function of $\omega_0 \pm \omega_v$, i.e., $n = n(\omega_0 \pm \omega_v)$. Since the number of molecules in high energy states is much lower than in lower states (and a function of the temperature), $n(\omega_0 - \omega_v) > n(\omega_0 + \omega_v)$, and the Stokes line is usually much more intense than the anti-Stokes line.

In more complex molecules, e.g., H_2O , there are more vibrational modes (three in H_2O) and therefore several values of ω_v . Also, the polarizability tensors in the molecular-fixed reference frame have no identical elements, and take the form

$$\boldsymbol{\alpha}^{(0)} = \begin{pmatrix} \alpha_x^{(0)} & 0 & 0 \\ 0 & \alpha_y^{(0)} & 0 \\ 0 & 0 & \alpha_z^{(0)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\alpha}^{(1)} = \begin{pmatrix} \alpha_x^{(1)} & 0 & 0 \\ 0 & \alpha_y^{(1)} & 0 \\ 0 & 0 & \alpha_z^{(1)} \end{pmatrix},$$

In this case Eqs. (3.61) and (3.62) still apply; however, α' and β' become

$$\begin{aligned} \alpha' &= \frac{1}{3} \times \frac{1}{2} (\alpha_x^{(1)} + \alpha_y^{(1)} + \alpha_z^{(1)}) \\ \beta'^2 &= \frac{1}{2} \times \frac{1}{4} [(\alpha_x^{(1)} - \alpha_y^{(1)})^2 + (\alpha_y^{(1)} - \alpha_z^{(1)})^2 + (\alpha_z^{(1)} - \alpha_x^{(1)})^2]. \end{aligned}$$

In a manner similar to Rayleigh scattering, the full phase matrix for Raman scattering is

$$\begin{aligned}
 \mathbf{P}_{\text{Ram}}(\Theta) &= \frac{4\pi\beta_{\text{Ram}}(\Theta)}{b_{\text{Ram}}} \\
 &= \frac{3}{2} \left[\frac{1 - \delta_{\text{Ram}}}{1 + \delta_{\text{Ram}}/2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2}\sin^2 \Theta & 0 & 0 \\ -\frac{1}{2}\sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \\
 &\quad + \frac{3}{2} \left[\frac{\delta_{\text{Ram}}}{1 + \delta_{\text{Ram}}/2} \right] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\cos \Theta \end{pmatrix}. \tag{3.63}
 \end{aligned}$$

3.8 Appendix 3: Useful Relationships for Rayleigh and Raman Scattering

The form of the Rayleigh and Raman volume scattering matrices are identical (Eqs. (3.19) and (3.63)). We show here that for this type of scattering there are simple relationships between b and the volume scattering functions derived with simple polarization states for incident and scattered light. These are useful in measurement of b and δ . Recall the definition of the volume scattering matrix (Chapter 2, Eq. (2.20)):

$$d\mathbf{I}_J(\hat{\xi}) = \beta_{\hat{\xi}_0 \rightarrow \hat{\xi}} \mathbf{I}_E(\hat{\xi}_0) dV,$$

where $d\mathbf{I}_J(\hat{\xi})$ is the scattered Stokes *intensity* vector in the direction $\hat{\xi}$ from a volume dV when $\mathbf{I}_E(\hat{\xi}_0)$ is the incident Stokes *irradiance* vector in the $\hat{\xi}_0$ direction. These Stokes vectors are defined with respect to the scattering plane (the plane containing $\hat{\xi}$ and $\hat{\xi}_0$).

Consider the case where the incident irradiance, E_{Inc} , is polarized perpendicular to the scattering plane. Its Stokes vector is

$$\mathbf{I}_E = \begin{pmatrix} +E_{\text{Inc}} \\ -E_{\text{Inc}} \\ 0 \\ 0 \end{pmatrix}.$$

We want to determine the polarization state of the scattered intensities. This can be effected with analyzing polarizers having pass directions perpendicular (r) and parallel (ℓ)

to the scattering plane. The Mueller matrices for these analyzers are (Chapter 2, Eq. (1.59))

$$\mathbf{M}_r = \frac{1}{2} \begin{pmatrix} +1 & -1 & 0 & 0 \\ -1 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{M}_\ell = \frac{1}{2} \begin{pmatrix} +1 & +1 & 0 & 0 \\ +1 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

respectively. The scattered Stokes intensity vector is

$$\begin{pmatrix} dI_J(\hat{\xi}) \\ dQ_J(\hat{\xi}) \\ dU_J(\hat{\xi}) \\ dV_J(\hat{\xi}) \end{pmatrix} = \mathbf{M}_{(r \text{ or } \ell)} \beta(\hat{\xi}_0 \rightarrow \hat{\xi}) \begin{pmatrix} +E_{\text{Inc}}(\hat{\xi}_0) \\ -E_{\text{Inc}}(\hat{\xi}_0) \\ 0 \\ 0 \end{pmatrix}.$$

When \mathbf{M}_r is used, we define $\beta_{r \rightarrow r}$ through,

$$\beta_{r \rightarrow r}(\hat{\xi}_0 \rightarrow \hat{\xi}) = \frac{(dI_J(\hat{\xi}))_r}{E_{\text{Inc}}(\hat{\xi}_0) dV}$$

and when \mathbf{M}_ℓ is used

$$\beta_{r \rightarrow \ell}(\hat{\xi}_0 \rightarrow \hat{\xi}) = \frac{(dI_J(\hat{\xi}))_\ell}{E_{\text{Inc}}(\hat{\xi}_0) dV}$$

Carrying out the various matrix multiplications with Eqs. (3.19) or (3.63) yields

$$\beta_{r \rightarrow r}(\Theta) = \frac{3b}{8\pi} \left(\frac{1 - \delta/2}{1 + \delta/2} \right), \quad \beta_{r \rightarrow \ell}(\Theta) = \frac{3b}{8\pi} \left(\frac{\delta/2}{1 + \delta/2} \right),$$

and

$$\frac{\beta_{r \rightarrow \ell}(\Theta)}{\beta_{r \rightarrow r}(\Theta)} = \frac{\delta/2}{1 - \delta/2} = \rho,$$

where ρ is the alternate definition of the depolarization ratio defined in Chapter 1.

These relationships are important because they facilitate the determination of b . First, to determine b , given ρ or δ , one need only carry out an experimental determination of $\beta_{r \rightarrow r}(\Theta)$ at *any* value of Θ , as $\beta_{r \rightarrow r}$ is independent of Θ . To determine ρ or δ one needs, in addition to $\beta_{r \rightarrow r}(\Theta)$, to measure $\beta_{r \rightarrow \ell}(\Theta)$ at any Θ , and not necessarily the same angle for $\beta_{r \rightarrow r}$ and $\beta_{r \rightarrow \ell}$, since neither depend on Θ . Typically both measurements are made at

$\Theta = 90^\circ$.⁴⁵

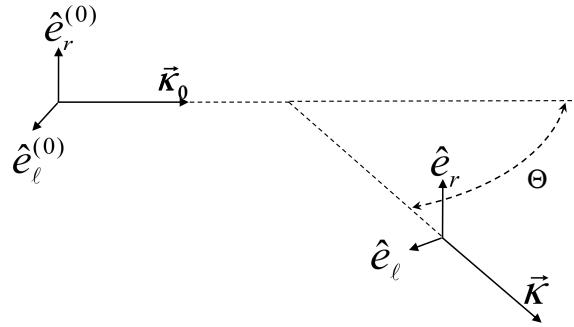
3.9 Appendix 4. Short Review of Basic Scattering Theory

Let the scattering center (particle) be located at $\vec{r} = 0$. The incident electric field propagating in the \vec{k}_0 direction is in the form of a plane wave given by (see figure below)

$$\vec{E}^{(0)}(\vec{r}, t) = \vec{E}^{(0)} \exp[i(\vec{k}_0 \bullet \vec{r} - \omega t)],$$

where the field amplitude $\vec{E}^{(0)}$ is resolved into components parallel and perpendicular to the scattering plane

$$\vec{E}^{(0)} = E_r^{(0)} \hat{e}_r^{(0)} + E_\ell^{(0)} \hat{e}_\ell^{(0)} = \begin{pmatrix} E_r^{(0)} \\ E_\ell^{(0)} \end{pmatrix}. \quad (3.64)$$



⁴⁵There are many other relationships that can be derived in this manner. Among them are

$$\begin{aligned} \beta_{r \rightarrow r}(\Theta) &= \text{Constant} \triangleq \beta_{r \rightarrow r} \\ \beta_{\ell \rightarrow r}(\Theta) &= \beta_{r \rightarrow \ell}(\Theta) = \text{Constant} \triangleq \beta_{\ell \rightarrow r} = \beta_{r \rightarrow \ell} \\ \beta_{\ell \rightarrow \ell}(\Theta) &= \beta_{r \rightarrow \ell} \sin^2 \Theta + \beta_{r \rightarrow r} \cos^2 \Theta \\ \beta_{n \rightarrow r} &= (\beta_{\ell \rightarrow r} + \beta_{r \rightarrow r}) = \beta_{r \rightarrow n} \\ \beta_{n \rightarrow \ell}(\Theta) &= (\beta_{r \rightarrow \ell} + \beta_{\ell \rightarrow \ell}(\Theta)) = \beta_{\ell \rightarrow n}(\Theta) \\ \beta_{n \rightarrow n}(\Theta) &= \frac{1}{2} (\beta_{r \rightarrow \ell} + \beta_{\ell \rightarrow \ell}(\Theta) + \beta_{\ell \rightarrow r} + \beta_{r \rightarrow r}) \end{aligned}$$

In these relationships, when the argument “ Θ ” is omitted, it means the corresponding β is independent of Θ . Also, the subscript “ n ” stands for “natural” and means that the incident or the scattered light is not passed through a polarizer. Thus $\beta_{n \rightarrow n}(\Theta)$ is the volume scattering function measured *without any polarizers* in the measurement path.

The scattered field at \vec{r} is in the form of a spherical wave and can be written

$$\vec{E}^{(s)}(\vec{r}, t) = \frac{1}{-i\kappa r} \mathbf{A} \vec{E}^{(0)} \exp[i(\kappa r - \omega t)],$$

where \mathbf{A} is the 2×2 (dimensionless) scattering amplitude matrix, and the field has also been resolved in to components parallel and perpendicular to the scattering plane (note that \hat{e}_ℓ and $\hat{e}_\ell^{(0)}$ are not parallel):

$$\vec{E}^{(s)} = E_r^{(s)} \hat{e}_r + E_\ell^{(s)} \hat{e}_\ell = \begin{pmatrix} E_r^{(s)} \\ E_\ell^{(s)} \end{pmatrix}. \quad (3.65)$$

To relate the scattered field to scattering cross sections, etc., we recall that the time averaged Poynting vector of the a field in the form of Eqs. (3.64) or (3.65) is

$$\langle \vec{S} \rangle = \frac{\hat{\kappa}}{2\mu_0 c_\ell} \langle E_r E_r^* + E_\ell E_\ell^* \rangle = \frac{dP}{dA},$$

where dP is the power crossing and area dA oriented normal to the propagation direction (i.e., $\langle \vec{S} \rangle$ is the irradiance associated with the propagating field).⁴⁶ The differential scattering cross section is defined to be the power scattered into a solid angle $d\Omega$ divided by the irradiance of the incident beam, i.e.,⁴⁷

$$\frac{d\sigma(\hat{\xi}_0 \rightarrow \hat{\xi})}{d\Omega(\hat{\xi})} = \frac{dP^{(s)}(\hat{\xi}) / d\Omega(\hat{\xi})}{dP^{(0)}(\hat{\xi}_0) / dA} = r^2 \frac{|\langle \vec{S}^{(s)}(\hat{\xi}) \rangle|}{|\langle \vec{S}^{(0)}(\hat{\xi}_0) \rangle|}$$

where the superscript “s” stands for “scattered,” the superscript “0” stands for “incident,” r is the distance from the scattering volume to the point of observation and $\hat{\xi}_0$ and $\hat{\xi}$ are in the directions of $\hat{\kappa}_0$ and $\hat{\kappa}$, respectively. The required Poynting vectors are given by

$$|\langle \vec{S}^{(0)}(\hat{\xi}_0) \rangle| = \frac{1}{2\mu_0 c_\ell} \langle \widetilde{\vec{E}^{(0)*}} \vec{E}^{(0)} \rangle \quad \text{and} \quad |\langle \vec{S}^{(s)}(\hat{\xi}) \rangle| = \frac{1}{2\mu_0 c_\ell} \frac{1}{\kappa^2 r^2} \langle \widetilde{\vec{E}^{(0)*}} \mathbf{A}^* \mathbf{A} \vec{E}^{(0)} \rangle$$

so

$$\frac{d\sigma}{d\Omega} = \frac{1}{\kappa^2} \frac{\langle \widetilde{\vec{E}^{(0)*}} \mathbf{A}^* \mathbf{A} \vec{E}^{(0)} \rangle}{\langle \widetilde{\vec{E}^{(0)*}} \vec{E}^{(0)} \rangle}$$

⁴⁶Here we assume that all quantities are restricted to the spectral band $\Delta\lambda$, which for compactness we will omit from the various quantities as they will all cancel from the final equations.

⁴⁷Although this definition appears different from that in Chapter 2, it is identical. Note $dP^{(s)}(\hat{\xi})/d\Omega$ is the intensity in the direction $\hat{\xi}$ and $dP^{(0)}/dA$ is the irradiance incident on the scattering volume, ΔV . If there is one particle in ΔV , i.e., the number density of particles in ΔV is $1/\Delta V$, then $\beta = nd\sigma/d\Omega$ indicates that $\beta(\hat{\xi}_0 \rightarrow \hat{\xi})$ equals the intensity in the direction $\hat{\xi}$ divided by the incident irradiance (direction $\hat{\xi}_0$) times ΔV . This is Eq. (2.19), the definition of β .

If the incident beam has a polarization specified by the Stokes vector \mathbf{I} , the differential cross section is a 4×4 matrix, i.e.,

$$\frac{d\boldsymbol{\sigma}}{d\Omega} = r^2 \mathbf{M},$$

where

$$\mathbf{I}^{(s)} = \mathbf{M} \mathbf{I}^{(0)},$$

with \mathbf{I} defined as usual:

$$\mathbf{I} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \frac{1}{2\mu_0 c_\ell} \begin{pmatrix} E_\ell^* E_\ell + E_r^* E_r \\ E_\ell^* E_\ell - \vec{E}_r^* E_r \\ E_r^* E_\ell + \vec{E}_\ell^* E_r \\ i(E_r^* E_\ell - \vec{E}_\ell^* E_r) \end{pmatrix}$$

The elements of \mathbf{M} are given by

$$M_{ij} = \frac{(-1)^{(i-j)}}{2\kappa^2 r^2} \text{Trace}(\boldsymbol{\sigma}_i \mathbf{A} \boldsymbol{\sigma}_j \widetilde{\mathbf{A}}^*).$$

The $\boldsymbol{\sigma}_i$'s in this equation should not be confused with $d\boldsymbol{\sigma}$. They are given by (Chapter 1),

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\sigma}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{\sigma}_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\sigma}_4 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

If n is the number density of (identical) particles in the scattering volume, then $\beta = n d\boldsymbol{\sigma}/d\Omega$. The extension to a distribution of particle sizes is immediate.

3.10 Bibliographic Notes

A considerable amount of the material in this Chapter is derived directly from the material presented in Chapters 1 and 2. The rest can be found in the books by [Mishchenko et al. \[2002a\]](#), [Bohren and Huffman \[1983\]](#) and [van de Hulst \[1957\]](#) or in the review paper by [Hansen and Travis \[1974\]](#). Here, we have chosen to expand on those portions of scattering theory that are of most interest in atmospheric and oceanic optics. Additional sources for portions of various sections are provided below.

3.2.3 Rayleigh Scattering by Liquids

The thermodynamic relationships used in this derivation were taken from [Morel \[1974\]](#).

3.3 Scattering by a Homogeneous Sphere

There are many freely available codes for computing scattering by a sphere. One is described in detail in [Mishchenko et al. \[2002a\]](#).

3.3.3.1 The Rayleigh-Gans Approximation

[Gordon \[2007a\]](#) provides examples showing that the Rayleigh-Gans approximation yields excellent results for disks with diameters much greater than λ as long as the thickness t satisfies $\kappa t|m-1| \ll 1$. In addition unpublished results show that the approximation works well for cylinders with length much greater than λ as long as the radius R satisfies $\kappa R|m-1| \ll 1$. These are the sources of the statement that the Rayleigh-Gans requirement $\kappa d|m-1| \ll 1$ need apply only to the *smallest* dimension d of the particle.

3.3.3.2 The van de Hulst Approximation

In addition to spheres, this approximation can be applied to particles of other morphologies. Some are provided by [Jonasz and Fournier \[2007\]](#) and references therein.

3.3.3.3 Diffraction

[Born and Wolf \[1975\]](#) cover the Huygens-Fresnel principle in considerable detail.

3.4 Scattering by Irregularly-Shaped Particles

Of additional methods for solving the scattering problem for non spherical particles to most famous is the T-Matrix method described in detail in [Mishchenko et al. \[2002a\]](#). Various other methods are outlined in [Mishchenko et al. \[2000\]](#).

3.4.1 Discrete-Dipole Approximation

The DDA was introduced by [Purcell and Pennypacker \[1973\]](#). The development here follows [Draine \[1988\]](#). A DDA code is available from B. Draine, Princeton University.

3.6 The Optical Theorem

Our basic derivation of the optical theorem follows that of [Bohren and Huffman \[1983\]](#). The σ_c Mueller matrix is derived in [Mishchenko et al. \[2002a\]](#). We took their matrix and converted it from their notation to ours. The expansion of a plane wave into two spherical waves (Footnote 44) was also taken from this source.

3.7 Raman Scattering

The classical explanation of the Raman effect as being due to a change in the polarizability tensor as a molecule vibrates is found in many texts. A good one is [Wilson et al. \[1980\]](#). For a full quantum mechanical treatment of Raman scattering, see [Placzek \[1962\]](#).

Part II

Optical Properties

Chapter 4

Inherent Optical Properties of the Atmosphere

4.1 Introduction

In this chapter we describe some in detail the inherent optical properties of the atmosphere. The processes of interest are scattering by the air molecules (Rayleigh scattering), scattering and absorption by aerosols (condensed liquid or solid particles suspended in the air) and absorption by the gaseous constituents of the atmosphere. The goal is to provide background required for later chapters regarding the influence of the atmosphere on the radiance observed from a space-borne or air-borne sensor viewing the water.

4.2 Molecular (Rayleigh) Scattering

The scattering properties of atoms and anisotropic molecules were developed in Chapters 1 and 3. For a gas with bulk refractive index m (which is real, as we will assume there are no absorption bands near the wavelength of interest), the scattering cross section *per molecule* is

$$\sigma_r = \frac{24\pi^3}{\lambda^4 n^2} \left(\frac{m^2 - 1}{m^2 + 1} \right)^2 \left[\frac{6 + 3\delta}{6 - 7\delta} \right],$$

where δ is the Rayleigh depolarization factor, and as

$$\frac{1}{n} \left(\frac{m^2 - 1}{m^2 + 2} \right) = \frac{\alpha}{3\epsilon_0}$$

is a property of a *single* atom, σ_r and does not depend on n . However, the scattering coefficient does depend on temperature and pressure. For an ideal gas the scattering coefficient is given by

$$b_r = n\sigma_r = \left(\frac{P}{k_B T} \right) \sigma_r,$$

where k_B is the Boltzmann constant. Finally, the phase matrix for molecular scattering is given by

$$\begin{aligned} \mathbf{P}_r(\Theta) &= \frac{4\pi\beta_r(\Theta)}{b_r} \\ &= \frac{3}{2} \left[\frac{1-\delta}{1+\delta/2} \right] \begin{pmatrix} \frac{1}{2}(1+\cos^2\Theta) & -\frac{1}{2}\sin^2\Theta & 0 & 0 \\ -\frac{1}{2}\sin^2\Theta & \frac{1}{2}(1+\cos^2\Theta) & 0 & 0 \\ 0 & 0 & \cos\Theta & 0 \\ 0 & 0 & 0 & \cos\Theta \end{pmatrix} \\ &\quad + \frac{3}{2} \left[\frac{\delta}{1+\delta/2} \right] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\cos\Theta \end{pmatrix}. \end{aligned}$$

4.2.1 Rayleigh Scattering by the Atmospheric Gas Mixture

As the atmosphere is a mixture of molecules of different species, we need to know how to combine the various gases to form “air.” The simplest method is to note that the scattering cross section of one molecule of species i is

$$(\sigma_r)_i = \frac{24\pi^3}{\lambda^4 n_i^2} \left(\frac{m_i^2 - 1}{m_i^2 + 1} \right)^2 \left[\frac{6 + 3\delta_i}{6 - 7\delta_i} \right]. \quad (4.1)$$

Since the quantity

$$\frac{1}{n_i^2} \left(\frac{m_i^2 - 1}{m_i^2 + 1} \right)^2$$

is independent of P , T , and n_i , we can replace it with values measured at standard conditions, i.e.,

$$\frac{1}{n_i^2} \left(\frac{m_i^2 - 1}{m_i^2 + 1} \right)^2 = \frac{1}{n_s^2} \left(\frac{m_i^2 - 1}{m_i^2 + 1} \right)_{\text{STP}}^2,$$

where STP refers to standard conditions: $T_s = 273.15$ °K, $P_s = 101,325$ Pa and $n_s = 2.687 \times 10^{19}$ molecules/cm³. The depolarization factor, however, δ_i is a property of the species only and can be measured at any P and T . Then, the scattering coefficient and the volume scattering matrix are

$$b_r = \sum_i (\sigma_r)_i n_i, \quad \beta(\Theta) = \sum_i \beta_i n_i.$$

Measurements of the refractive index of gases (m_g) show that for the principal components of the atmosphere, N₂, O₂, and Ar, the *refractivity* ($m_g - 1$) has the values 275.9×10^{-6} , 249.5×10^{-6} , and 260.6×10^{-6} , respectively at $T = 293.15$ °K, a pressure of 101,325 Pa and a wavelength of 633 nm. In contrast, for CO₂, $m_g - 1 = 413.4 \times 10^{-6}$ under the same conditions. Because of the similarity of m_g for N₂, O₂, and Ar, one usually replaces their values with the value given for air: $m_g - 1 = 270.3 \times 10^{-6}$.¹ When this is effected,

$$(\sigma_r)_i = \frac{24\pi^3}{\lambda^4 n_{\text{air}}^2} \left(\frac{m_{\text{air}}^2 - 1}{m_{\text{air}}^2 + 1} \right)^2 \left[\frac{6 + 3\delta_i}{6 - 7\delta_i} \right] = \frac{24\pi^3}{\lambda^4 n_{\text{air}}^2} \left(\frac{m_{\text{air}}^2 - 1}{m_{\text{air}}^2 + 1} \right)^2 F_i(\lambda),$$

where F_i is called the King factor. Thus,

$$b_r = \sum_i (\sigma_r)_i n_i = n_{\text{air}} \frac{24\pi^3}{\lambda^4 n_{\text{air}}^2} \left(\frac{m_{\text{air}}^2 - 1}{m_{\text{air}}^2 + 1} \right)^2 \sum_i \frac{n_i}{n_{\text{air}}} F_i = n_{\text{air}} \frac{24\pi^3}{\lambda^4 n_{\text{air}}^2} \left(\frac{m_{\text{air}}^2 - 1}{m_{\text{air}}^2 + 1} \right)^2 \sum_i v_i F_i, \quad (4.2)$$

where v_i is the fraction by volume of component i . Note that this is equivalent to defining an overall King factor

$$F_{\text{air}} = \sum_i v_i F_i \quad \text{or} \quad \left[\frac{6 + 3\delta_{\text{air}}}{6 - 7\delta_{\text{air}}} \right] = \sum_i v_i \left[\frac{6 + 3\delta_i}{6 - 7\delta_i} \right]. \quad (4.3)$$

This last equation can be solved to provide δ_{air} for any set of volume concentrations. As the King factors F_i depend on wavelength, the depolarization factor δ_{air} will as well. Formulae for the King factors for N₂ and O₂ are

$$F(\text{N}_2) = 1.034 + (3.17 \times 10^{-4})/\lambda^2$$

$$F(\text{O}_2) = 1.096 + (1.385 \times 10^{-3})/\lambda^2 + (1.448 \times 10^{-4})/\lambda^4,$$

where λ is in μm . Since Ar is monatomic, and therefore isotropic, $F(\text{Ar}) = 0$. Using these, δ_{air} is easily found to be

$$\delta_{\text{air}} = \frac{6F_{\text{air}} - 6}{7F_{\text{air}} + 3},$$

¹Note that the refractive index of air is very close to the sum of the indices of N₂, O₂, and Ar weighted by their volume concentrations: $m_g - 1 = 270.1 \times 10^{-6}$.

and for a mixture (by volume) of 78.084% N₂, 20.946% O₂ and 0.934% Ar, yields $\delta_{\text{air}} = 0.02973$ and 0.02758 at 400 and 900 nm, respectively. We have ignored CO₂ in our estimates, which is justified to the accuracy we require (including it with $F(\text{CO}_2) = 1.15$ decreases δ_{air} at 400 nm to 0.02969). Then, to the same order of approximation,

$$\begin{aligned} \mathbf{P}_r(\Theta) &= \frac{4\pi\beta(\Theta)}{b_r} \\ &= \frac{3}{2} \left[\frac{1 - \delta_{\text{air}}}{1 + \delta_{\text{air}}/2} \right] \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2}\sin^2 \Theta & 0 & 0 \\ -\frac{1}{2}\sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \\ &\quad + \frac{3}{2} \left[\frac{\delta_{\text{air}}}{1 + \delta_{\text{air}}/2} \right] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\cos \Theta \end{pmatrix}. \end{aligned} \quad (4.4)$$

The first element (11) of β is the scalar volume scattering function

$$\beta_r(\Theta) = \frac{3b_r}{16\pi} \left[\frac{(1 + \delta_{\text{air}}) + (1 - \delta_{\text{air}})\cos^2 \Theta}{1 + \delta_{\text{air}}/2} \right]. \quad (4.5)$$

Henceforth, when we used the term “Rayleigh scattering by air” we are referring to Eqs. (4.2) through Eq. (4.5) and will use the notation b_r , \mathbf{P}_r and β_r , respectively.

4.2.2 The Rayleigh Optical Depth

The optical depth of the atmosphere resulting from Rayleigh scattering is

$$\tau_r \triangleq \int_0^\infty b_r(h) dh = \int_0^\infty \sum_i (\sigma_r)_i n_i(h) dh$$

Now, $n_i(h) = v_i(h)n(h)$, where n is the actual number density at altitude h . The principal atmospheric gases are homogeneously mixed in the atmosphere (O₃ and H₂O are notable exceptions), so for these v_i is independent of h . Thus,

$$\tau_r = \sum_i \sigma_i v_i \int_0^\infty n(h) dh. \quad (4.6)$$

This can be related to the surface pressure, which is given by

$$P = \int_0^\infty m(h)n(h)g dh,$$

where $m(h)$ is the average mass of a molecule at altitude h . It is determined from the molecular weight of air, m_{air}^* (28.96 g/mole), and the Avogadro number N_s (6.022×10^{23} atoms/mole): $m(h) = m_{\text{air}}^*/N_s$. Note that $m(h)$ is independent of h (true only for well mixed gases). Thus,

$$P = \frac{m_{\text{air}}^* g}{N_s} \int_0^\infty n(h) dh, \quad (4.7)$$

and eliminating the integral from Eqs. (4.6) and (4.7) yields

$$\tau_r = \frac{N_s P}{m_{\text{air}}^* g} \sum_i \sigma_i v_i = \frac{N_s P}{m_{\text{air}}^* g} \sigma_r, \quad (4.8)$$

and the optical thickness of the atmosphere is proportional to the surface pressure, and *independent* of the variation of n with altitude. Note that one can compute the optical thickness from some altitude H to the top of the atmosphere by simply replacing the surface pressure with the pressure at the altitude H .

Most determinations of the Rayleigh optical depth of the atmosphere employ Eq. (4.2) to compute the scattering coefficient and Eq. (4.8) to compute τ_r . As discussed above, the determinations are based on precise measurements of the index of refraction of dry air and of δ_i for the various atmospheric gases. Convenient fits to tabulated determinations for sea level ($P = 1013.25$ mb) are

$$\tau_r(\lambda) = 0.008569\lambda^{-4}(1 + 0.0113\lambda^{-2} + 0.00013\lambda^{-4}) \quad (4.9)$$

and

$$\tau_r(\lambda) = 0.002152 \left(\frac{1.0455996 - 341.29061\lambda^{-2} - 0.9023850\lambda^{+2}}{1 + 0.0027059889\lambda^{-2} - 85.968563\lambda^{+2}} \right). \quad (4.10)$$

The optical depth depends on the gravitational acceleration, which at sea level varies with latitude (Lat) according to

$$g = 980.6160(1 - 0.0026373 \cos(2\text{Lat}) + 0.0000059 \cos^2(2\text{Lat})),$$

where Lat is in degrees, and g is in cm/sec². Equation (4.10) is for Lat = 45°.

Figure 4.1 shows the atmospheric direct transmittance of the solar beam from the top of the atmosphere to the surface (Sun at the zenith) as affected by Rayleigh scattering,

$$\text{Transmittance} = \exp[-\tau_r(\lambda)].$$

Since the gases in the atmosphere also absorb light, this is the upper limit to the direct transmittance of the atmosphere for any wavelength λ .

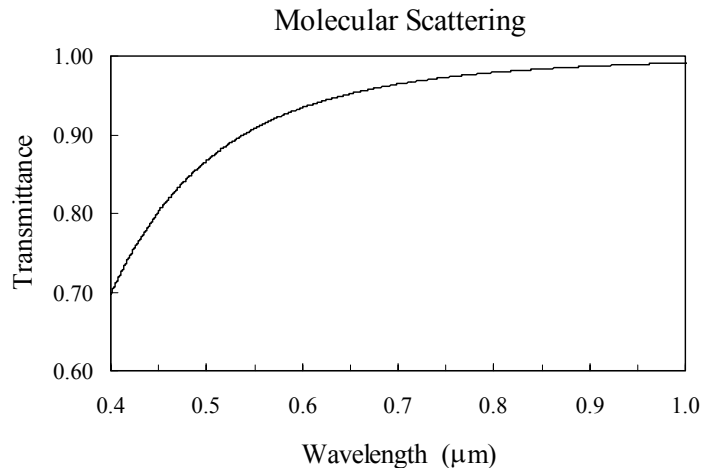


Figure 4.1: Direct transmittance of the atmosphere resulting from Rayleigh (molecular) scattering along a vertical path from the surface to space.

4.3 Scattering by Particles (Aerosols)

The various species of particulate matter suspended in the air are referred to as *aerosols*. Examples of such particles are wind-blown dust usually originating in desert areas and transported over the oceans; sea-salt aerosol, resulting from the breaking of waves on the water surface; sulfuric acid particles formed from condensed sulfur-containing compounds released during volcanic eruptions; carbonaceous particles released through the burning of fossil fuels, etc. As these particles remain suspended in the air for substantial periods of time (a few days), they are of necessity small, i.e., *maximum* dimension in the tens of micrometers (microns) size range. The aerosol particles are distributed in size from this upper limit down to a fraction of a micron.

As we shall see below, aerosols typically are present in very low concentrations (~ 100 particles/ml), and thus very difficult to sample. In addition, the individual aerosol particles are often complex in morphology and inhomogeneous in composition — an exception being the maritime aerosol, which is composed of spherical, mostly homogeneous, salt-water droplets. In order to render the optical effects of aerosols tractable, it is almost always assumed that the aerosol is composed of homogeneous spherical particles.² In addition, this enables the optical observations to characterize aerosol properties; however, one must remember that any physical property derived from optical measurements is subject to the proviso that the particles have been assumed to be homogeneous spheres.

²This assumption is employed in this work unless otherwise indicated.

Employing the homogeneous-spherical particle assumption, for a particular species (the i^{th}) of particle we define its size frequency distribution $n_i(D)$ as $dN_i(D)/dD$, where $dN_i(D)$ is the number of particles per unit volume with diameters between D and $D + dD$. The scattering and absorption properties of this species of particle is determined by an integral of the product of the size frequency distribution and the appropriate scattering (or absorption) quantity provided by Mie theory (Chapter 3), e.g., the volume scattering function,

$$\beta_i(\Theta, \lambda) = \frac{\lambda^2}{4\pi^2} \int_0^\infty M_{11}(\Theta, D, \tilde{m}_i) \frac{dN_i(D)}{dD} dD, \quad (4.11)$$

where \tilde{m}_i is the complex refractive index of the species (i) relative to the medium (air).³ Thus, for homogeneous spheres, the two properties of the particles that determine the optical properties are the size distribution of the species and the index of refraction of the material of which the particles are composed.

4.3.1 Aerosol Particle Size Distributions

There are many ways of estimating particle size distributions. The most direct is to pump air containing the particles through a small-pore filter, and to examine the filter with either an optical microscope (for larger particles) or an electron microscope (for smaller particles). An obvious disadvantage is the loss of volatile particles during the process.

Another method is to use a *cascade impactor*. It works in the following manner. Air containing the particles to be sized flows through an orifice and immediately impinges on

³Equation (4.11) provides the seeds for estimating the physical properties from optical measurements. Assuming there is only one component and that its refractive index is \tilde{m} , if one measures $\beta(\Theta)$ at several scattering angles, Θ_i , at a single wavelength, then

$$\beta(\Theta_i) = \frac{\lambda^2}{4\pi^2} \int_0^\infty M_{11}(\Theta_i, D, \tilde{m}) \frac{dN(D)}{dD} dD.$$

Converting this to a sum

$$\beta(\Theta_i) = \sum_j K_{ij} n(D_j), \text{ where } K_{ij} = \frac{\lambda^2}{4\pi^2} M_{11}(\Theta_i, D_j, \tilde{m}) \Delta D_j.$$

If we choose the number of D_j 's equal to the number of Θ_i 's, this set linear of equations can be inverted to yield $n(D_j)$. This inversion is equivalent to finding the inverse of the matrix \mathbf{K} with elements K_{ij} . Although straightforward, there is a difficulty to be faced with this procedure. The determinant of \mathbf{K} is often close to zero because β over some range of angles may be insensitive to variations in D . In this case the system of linear equations is said to be "ill conditioned," and this causes large fluctuations in $n(D_j)$ with D_j . This can be partially overcome by placing constraints on $n(D)$, e.g., that is be "smooth" in some quantifiable sense. This procedure is called "constrained linear inversion" or "regularization." Note that these ideas can be applied to other optical measurements as well, e.g., c as a function of λ .

a flat plate obstructing the flow. The larger particles, because of their momentum strike and stick to the plate. The smaller particles remain with the air as it flows around the plate where it encounters a smaller orifice. The air speed increases as it flows through the second orifice where it immediately encounters a second flat plate. This plate collects particles smaller than the first; however, still smaller particles remain in the airstream, which is again channeled around the plate to a still smaller orifice, etc. By cascading several of these assemblies with smaller and smaller orifices, a fractionation of particle sizes is effected.

A third method is the resistive-pulse particle counter. In this apparatus, particles are suspended in saline (salt water). Two electrodes are involved, one in the sample and one in a test tube suspended in the sample. The suspended test tube also contains saline and has a small hole (diameter $\sim 100\mu\text{m}$) in its side. Both electrodes are connected to a DC power source causing current to flow through the hole. The solution is pumped through the hole and when a particle passes through the orifice the current is reduced in proportion to the particle's volume (V). On an oscilloscope the drop in current is measured as a pulse with height (depth) proportional to V . By analyzing the height of the pulses as a given volume of fluid is pumped through the aperture, the distribution of particle volume is determined. Obviously water-soluble particles are lost in the process. (However, the method is clearly applicable to particles suspended in sea water.)

Optical methods are obviously available in which one measures the differential scattering cross section of particles in a flow and interprets the result using light scattering theory. One of the most popular methods (when the composition of the particles is known or can be estimated) involves the analysis of light scattering by a volume of sample containing many particles. One measures a light scattering quantity such as the scattering coefficient (b) or the volume scattering function ($\beta(\Theta)$), assumes a mathematical form for the size distribution (with adjustable parameters), and uses the measurements along with relationships such as Eq. (4.11), to estimate the parameters of the assumed size distribution. Unfortunately, the concentration of aerosol particles in the marine atmosphere (our principal interest) is only of the order of a few hundred per ml, making it very difficult to obtain measurements of c or β . A simple example will help clarify this. Consider a spherical particle with $\tilde{n} = 1.50 - 0.0i$ and $D = 1\mu\text{m}$. At 500 nm, Mie theory provides $Q_c = Q_b \approx 2.3$. Thus, for one particle, $\sigma_c = \sigma_b = \pi D^2 Q_c / 4 \approx 2\mu\text{m}^2$. For 100 particles per ml, which is 10^{-10} particles per μm^3 , $c = b = 2 \times 10^{-12} \mu\text{m}^{-1} = 0.002 \text{ km}^{-1}$. For $D = 5\mu\text{m}$, $c \approx 0.05 \text{ km}^{-1}$. Clearly, extremely long paths (difficult to obtain in the marine environment) are required to accurately measure b or c , and the measurement of β is virtually impossible except at very small angles.

Very long paths are available if one uses the Sun as the source. In this case, quantities related to c and β can be estimated by measuring the radiance of the solar disk and the sky,

respectively, as follows. The downward radiance at the bottom of a slab (the atmosphere) of optical thickness τ_1 , and illuminated by a beam source (the Sun), was derived in the single-scattering approximation in Chapter 2 (Eq. (2.103)):

$$L(\tau_1, u, \phi) = F_0 \delta(u - u_0) \delta(\phi - \phi_0) \left[1 - \frac{\tau_1}{u} \right] + \frac{\omega_0 F_0 \tau_1}{4\pi u} P(u_0 \rightarrow u, \phi_0 \rightarrow \phi).$$

In this equation, $u = \cos \theta$, where θ and ϕ are the polar and azimuth angles of the propagation vector of the radiance ($\hat{\xi}$), and $u_0 = \cos \theta_0$, where θ_0 and ϕ_0 are the polar and azimuth angles of the propagation vector of the solar beam ($\hat{\xi}_0$). The scattering phase function of the slab is P , and the single scattering albedo is ω_0 . Note that the spherical coordinate system in which these angles are measured has its z -axis directed downward. Also, the radiometer carrying out the measurements is directed toward $-\hat{\xi}$. Rewriting the equation and noting that the δ -function term is the first order expansion of the exponential function, we have

$$L(\tau_1, \theta, \phi) = F_0 \delta(\cos \theta - \cos \theta_0) \delta(\phi - \phi_0) \exp \left[-\frac{\tau_1}{\cos \theta_0} \right] + \frac{\omega_0 F_0 \tau_1}{4\pi \cos \theta} P(\Theta),$$

where $\cos \Theta = \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\phi - \phi_0)$. In the absence of absorbing gases, $\tau_1 = \tau_r + \tau_a$, where τ_r is the optical depth of the air (the “Rayleigh” optical depth of the atmosphere) and τ_a is the optical depth of the aerosol:

$$\tau_a = \int_0^\infty c_a(h) dh$$

with h being the altitude. Likewise, $\omega_0 \tau_1 P(\Theta) = \tau_r P_r(\Theta) + \omega_a \tau_a P_a(\Theta)$, where the subscripts “ r ” and “ a ” refer to “Rayleigh” and “aerosol,” respectively. Since $P = 4\pi\beta/b$, $\omega_0 = b/c$, and for a homogeneous slab with no absorption $\tau_1 = bz_1$, we have (again, assuming no gaseous absorption and a homogeneous slab) $\tau_a \omega_a P_a(\Theta) = 4\pi\beta_a(\Theta)$.⁴ Since aerosols from all altitudes contribute to the sky radiance, the values of τ_a and $\tau_a \omega_a P_a(\Theta)$ are referred to as *columnar* optical properties, and using them in Eq. (4.11) will provide a *columnar aerosol size distribution*, i.e., the aerosol size distribution for a vertical column of atmosphere. The

⁴The assumptions of single scattering and a homogeneous slab are not necessary to extract the aerosol phase function from such measurements. Iterative methods are now available to effect such a retrieval in a multiple scattering layered atmosphere, and have been extended to permit retrieval the full scattering phase matrix. These are developed in the appendix to Chapter 12 (Section 12.5).

columnar aerosol size distribution is most appropriate for our purposes, as the ultimate goal is to find aerosol models that reproduce the optical properties that, when combined with radiative transfer, can predict the radiation reflected from the Earth's atmosphere. An active area of research is to achieve *closure* between the columnar aerosol size distribution and the aerosol size distribution in each layer forming the column.

One of the earliest proposed particle size distributions for aerosols was the power-law (often called Junge, although strictly speaking, a Junge distribution refers only to a power law with $\nu = 3$, i.e., D^{-4} in Eq. (4.12)):

$$n(D) = \frac{dN(D)}{dD} = \frac{K}{D^{\nu+1}}, \quad (4.12)$$

where K , and ν are constants. The parameter ν is typically between 2 and 4. This distribution cannot of course be valid over the entire size range $0 \leq D \leq \infty$, e.g., for typical values of ν the total number of particles per unit volume $\int_0^\infty n(D) dD$ would be infinite. Thus, the distribution is usually truncated, holding only for $D_S \leq D \leq D_L$ and vanishing outside this range. For the truncated distribution the constant K is related to the total number of particles:

$$N_0 = \int_0^\infty dN = K \int_{D_S}^{D_L} D^{-(\nu+1)} dD,$$

so

$$K = \frac{N_0 \nu}{(D_S^{-\nu} - D_L^{-\nu})}.$$

In aerosol studies, rather than using the distribution of the number of particles per unit volume between D and $D + dD$, it is popular to use the distribution of the *volume* of particles per unit volume between D and $D + dD$. This is

$$\frac{dV(D)}{dD} = \frac{\pi D^3}{6} \frac{dN(D)}{dD},$$

which, for a power-law distribution yields

$$\frac{dV(D)}{dD} = \frac{\pi D^3}{6} K D^{-(\nu+1)} \triangleq K' D^{-(\nu-2)},$$

where K' usually related to the total volume of particles, V_0 , through

$$V_0 = K' \int_{D_S}^{D_L} D^{-(\nu-2)} dD,$$

which gives

$$\begin{aligned} K' &= \frac{(3-\nu)V_0}{(D_L^{3-\nu} - D_S^{3-\nu})}, & \nu \neq 3, \\ &= \frac{V_0}{\ln(D_L/D_S)}, & \nu = 3. \end{aligned}$$

Or, in terms of $\ln D$,

$$\frac{dV(D)}{d\ln D} = K' D^{3-\nu}.$$

Thus, for a particle size distribution with $\nu = 3$, the volume distribution $dV(D)/d\ln D$ is independent of D .

A more realistic representation of the size spectrum (no truncation is required) is the *log-normal* distribution:

$$\frac{dN}{d\ln D} = \frac{N_T}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{\ln D - \ln D_N}{\sigma} \right)^2 \right], \quad (4.13)$$

where N_T (the total number of particles per unit volume), D_N (the *median* or *modal* diameter) and σ are constant.⁵ An equivalent form of this distribution is used later in this chapter:

$$\frac{dN(D)}{dD} = \frac{N_T}{\log_e(10)\sqrt{2\pi}\sigma_{10}D} \exp \left[-\frac{1}{2} \left(\frac{\log_{10}(D/D_N)}{\sigma_{10}} \right)^2 \right], \quad (4.14)$$

where $\sigma_{10} = \sigma \log_{10}(e)$. The log-normal distribution can also be converted to a volume distribution. The result is

$$\frac{dV}{d\ln D} = \frac{V_T}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{\ln D - \ln D_V}{\sigma} \right)^2 \right], \quad (4.15)$$

where V_T (the total particle volume) and D_V (the median diameter for volume) are related to N_T and D_N through

$$V_T = \frac{N_T}{6} D_N^3 \exp \left[\frac{9}{2} \sigma^2 \right] \quad \text{and} \quad D_V = D_N \exp[3\sigma^2].$$

⁵Note the similarity between Eq. (5.28) and the *Gaussian* or *Normal* distribution of a random variable x :

$$\frac{d\mathcal{P}(x)}{dx} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x - x_m}{\sigma} \right)^2 \right],$$

where $d\mathcal{P}(x)$ is the probability that x falls between x and $x + dx$. For this distribution the mean of x , $\langle x \rangle = x_m$ and the variance $\langle (x - \langle x \rangle)^2 \rangle = \sigma^2$. Thus, for the log-normal distribution (Eq. (5.28)) $\langle \ln D \rangle = \ln D_N$ and $\langle (\ln D - \langle \ln D \rangle)^2 \rangle = \langle (\ln D - \ln D_N)^2 \rangle = \sigma^2$.

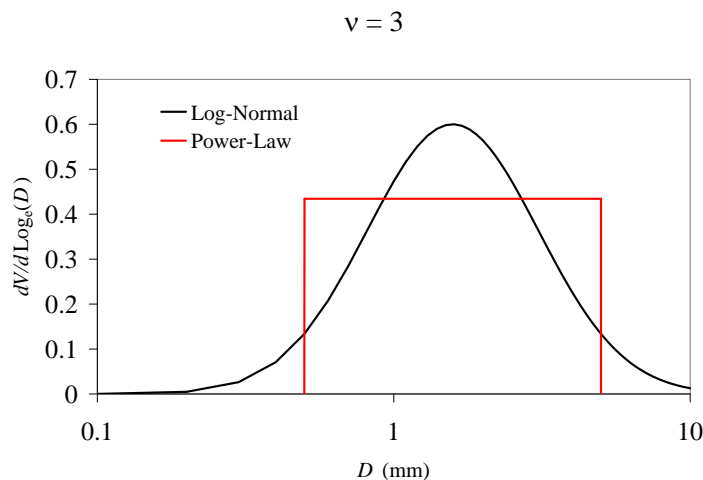


Figure 4.2: Comparison between power-law and log-normal volume distributions. For the power-law distribution, $\nu = 3$, $D_S = 0.5 \mu\text{m}$, and $D_L = 5.0 \mu\text{m}$. The log-normal volume distribution has the same D_V , σ , and total volume as the power-law distribution.

Figure 4.2 compares the volume distribution for the power-law distribution ($\nu = 3$, $D_S = 0.5 \mu\text{m}$, and $D_L = 5.0 \mu\text{m}$) and the log-normal distribution that has the same D_V and σ as the Junge distribution.

4.3.2 Aerosol Refractive Indices

Aerosol particles found in the marine atmosphere are composed of four broad types of materials: (1) water; (2) sea salt; (3) mineral particles, e.g., dust; and (4) carbonaceous particles usually originating from urban pollution or biomass burning, e.g., forest fires. The refractive index of water throughout the visible spectrum is approximately 1.33, with an insignificant imaginary part ($|m_i| < 4 \times 10^{-7}$). Sea salt (dry) has a refractive index of about 1.5 with negligible absorption ($|m_i| < 10^{-8}$). Mineral particles have $m_r \sim 1.53$ in the visible with $m_i \sim 10^{-3} - 10^{-2}$. The spectral dependence of the absorption of minerals depends significantly on their composition, e.g., minerals with a significant quantity of hematite have a strongly increasing $|m_i|$ from $\lambda \sim 1 \mu\text{m}$ into the UV, whereas quartz has low $|m_i|$ throughout the visible. Carbonaceous particles (black carbon) on the other hand have large absorption throughout the visible ($|m_i| \sim 0.1$) with little dependence on

wavelength. Thus refractive indices span the range $1.33 \leq m_r \leq 1.53$ with the absorption index $10^{-8} \leq |m_i| \leq 10^{-1}$.

4.3.3 Aerosol Models

An aerosol physical model is a combination of an aerosol size distribution, refractive index, and shape. For simplicity, the particle shape is usually taken to be a homogeneous sphere. This model is then used in combination with scattering theory to determine the scattering and absorption properties of the aerosol. The combination aerosol physical model and scattering theory is herein called an “aerosol model.”⁶

⁶The fact that extinction and absorption cross sections can be written $\sigma_{c,a} = GQ_{c,a}$, where G is the geometric cross sectional area of the particle (the “shadow” of the particle in the geometrical optics approximation: πR^2) and $Q_{c,a}$ is approximately constant for particles that are large compared to the wavelength, suggests that for the purpose of optical properties, the characteristic or *effective* radius of a spherical particle should be defined according to

$$R_{\text{eff}} = \frac{1}{G} \int_0^\infty R[\pi R^2 n(R)] dR, \text{ where } G = \int_0^\infty [\pi R^2 n(R)] dR.$$

That is, for optical purposes, the effective distribution of sizes is $\pi R^2 n(R)$. Similarly, the effective variance, v_{eff} is defined by

$$v_{\text{eff}} = \frac{1}{GR_{\text{eff}}^2} \int_0^\infty (R - R_{\text{eff}})^2 [\pi R^2 n(R)] dR.$$

For the log-normal distribution these are related to R_N and σ through

$$R_N = \frac{R_{\text{eff}}}{(1 + v_{\text{eff}})^{5/2}} \text{ and } \sigma^2 = \ln(1 + v_{\text{eff}}).$$

Empirically, it has been shown that dimensionless optical quantities such as the scattering phase function or the size-averaged extinction efficiency

$$\langle Q_c \rangle \triangleq \frac{\langle \sigma_c \rangle}{\pi R_{\text{eff}}^2} = \frac{\int_0^\infty Q_c [\pi R^2 n(R)] dR}{\pi R_{\text{eff}}^2},$$

are similar for size distributions with the same values of R_{eff} and v_{eff} , but with significantly different functional forms. This suggests that the scattering and absorption properties of a given species of particle (i.e., a given \tilde{m}) can be characterized by the two parameters R_{eff} and v_{eff} or equivalently D_{eff} and v_{eff} .

4.3.3.1 Haze C Aerosol Model

One of the earliest models used to describe aerosol scattering was the so-called “Haze C” model for which

$$\begin{aligned} n(D) &= \frac{dN(D)}{dD} = K, & D_0 < D < D_1, \\ &= K \left(\frac{D_1}{D} \right)^{\nu+1}, & D_1 < D < D_2, \\ &= 0, & D > D_2, \end{aligned} \quad (4.16)$$

with $D_0 = 0.06 \mu\text{m}$, $D_1 = 0.20 \mu\text{m}$, $D_2 = 20 \mu\text{m}$ and $\nu = 2, 3$, or 4 . This size distribution has been used with a refractive index of 1.333 or 1.50 to examine the scattering properties of sea salt and mineral (non-absorbing) aerosols.

One aspect of power-law size distributions is that, if the refractive index is independent of wavelength, the scattering phase function is nearly independent of wavelength. Consider the size distribution given by Eq. (4.12). If this size distribution was valid for the entire size range $0 \leq D \leq \infty$, it would lead to some interesting optical properties. The volume scattering function would be

$$\begin{aligned} \beta(\Theta, \lambda) &= \frac{\lambda^2}{4\pi^2} \int_0^\infty M_{11}(\Theta, D, \tilde{m}) K \left(\frac{1}{D} \right)^{\nu+1} dD \\ &= \left[\frac{K\pi^{\nu-2}}{4} \right] \lambda^{2-\nu} \int_0^\infty M_{11}(\Theta, x, \tilde{m}) x^{-(\nu+1)} dx, \end{aligned} \quad (4.17)$$

where we have changed the variable of integration from D to $x = \pi D/\lambda$. If \tilde{m} is independent of λ then the integral is as well, and the full wavelength dependence of the volume scattering function is provided by the $\lambda^{2-\nu}$ term. Thus, for such a size-refractive index distribution, the volume scattering function, the total scattering coefficient, and the absorption coefficient, are all proportional to $\lambda^{2-\nu}$, and the scattering phase function is *independent* of wavelength. Reviewing the derivation of this result, it is seen that, in addition to a wavelength-independent refractive index, the key ingredient is that the limits on the integral do not depend on λ . If the size distribution were cut off at some lower size D_S and/or some large size D_L , then the upper and lower limits on the second integral in (4.17) would be $\pi D_L/\lambda$ and $\pi D_S/\lambda$, respectively, so the integral would have a λ dependence. Recall that for small spherical particles, the scattering is $\propto x^6$, while for large particles, it is $\propto x^2$. Thus, the integrals for the scattering quantities will converge for $2 < \nu < 6$ (with integration over the full size range $0 \rightarrow \infty$). More importantly, if ν is significantly smaller than 6 , the integral will be relatively insensitive to the lower limit, while if ν is significantly larger than 2 the integral will be insensitive to the upper limit.⁷ Thus, one

⁷The dependence of the scattering quantities on D_S and D_L is investigated for Junge distributions of particles suspended in water in Chapter 5 (in particular, see Section 5.5.2.2).

might reasonably expect that for some range in ν , a size distribution truncated by D_S and D_L will yield scattering properties that are somewhat insensitive to the exact values of D_S and D_L , i.e., a truncated power-law distribution should yield a phase function that depends only weakly on wavelength and for which the scattering coefficient is approximately $\propto \lambda^{2-\nu}$. This is in fact what is often found for aerosols. Measurements of atmospheric transmittance processed to yield the aerosol optical depth $\tau_a(\lambda) = \int_0^\infty b_a(h, \lambda) dh$, usually closely follow Ångström's law:

$$\tau_a(\lambda) = \tau_a(\lambda_0) \left(\frac{\lambda_0}{\lambda} \right)^p$$

where p is a constant.⁸

Do the scattering properties of the Haze C distribution follow these observations, i.e., Ångström's law with $p = \nu - 2$? Figure 4.3 provides the scattering coefficient for the Haze C distributions with $\tilde{m} = 1.33 - 0.0i$ and $1.50 - 0.0i$. This figure shows that the scattering coefficient follows Ångström's law over the wavelength range 400-2200 nm very well. The prediction of $p = \nu - 2$ is certainly well approximated. The other prediction — that the phase function should be independent of wavelength — is examined in Figure 4.4, which compares the scattering phase functions at 412 and 865 nm. The figure shows that for $\nu = 2$ the phase function is indeed a weak function of wavelength; however, this is not the case for $\nu = 4$, especially for scattering angles $\gtrsim 80^\circ$. For completeness, Figure 4.5 shows the range of variation of the phase function at 510 nm for the full range of variation of ν and \tilde{m} . As expected, smaller values of ν (which yield relatively more large particles) show more scattering in the near forward direction (diffraction regime), while large values of ν (which yield relatively more small particles) produce an elevated phase function over the angular range $30^\circ - 120^\circ$. The phase function maxima near 140° ($\nu = 2$, $\tilde{m} = 1.33 - 0.0i$) and 160° ($\nu = 2$, $\tilde{m} = 1.50 - 0.0i$) are near the positions of the primary rainbows for the given \tilde{m} (see Chapter 3).

⁸This explains why Junge distributions were popular in early aerosol studies: determination of one parameter, p , that is relatively easy to measure, provided quantitative information on the columnar size distribution.

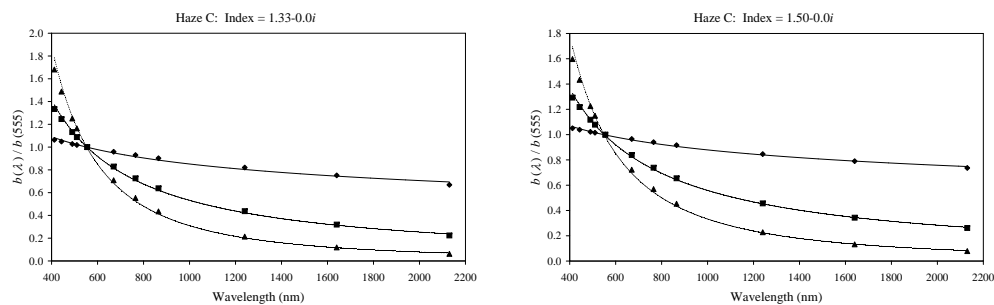


Figure 4.3: Spectral variation of the scattering coefficient for the Haze C models (left, $\tilde{m} = 1.33 - 0.0i$; right, $\tilde{m} = 1.50 - 0.0i$). The diamonds are for $\nu = 2$, the squares for $\nu = 3$, and the triangles for $\nu = 4$. The smooth curves are regression fits to $b(\lambda)/b(555) = C\lambda^{-p}$. The values of p and R^2 from the regression fits: 0.27 and 0.9864, 1.07 and 0.9978, and 1.98 and 0.9977, respectively for $\nu = 2, 3$, and 4, with $\tilde{m} = 1.33 - 0.0i$; and 0.21 and 0.9945, 0.98 and 0.9992, and 1.84 and 0.9987, respectively for $\nu = 2, 3$, and 4, with $\tilde{m} = 1.50 - 0.0i$.

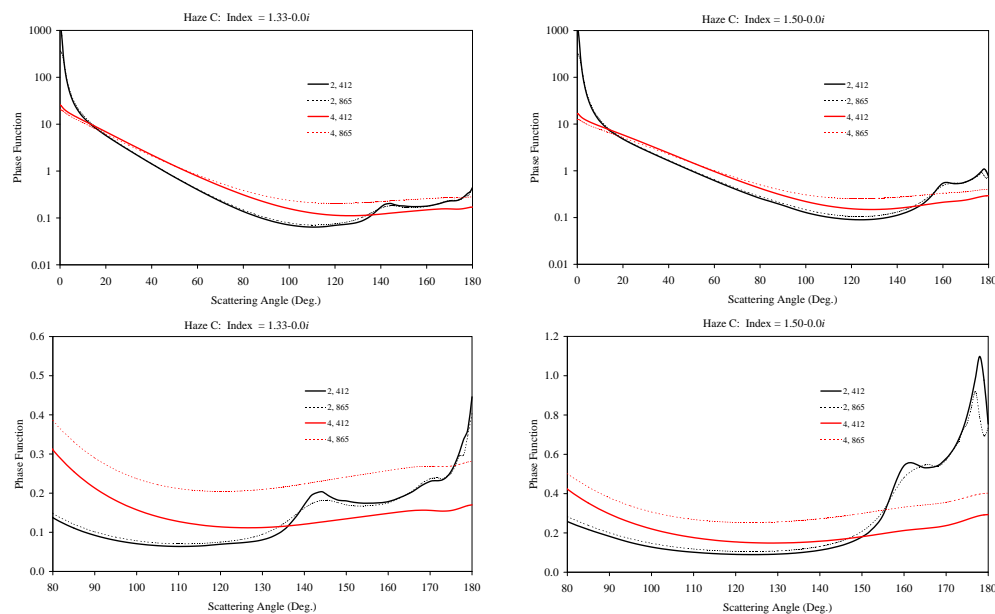


Figure 4.4: Comparison of the phase functions at $\lambda = 412$ and 865 nm for Haze C models. Left, $\tilde{m} = 1.33 - 0.0i$; right, $\tilde{m} = 1.50 - 0.0i$. In the legends, “2, 412” stands for $\nu = 2$ and $\lambda = 412$, etc.

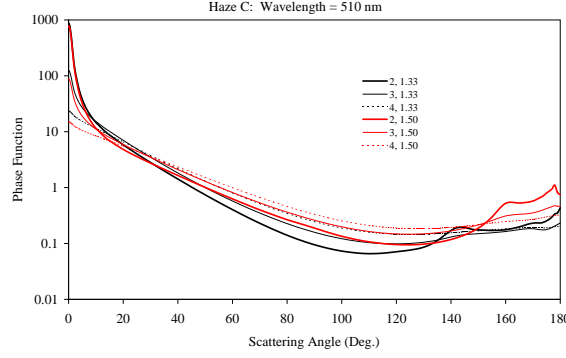


Figure 4.5: Comparison of the phase functions at $\lambda = 510$ nm for Haze C models. In the legend, “2, 1.33” stands for $\nu = 2$ and $\tilde{m} = 1.33 - 0.0i$, etc.

4.3.3.2 Shettle and Fenn Aerosol Model(s)

The Shettle and Fenn models consist of particles distributed in size according to combinations of log-normal distributions. The size frequency distribution $n(D)$ is given by

$$n(D) = \sum_{i=1}^2 n_i(D),$$

where n_i are given by equations of the form of Eq. (5.29). The models are based on the hypothesis that there is a background aerosol consisting of fine particles in the size range $0.05 \leq D \leq 1\mu\text{m}$. Their small size enables them to stay in suspension over an extended period of time. This background aerosol is accompanied by a component in a larger size range, $D > 1\mu\text{m}$, that is locally generated, and because of their larger size, settle out of the air much more rapidly. An important example for us is the marine aerosol consisting of small particles transported from the land and larger, sea salt, particles generated by breaking waves. Shettle and Fenn use knowledge of the possible composition of the aerosol to estimate their (dry) refractive indices, and the manner in which their size and refractive index vary with relative humidity.

The background component (called “Rural 1” by Shettle and Fenn) is a mixture of 70% water soluble material and 30% dust (quartz like). It is usually called the “Tropospheric”

component. In urban areas this component is replaced by a carbonaceous aerosol called “Urban 1.” The large-particle component consists of a sea-salt aerosol (“Oceanic”) in the marine environment, a larger-sized component similar in composition to Rural 1 (“Rural 2”) in the terrestrial environment, and a large-sized component similar to Urban 1 (“Urban 2”) in the polluted air masses of the urban environment.

The particles are assumed to be hygroscopic, so both the particle size and refractive index vary with relative humidity (RH in %). If $D_i(RH)$ is the value of D_i at a given RH , and $D_i(0)$ the value for a dry particle, then the refractive index will vary with RH according to the volume concentrations of water and dry particles, i.e.,

$$\begin{aligned} m_i(RH) &= \frac{m^{\text{water}}V_{\text{water}} + m_i^{\text{dry}}V_{\text{dry}}}{V_{\text{water}} + V_{\text{dry}}} \\ &= m^{\text{water}} + (m^{\text{dry}} - m_i^{\text{water}}) \left[\frac{D_i(0)}{D_i(RH)} \right]^3 \end{aligned}$$

Thus, $m_i(0) = m^{\text{dry}}$ and as RH approaches 100%, $m_i(RH)$ approaches that of water. The refractive indices (dry) of the various components are provided in Figure 4.6. The dry Oceanic is sea salt. Like water it is characterized by very low absorption throughout the visible. In contrast, the other components have moderate to strong absorption through the visible into the short-wave infrared (SWIR). “Mineral” is a hematite-containing mineral (as opposed to quartz) and is characteristic of Saharan dust, but is not actually used in the Shettle and Fenn models. With the exception of water, all components have $m_r \approx 1.5$ throughout the visible and near infrared (NIR).

Figure 4.7 provides the variation of the modal diameter for each component of the size distribution, and shows that the diameter increases rapidly with RH for $RH \gtrsim 70\%$. It also shows that there is essentially no variation in size for $RH \leq 50\%$. The remaining parameter in the size distributions σ_{10} , is 0.35 for the small fractions and 0.40 for the large size fractions.

From these components, four basic models were constructed: the Tropospheric model with no Oceanic contribution; the Maritime model for which 99% of the particles (in number) have the Tropospheric characteristics and 1% the Oceanic; and the Coastal model for which 99.5% of the particles have the Tropospheric characteristics and 0.5% the Oceanic, and the Oceanic model with no Tropospheric contribution. The Coastal model was introduced to represent the aerosol over the ocean nearer the coast (relatively more Tropospheric contribution). The Urban models are rarely used in remote sensing; usually only to show

the effects of a very strongly absorbing aerosol. They are particles distributed in a manner similar to the Rural models, but are composed of 80% mineral-like Rural particles and 20% carbonaceous particles. Figure 4.8 provides the volume size distributions of several of these models. There, and henceforth, the notation M99 stands for the Maritime model with 99% relative humidity, etc. In the figure the models are all normalized to unity at their maximum value. At the scale of the drawing O99 and M99 are virtually identical.

Assuming again that the aerosol particles are homogeneous spheres, their optical properties can be determined using Mie theory. Figure 4.9 shows the spectral variation of the extinction coefficient for some of these models. In the figure, the lines are regression fits to $c(\lambda) \propto \lambda^{-p}$. Clearly, these models fit Angstrom's law throughout the visible and NIR as well as the Haze C models. The scattering phase functions for O99 and T80 are shown in Figure 4.10. Note that the spectral variation of the Shettle and Fenn models is also similar to that for the Haze C models. Figure 4.11 (left panel) provides the range of variability of the phase function at 865 nm for the Shettle and Fenn models, and shows the progression from a flatter to a more forward peaked phase function as one progresses from T80 to O99 in much the same manner as progressing from $\nu = 4$ to $\nu = 2$ with the Haze C models (Figure 4.4).

Finally, Figure 4.11 (right panel) shows the degree of polarization ($-M_{12}(\Theta)/M_{11}(\Theta)$) at 865 nm for the models in Figure 4.11. Recall from Chapter 2 that, in the single scattering approximation, the polarization properties of scattered solar irradiance (unpolarized) are determined by $M_{11}(\Theta)$ and $M_{12}(\Theta)$, and in fact the degree of polarization (DOP) is just $(-M_{12}(\Theta)/M_{11}(\Theta))$. The figure shows that when the large particle component is missing, as in T80, the DOP resembles that for Rayleigh scattering (maximum at $\Theta = 90^\circ$). As more and more of the large particle component is added (M50→O99), the DOP is reduced near $\Theta = 90^\circ$ and a second maximum is developed near $\Theta \approx 140^\circ$, close to the “rainbow” angle for \tilde{m} close to that of water. For positive values of $-M_{12}/M_{11}$, the scattered light will have a stronger component perpendicular to the scattering plane rather than parallel to it, and vice versa for negative values.

The Shettle and Fenn models all have some absorption, therefore in general the single scattering albedo, $\omega_0 \neq 1$. The range of variation of ω_0 for the Tropospheric and Maritime aerosol models is shown in Figure 4.12. The Oceanic models all have $\omega_0 = 1$ (i.e., $1 - \omega_0 < 10^{-4}$). As Figure 4.7 shows that for $RH < 50\%$ the influence of RH on the particle size is negligible, the T50 and M50 models essentially represent dry particles. Over the visible and NIR portions of the spectrum, the Urban models (not shown) have $0.60 \leq \omega_0 \leq 0.66$ for dry particles and $0.936 \leq \omega_0 \leq 0.944$ for $RH = 99\%$.

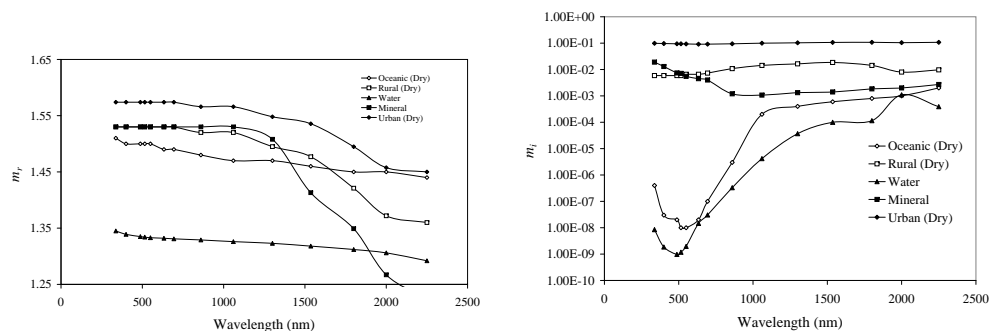


Figure 4.6: Real (left) and imaginary (right) parts of the complex index of refraction for the Shettle and Fenn aerosol model components. With the exception of water, all components are in the dry state.

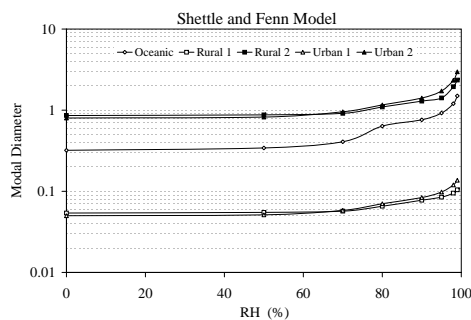


Figure 4.7: Variation of the modal diameter D_i with relative humidity for the various components in the Shettle and Fenn aerosol model.

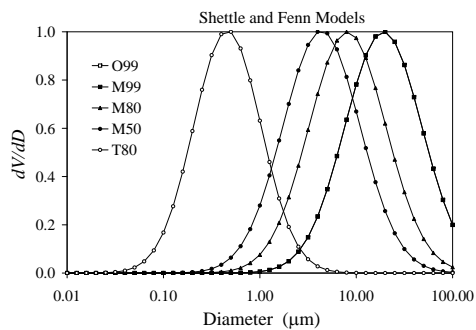


Figure 4.8: Volume size distribution for several of the Shettle and Fenn aerosol models. Note that M99 and O99 are virtually identical on the scale of the drawing. They differ only for particle sizes less than about $0.10 \mu\text{m}$

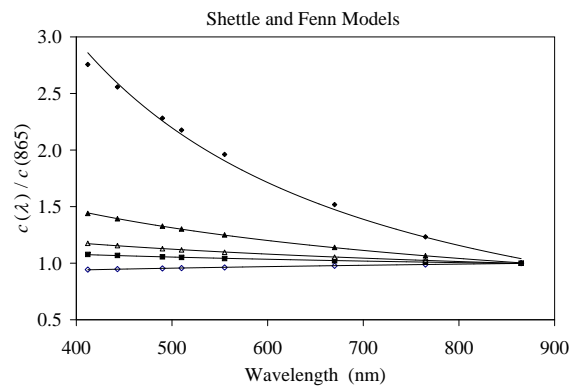


Figure 4.9: Spectral variation of the extinction coefficient for the Shettle and Fenn models: filled diamonds – T80; filled triangles – M50; open triangles – M80; filled squares – M99; open diamonds – O99. The lines are regression fits to $c(\lambda) \propto \lambda^{-p}$. The values of p and R^2 (of the fit) are 1.35 and 0.9943, 0.49 and 0.9997, 0.22 and 0.9990, 0.10 and 0.9951, and -0.08 and 0.9951, respectively for T80, M50, M80, M99, and O99.

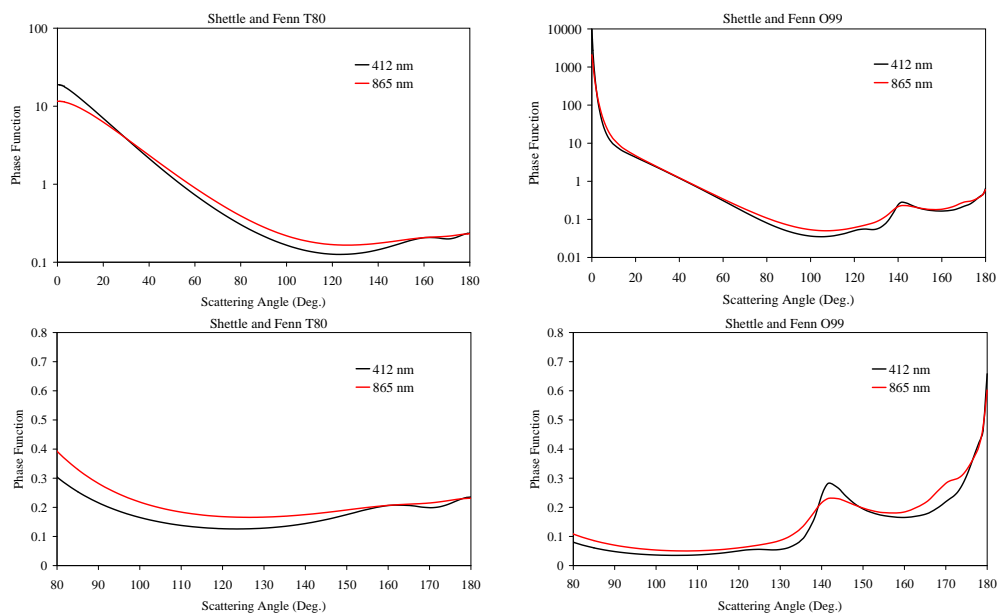


Figure 4.10: Comparison of the phase functions at $\lambda = 412$ and 865 nm for the Shettle and Fenn T80 and O99 models.

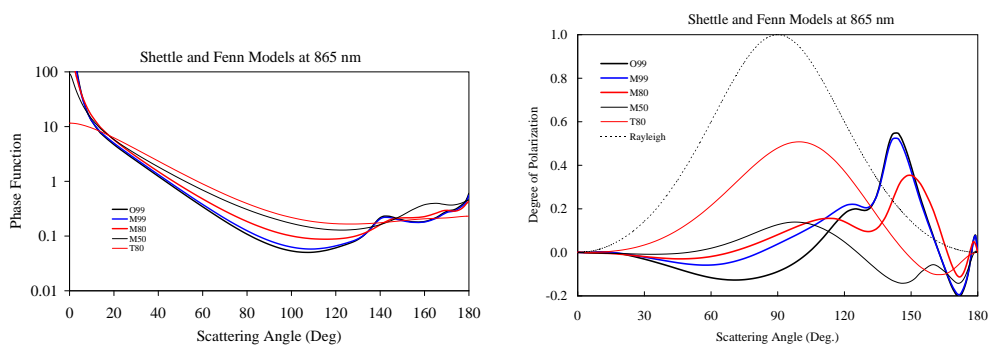


Figure 4.11: Left Panel: Comparison of the phase functions at $\lambda = 865$ nm for various Shettle and Fenn aerosol models. Right Panel: Degree of polarization at 865 nm for various Shettle and Fenn aerosol models. Actually plotted is $-M_{12}(\Theta)/M_{11}(\Theta)$ for these models.

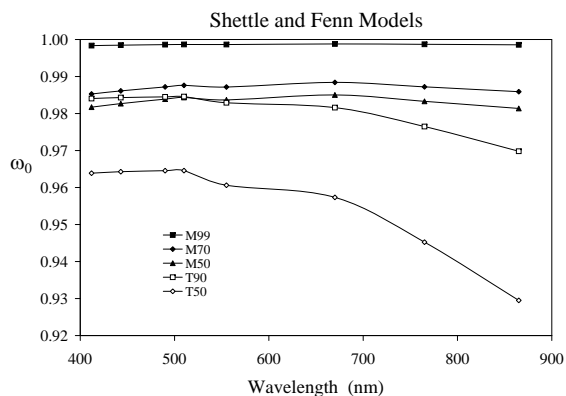


Figure 4.12: Single scattering albedo for various Shettle and Fenn aerosol models.

4.3.3.3 Smirnov et al. Model

To achieve a better understanding of the aerosol in a “clean” maritime environment, Smirnov and coworkers analyzed a large set of size distribution retrievals from three remote areas: Lanai in the Pacific (Hawaii), Bermuda in the Atlantic, and Kaashidoo in the Indian Ocean (Maldives). The size distributions were retrieved from measurements of the sun and sky radiance made with the Aerosol Robotic Network (AERONET).⁹ Although a broad distribution of aerosol optical depth at 500 nm ($\tau_a(500)$) was observed (mostly less than 0.3) the values retrieved with the highest frequency were 0.06, 0.09, and 0.11 at Lanai, Bermuda, and Kaashidoo, respectively. Limiting their analysis to retrievals with $\tau_a(500) \leq 0.15$, they retrieved the mean columnar volume size distribution shown in the left panel of Figure 4.13, and fit these to a log-normal volume size distribution

$$\frac{d\mathcal{V}}{d\ln R} = \left(\frac{d\mathcal{V}}{d\ln R} \right)_{\text{fine}} + \left(\frac{d\mathcal{V}}{d\ln R} \right)_{\text{coarse}},$$

where the individual $d\mathcal{V}/d\ln R$'s are given by Eq. (4.15) with the diameter D replaced by the radius R , and V replaced by \mathcal{V} . Here \mathcal{V} is the volume of particles contained in a vertical column of atmosphere with a base of $1 \mu\text{m}^2$, i.e., the units of \mathcal{V} are $\mu\text{m}^3/\mu\text{m}^2$. The right panel of Figure 4.13 compares the average Lanai distribution with the analytical fit. Figure 4.14 compares the fine and coarse modes with the similar components from the Shettle and Fenn models with $RH \leq 50\%$. We see that the modal sizes are similar; however, the Shettle and Fenn distributions are considerably wider, especially the fine mode. It should be noted that the measurements are for *ambient* RH , and the particles are expected to increase in size with increasing RH . Figure 4.13 is for what is considered to be a pure maritime aerosol; however, when compared to the full data set from the three locations, most of the variability is in the relative contributions of the two modes, i.e., the relative values of V_T for the fine and coarse modes. An analysis of the correlation between the size distribution and the wind speed at Midway Island, shows that the modal size of both modes is virtually independent of $\langle w \rangle$, the wind speed averaged over 24 hrs, as is V_T for the

⁹These retrievals yield the *columnar* size distribution. Note that the measurement of $\tau_a(\lambda)$ yields

$$\begin{aligned} \tau_a(\lambda) &= \int_0^\infty c(h, \lambda) dh = \int_0^\infty dh \int_0^\infty \pi r^2 Q_c(\tilde{m}(\lambda, h), \lambda, R) \frac{dN(h, R)}{dR} dR \\ &= \int_0^\infty f(h) dh \int_0^\infty \pi R^2 Q_c(\tilde{m}(\lambda), \lambda, R) \frac{dN(R)}{dR} dR, \end{aligned} \quad (a)$$

where we assume $dN(h, R) = f(h)dN(R)$, i.e., a single size distribution characterizes the entire atmospheric column, and $\tilde{m}(\lambda, h) = \tilde{m}(\lambda)$ – the refractive index is independent of altitude. Note that $\int f(h) dh = H$, the height of the atmospheric column containing *all* of the particles. Upon inverting Eq. (a) to retrieve dN/dR , what we actually end up with is HdN/dR , which we designate as $d\mathcal{N}/dR$, $d\mathcal{N}(R)$ being the number of particles in dR in a column of atmosphere with a base area A . If the units of R are μm then the units of \mathcal{N} are μm^{-2} . Likewise, given \mathcal{N} we represent the volume of particles in the column by \mathcal{V} with units $\mu\text{m}^3/\mu\text{m}^2$.

fine mode. In contrast, the total volume of particles of the coarse mode increases roughly linearly with $\langle w \rangle$ at a rate of $0.0073 \mu\text{m}^3/\mu\text{m}^2$ per m/s of wind.

More recent data from AERONET island stations provide mean values for the aerosol optical thickness $\tau_a(500)$ and the Ångström power p for several of the world's oceans. The global average is $\tau_a(500) = 0.108$ and $p = 0.573$. The Southern Ocean has the clearest atmosphere ($\tau_a(500) = 0.060$ and $p = 0.380$) and lowest variability, while the Atlantic Ocean has the most turbid atmosphere ($\tau_a(500) = 0.190$ and $p = 0.604$) and the highest variability. An important conclusion to be drawn from these studies is that for most of the World oceans, $\tau_a(500) \leq 0.2$, and so with the Atlantic value of p , $\tau_a(865) \leq 0.14$. In a later chapter, where we develop the atmospheric correction algorithm for remotely sensing ocean color, we take the upper limit of τ_a to be 0.20. These results show that $\tau_a(865) \leq 0.20$ for the open ocean is quite conservative.

The Smirnov et al. model leads to optical properties that are similar to those of Shettle and Fenn. The scattering phase function and spectral variation of the extinction coefficient, computed assuming spherical particles with $\tilde{m} = 1.37 - 0.001i$ and the Lanai size distribution are shown and compared with the same quantities for several Shettle and Fenn models in Figures 4.15 and 4.16. The Smirnov et al. model clearly yields optical properties that fall within the range of variability of the Shettle and Fenn models.

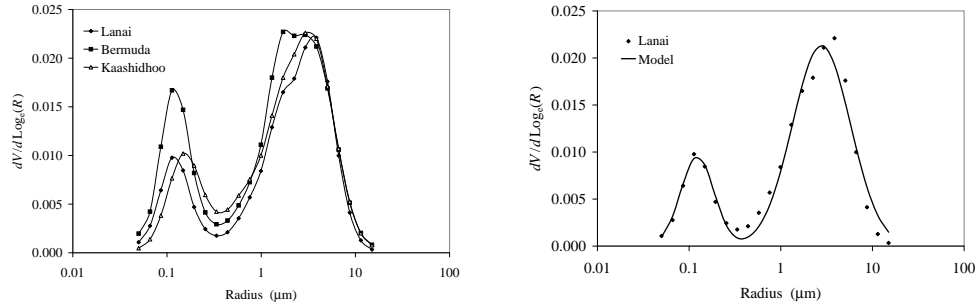


Figure 4.13: Volume size distributions, $dV/d\ln R$ in units of $\mu\text{m}^3/\mu\text{m}^2$, for a pure maritime regime. Left panel: average of measurements at three remote locations. Right panel: comparison between the measurements at the Lanai location and the analytic fit to the sum of two log-normal volume distributions. Note that these plots use radius as the independent variable as opposed to diameter in most of this text.

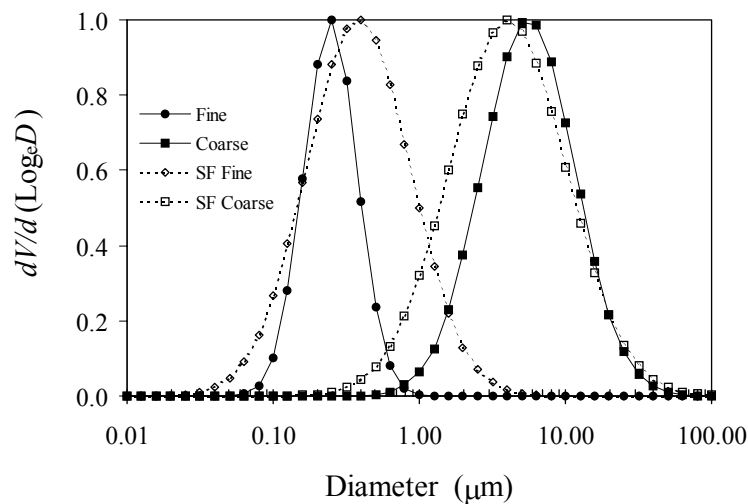


Figure 4.14: Comparison between the Shettle and Fenn [$dV/d\ln D$] and the Smirnov et al. [$dV/d\ln D$] (Lanai) size distributions for the fine and coarse modes. Each mode has been normalized to unity at the size of maximum $dV/d\ln D$. The values of V_T for the fine and coarse modes at Lanai are 0.010 and $0.039 \mu\text{m}^3/\mu\text{m}^2$, respectively.

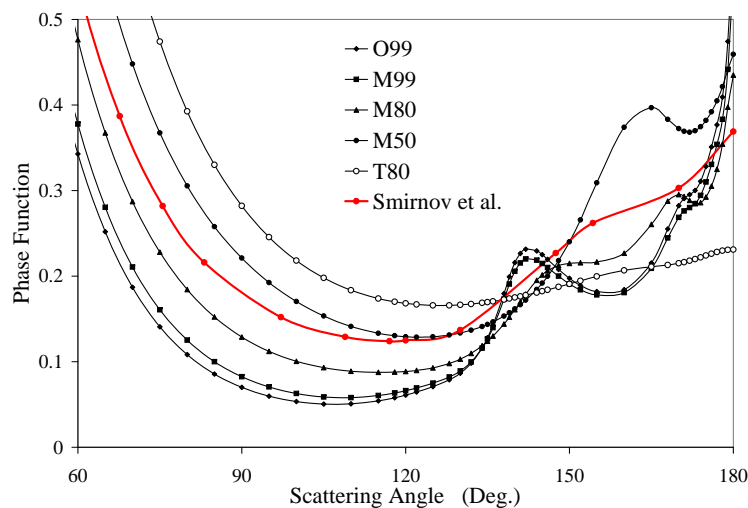


Figure 4.15: Phase function (865 nm) computed using the Lanai size distribution compared with several Shettle and Fenn phase functions.

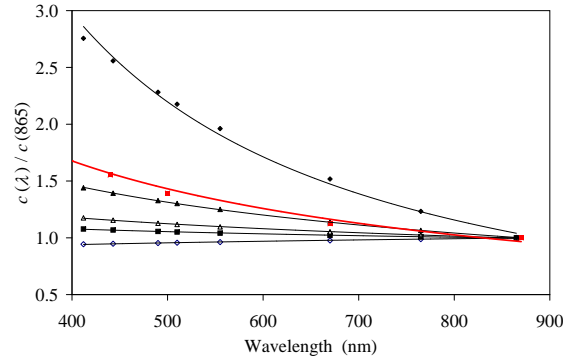


Figure 4.16: Spectral variation of the extinction coefficient for the Lanai size distribution (red) compared with several Shettle and Fenn models: filled diamonds – T80; filled triangles – M50; open triangles – M80; filled squares – M99; open diamonds – O99. The lines are regression fits to $c(\lambda) \propto \lambda^{-p}$. The regression coefficients for the Shettle and Fenn models are given in the caption to Figure 4.9. For the Lanai spectrum, $p = 0.708$ and $R^2 = 0.9806$.

4.3.3.4 Saharan Dust Models

Massive dust storms in desert areas such as the Sahara, Sahel, and Gobi regions, provide a source for large quantities of dust transported (thousands of km) over the oceans by the wind. This dust is important in an optical sense due to very high particle concentrations. In addition, the dust contains absorbing minerals, such that its absorption coefficient increases rapidly from the green portion of the spectrum into the ultraviolet. It is through these two properties – high concentration and strong absorption – that dust influences ocean remote sensing.

The particle size distribution of this dust is generally represented as a combination of three log-normal distributions as in Eq. (5.29). In the models we will discuss, $D_N = 0.002$, 0.044 , and $12.48 \mu\text{m}$, and $\sigma_{10} = 2.13$, 3.20 , and 1.89 for the ultrafine, fine, and coarse models, respectively. The fraction by number in the models labeled “1” for these modes are 54.21 , 45.79 , and $3.9 \times 10^{-5}\%$. We shall also examine two other models where the coarse mode has been increased by factors of 10 (“2”) and 20 (“3”). These latter models are expected to be extant near the source of the dust, and absent far from the source. Being mineral in nature, the real part of \tilde{m} is near 1.5. The imaginary part varies with composition, and for Saharan dust we consider two estimates shown in Figure 4.17. The estimate “Patterson” is based on a survey of measurements of the absorbing properties

of dust, and the estimate “Moulin et al.” is based on comparing the reflectance of dust clouds (as observed with space-borne sensors) with the size distribution described above and various combinations of \tilde{m} and the vertical distribution of the dust. (In contrast to weakly absorbing aerosols, for strongly absorbing aerosols like dust, the reflectance is considerably influenced by its vertical distribution.) For these two refractive indices and the three size distribution models, we have six individual cases: “BDS1”, “BDS2”, and “BDS3”, which use the Patterson m_i along with size distributions “1”, “2”, and “3”; and the corresponding “BDB1”, “BDB2”, and “BDB3”, which use the Moulin et al. m_i . Figure 4.18 provides ω_0 and extinction coefficient spectra and Figure 4.19 provides examples of the phase function for these dust models using 1.53 for m_r . These were computed assuming that the particles are spherical, which is a poor assumption for dust. Of primary importance here is the fact that for the BDS and BDB models, the extinction coefficient spectrum and the scattering phase functions are essentially identical, i.e., they do not depend in any important way on m_i . Note also that these parameters are quite similar to similar parameters for the earlier models: Haze C; Shettle and Fenn; and Smirnov et al. In contrast, ω_0 is strongly dependent on m_i and unlike the Shettle and Fenn models it decreases from the red to the blue rather than increasing (Figure 4.12). Thus, the main difference between the optical properties of dust and other aerosols is the former’s strong decrease in single scattering albedo from near unity in the red to values close to 0.80 in the blue portions of the visible spectrum. As we shall see later, in contrast to the aerosols discussed earlier, for aerosols that are strongly absorbing like dust, the radiance reflected from the ocean-atmosphere system is strongly dependent on their vertical distribution. The presence of dust, which can be mixed to altitude of 5-6 km, considerably complicates the analysis of ocean color as observed from space.

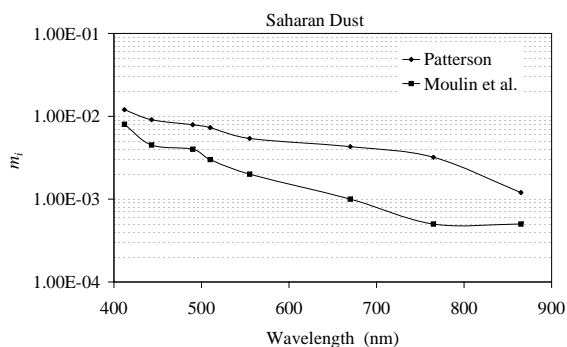


Figure 4.17: Two models of the imaginary part of the refractive index of Saharan dust, as discussed in the text.

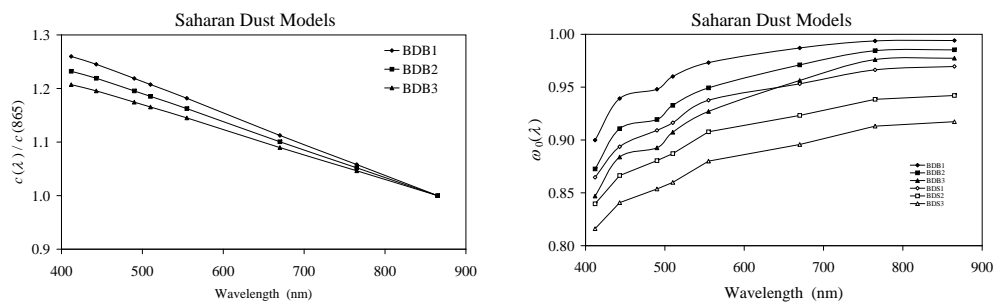


Figure 4.18: Spectra of the extinction coefficient (left) and ω_0 (right) for the dust models described in the text. The extinction spectra for the BDS models (not shown) are virtually identical to the BDB models. The lines are *not* the fits to Angstrom's law as in earlier figures. The extinction coefficients do not follow Angstrom's law as well as the earlier models. Regression gives $p = 0.3109, 0.2812$, and 0.2538 , and $R^2 = 0.9812, 0.9818$, and 0.9824 , for BDB1, BDB2, and BDB3, respectively.

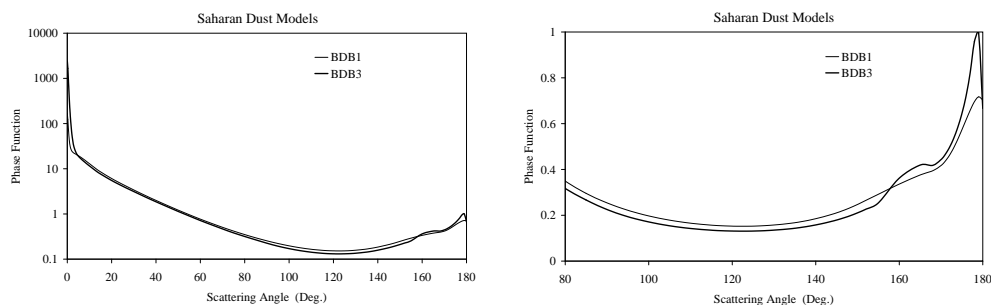


Figure 4.19: Phase functions at $\lambda = 865$ nm for the BDB1 and BDB3 dust models. The associated BDS models are identical to these on the scale of the drawing.

4.3.3.5 Stratospheric Aerosol Models

The models discussed so far are for aerosols in the *troposphere*: the surface to 8-10 km. (More on the vertical distribution of aerosols in the atmosphere later.) However, aerosols also exist in the *stratosphere*: a layer extending from approximately 10 to 50 km. With the exception of a very dilute background stratospheric aerosol (which is of little consequence

to remote sensing), the principal components of the stratospheric aerosol are volcanically injected particles and gases and thin cirrus clouds.

Volcanic eruptions can inject gases and particulate debris to great heights in the atmosphere. The large particles settle out quickly; however, the smaller particles as well as the aerosol created by condensing gases mixed with water vapor (principally H_2SO_4) can remain in the stratosphere for a few years, and generally spread over the entire globe. Global mean stratospheric optical thicknesses (550 nm) from a strong eruption are in the range 0.05–0.15, comparable to the optical thickness of a *clean* maritime atmosphere (Section @.3.3.3). A well-studied example is El Chichon (Chiapas, Mexico) which erupted March 28 to April 4, 1982. Three months after the eruption, spectra of aerosol optical depth at the Mauna Loa Observatory (Hawaii, altitude 3397 m) showed a maximum of about 0.25 at 500 nm. Although the Observatory is actually in the troposphere, as we shall see in the (pure) maritime environment, most of the aerosol is near the surface (altitude ~ 0 –1 km), so the measurement contains very little tropospheric aerosol. Note that this optical depth is much higher than the global averages mentioned above, as the aerosol had yet to spread globally. Inversion of the optical depth spectrum, using methods similar to those described in Sections @.3.1 and @.3.3.3, and assuming spherical particles of composition 25% water and 75% H_2SO_4 , found that the columnar particle size distribution fit the modified gamma distribution:

$$\frac{d\mathcal{N}}{d\ln R} = CR^{(1-3b)/b} \exp[-R/(ab)],$$

where R is the particle radius, rather well for $a = 0.4 \mu\text{m}$ and $b = 0.064$.¹⁰ Using this size distribution, the refractive index of the $\text{H}_2\text{O} - \text{H}_2\text{SO}_4$ mixture ($\sim 1.42 - 0.0i$), and Mie theory, yields the extinction spectrum and the scattering phase functions provided in Figure 4.20. Of course, the extinction spectrum was already known, as it was used to derive the size distribution; however, the phase function was not measured, and in a very direct sense, is a product of the measurement of the extinction spectrum. Note the departure of the extinction spectrum from Angstrom’s law, even to the extent of *decreasing* extinction with *decreasing* wavelength for $\lambda \lesssim 550$ nm. This effect is due to the narrowness of the size distribution which also causes the considerable variation of the phase function with λ for $\Theta \gtrsim 120^\circ$.

Cirrus clouds are composed of ice particles, which in general have the shape of hexagonal cylinders that are large compared to the wavelength. Their scattering can be computed

¹⁰If \mathcal{N}_0 is the total number of particles (in the column above the observations), then

$$C = \mathcal{N}_0 \frac{(ab)^{-x}}{\Gamma(x)},$$

where $x = (1 - 2b)/b$ and Γ is the gamma function. The modal radius $R_M = (1 - 2b)a$, which for the retrieved parameters is $0.35 \mu\text{m}$. For the modified gamma distribution, $R_{\text{eff}} = a$ and $v_{\text{eff}} = b$.

from a combination of geometrical optics and diffraction theory (Chapter 3). Assuming the particles are in random orientation and $\tilde{m} \sim 1.31 - 0.0i$, the resulting phase function is shown in Figure 4.21. The sharp maxima at $\Theta \sim 22^\circ$ and 46° correspond to the light entering one face of the hexagon and exiting another face (22°), and the light entering one face of the hexagon and exiting through the base of the column (46°). They give rise to often-observed atmospheric phenomena such as Sun dogs, Sun halos, etc. Because the particles are large compared to the wavelength, the extinction is essentially independent of wavelength over the visible and NIR portions of the spectrum. This observation and Figure 4.21 can be considered first-order approximations for examining the effects of thin cirrus clouds on radiative transfer in the atmosphere.

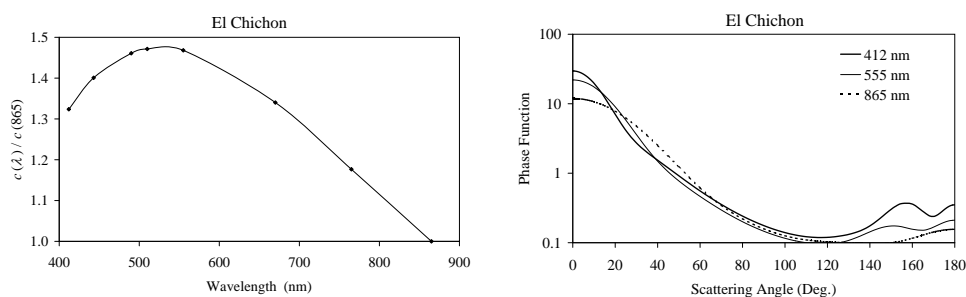


Figure 4.20: Spectral variation of the extinction coefficient (left) and the phase function (right) for the El Chichon aerosol.

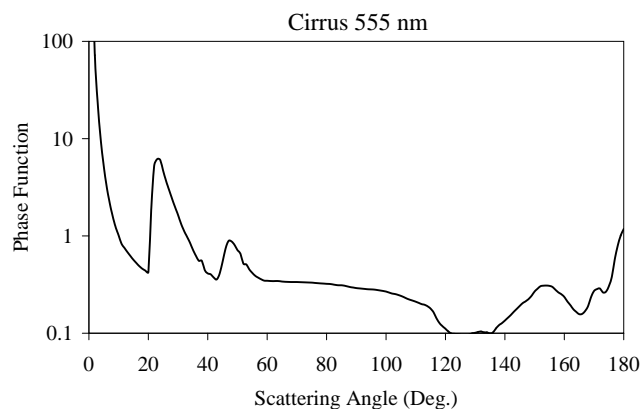


Figure 4.21: Scattering phase function for cirrus clouds at 555 nm.

4.3.4 Aerosol Vertical Structure

As suggested in the last section, aerosols exist at all altitudes in the atmosphere from the surface to the stratosphere. However, it is important for us to understand how the bulk of the aerosol is typically distributed with altitude. Although measurements have been made from aircraft and from ship- and island-based LIDAR (Light Detection and Ranging) instruments, for our purposes we will discuss data provided by the Lidar-In-space Technology Experiment (LITE). LITE was a three wavelength LIDAR (355, 532, and 1064 nm) flown on the Space Shuttle (STS-64), September 9-20, 1994. Light pulses emitted from the laser enter the top of the atmosphere and are backscattered from air molecules and from aerosols (and clouds). The time delay between the initial pulse and the return provide the distance from the LIDAR instrument and the scatterer, or equivalently the distance from the scatter to the sea surface.¹¹

For examples of the vertical distribution of aerosols over the oceans, we show samples of data from several LITE orbits over the North Atlantic Ocean. The LITE orbits to be examined are shown in Figure 4.22. We will discuss orbits 115, 116, and 117. All of these orbits occurred during the time of a Saharan dust outbreak, and the portions of the orbits colored in red are over regions containing dust. The northern portion of orbit 115 (colored black) is free of dust and characteristic of a clean maritime environment. Figure 4.23 shows the LIDAR return from the northern portion of orbit 115. What is shown in the figure is proportional to the returned power as a function of altitude (see color scale at the top). At an altitude of 20 km almost all of the return is due to Rayleigh scattering by the air and background stratospheric aerosol. At this period of time there was little volcanic aerosol in the stratosphere. The Rayleigh return (proportional to the atmospheric pressure, Eq. (4.8)) increases with decreasing altitude; hence, the gradual change in color from deep blue to a pale blue progressing downward from 20 to 5 km. The backscattered return increases dramatically at altitudes below about 1 km (the height of the marine boundary layer). In this layer the return is principally from clouds and aerosols. The cloud return is white (signal > 3203) in the figure (very strong backscattering with little return from below the

¹¹Let the laser source be located at an altitude H above the sea surface. Then, given the definition of the volume scattering function in Chapter 2 ($\beta(\Theta) = dJ(\Theta)/E_0 dV$), the power $dP_b(z)$ backscattered from a layer of thickness dz located at an altitude z above the sea surface is easily found to be

$$dP_b(z) = \frac{\beta(180^\circ) \exp[-2 \int_z^{H-z} c(z) dz]}{(H-z)^2} A_R P_L dz,$$

where A_R is the area of the receiver and P_L is the power emitted by the laser. The exponential term accounts for the attenuation of the laser pulse on its way to the target and the return signal from the target to the receiver. Now, the laser pulse emitted at time t_0 will be received at time $t = 2(H-z)c_\ell$, where c_ℓ is the speed of light. Therefore, $dz = 2c_\ell |dt|$, where $|dt|$ is the time interval over which the return is measured, so if the scattered return is broken into small time intervals, and $c(z)$ is not too large, this provides a direct measurement of $\beta(180^\circ)$ as a function of z .

clouds) and the aerosol is in the return power range of 985-1791. This figure shows that in a clear maritime regime most of the aerosol is located in the marine boundary layer (altitude $\lesssim 1$ km. In contrast, Figure 4.24 shows the return from the portions of the three orbits that contain Saharan dust. For orbit 115, the returns from latitudes greater than about 20° originate over water, while those from lower latitudes are over land. Note the particularly strong return from altitudes between about 1 and 5 km. This is from the dust (close to the source). There is essentially no LIDAR return from below about 10 km at the extreme right of the image, with almost all of the return coming from about 14 km. Here clouds exist that are almost opaque to the laser and to its return. This is a characteristic of optically thick clouds. Thin cirrus clouds are not optically thick and there is return from below the cloud. An example of thin cirrus is the strong return at a latitude of $\sim 17.6^\circ$ where the backscattering from below is clearly unaffected. The returns from orbits 116 and 117 show the spread of the dust across the Atlantic. For most of its journey across the ocean the dust is in a layer between the marine boundary layer and about 5 km.

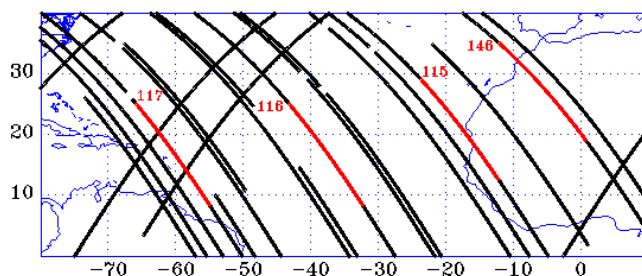


Figure 4.22: Orbital map for STS-64 carrying the LITE instrument. In the following figures we examine data from orbits 115, 116, and 117. From LITE web site: https://www-lite.larc.nasa.gov/n_saharan_dust.html.

The LITE data are characteristic of aerosol vertical structure over oceans. In clean regimes, where the aerosols are in large part locally generated, they are generally confined to the marine boundary layer, while aerosols like dust that is generated over land and transported over the water by the winds are usually in a layer above the boundary layer and below the stratosphere. The latter also includes aerosols derived from polluted urban air. Dust and urban aerosols are generally strongly absorbing in some or all of the visible spectrum. Cirrus clouds are generally in the stratosphere. These LITE data suggest that for situations in which viewing the sea surface from space is possible (i.e., in the absence of cloud cover), and in the absence of volcanic aerosol, dust or urban pollution, for optical purposes the atmosphere can be modeled as a 2-layer system, with aerosols and

air (Rayleigh scattering) in the lower layer and air alone in the upper layer. If volcanic aerosols, dust (and/or urban pollution), and cirrus clouds need to be taken into account, then a 3-layer model would be appropriate with air and dust (and/or urban pollution) in a middle layer and air and cirrus and/or volcanic aerosol in the top layer.

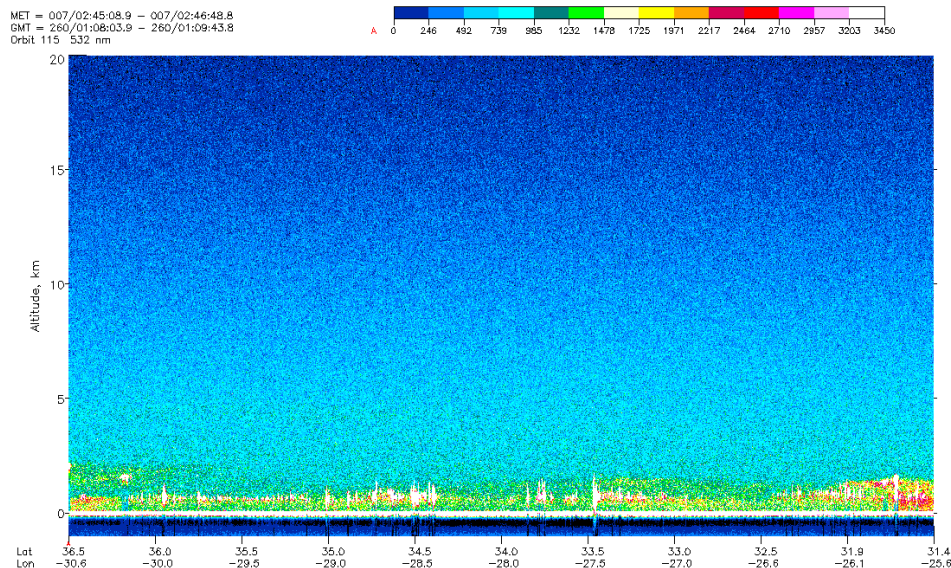


Figure 4.23: LITE backscattering return for the northern portion of orbit 115 over the open ocean: The horizontal axis is latitude and longitude. From the left to the right, latitude goes from 36.5°N to 31.4°N, and longitude from 30.6°W to 25.6°W in 0.5°increments. The vertical axis is the altitude of the return from 0 to 20 km in increments of 5 km. The color scale at the top provides a measure of the intensity of the backscattered return. From LITE web site: https://www-lite.larc.nasa.gov/n_saharan_dust.html.

4.4 Absorption by Gases in the Atmosphere

We now describe the absorption of gases in the atmosphere. As we shall see, gaseous absorption is usually confined to narrow (compared to the whole visible spectrum), well-defined spectral regions, which are to be avoided in remote sensing systems designed to view water bodies. Thus, gaseous absorption is less important than gaseous scattering or aerosol scattering and absorption. Exceptions are NO₂ and O₃. For these, the absorption

is more broad band, but fortunately consists of narrow, weakly absorbing spectral features, which for a typical remote-sensing spectral band, is manifest as a simple, small increase in the total optical thickness of the atmosphere in the band.

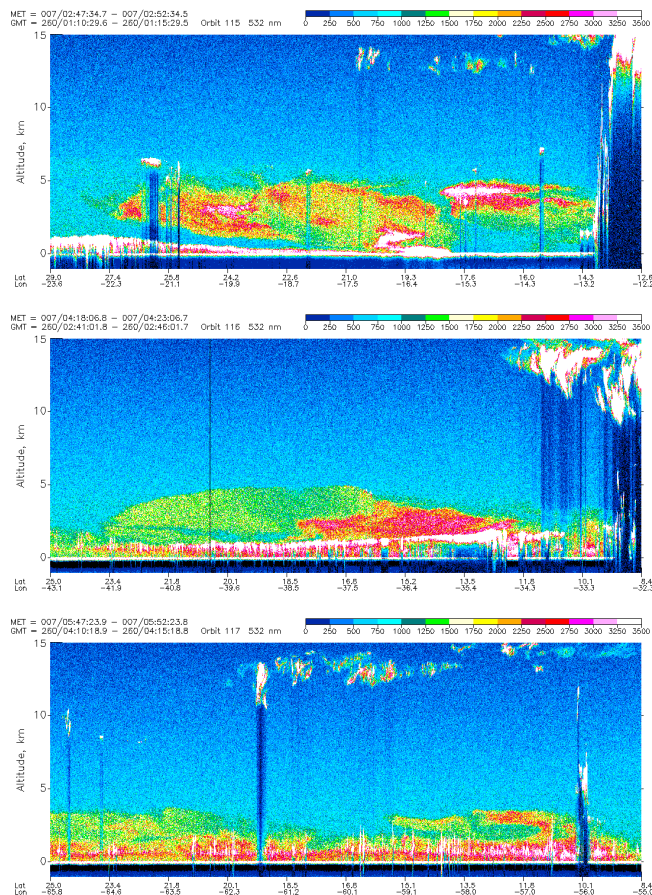


Figure 4.24: LITE backscattering return for the portions of orbit 115, 116, and 117 labeled in red in Figure 4.22: top orbit 115; middle orbit 116; bottom orbit 117. The horizontal axis is latitude and longitude. From the left to the right, latitude goes from 29.0°N to 24.6°N , and longitude from 23.6°W to 12.2°W , for orbit 115; latitude from 25.0°N to 8.4°N , and longitude from 43.1°W to 32.3°W for orbit 116; and latitude from 25.0°N to 8.4°N , and longitude from 65.8°W to 55.0°W , for orbit 117. The vertical axes are the altitude of the returns from 0 to 15 km in increments of 5 km. The color scale at the top provides a measure of the intensity of the backscattered return. From LITE web site: https://www-lite.larc.nasa.gov/n_saharan_dust.html.

4.4.1 General Properties of Gaseous Absorption.

In Chapter 1 we found that for radiation traversing a medium the imaginary part (absorptive part) of the refractive index for a transition from a single quantum state i to a single state j was given by (Eq. (1.31))

$$m_i = \frac{1}{2\epsilon_0} \frac{f_{ij}\gamma_{ij}\omega}{(\omega_{ij}^2 - \omega^2)^2 + (\gamma_{ij}\omega)^2} (n_i - n_j),$$

where n_i and n_j are the number densities of atoms (molecules) in the lower i and upper j states, respectively, and $\omega_{ij} = 2\pi(e_j - e_i)/h$, where e_j and e_i are the energies of the two states with $e_j > e_i$. Note here that i and j refer to the full set of quantum numbers needed to specify each state. The absorption coefficient a was (Eq. (1.26))

$$a(\omega) = \frac{4\pi m_i}{\lambda_v} = \frac{2\omega m_i}{c_\ell},$$

where c_ℓ is the speed of light, so

$$a(\omega) = \frac{1}{2c_\ell\epsilon_0} \frac{f_{ij}\gamma_{ij}\omega^2}{(\omega_{ij}^2 - \omega^2)^2 + (\gamma_{ij}\omega)^2} (n_i - n_j).$$

Note that a is for the given transition $i \rightarrow j$ only. Absorption features such as those described by the equation above are referred to as *absorption lines*, the term originating from the fact that in spectrographs, light enters through a slit and, after traversing a dispersing element (prism or diffraction grating), appears as a “line” (the image of the slit) on a photographic plate placed in the focal plane of the instrument.

From statistical thermodynamics we know that at thermal equilibrium $n_i = C \exp[-e_i/kT]$, where C is a constant. If n is the total number of atoms (molecules) per unit volume, then

$$n_i - n_j = n \left(\exp[-e_i/kT] - \exp[-e_j/kT] \right) / Z,$$

where Z is the partition function:¹²

$$Z = \sum_{\text{All } i} \exp[-e_i/kT].$$

¹²This is easy to show. Since $n_i = C \exp[-e_i/kT]$, and $n_i/n = \mathcal{P}(n_i)$, the probability of n_i , we have

$$\sum_{\text{All } i} \mathcal{P}(n_i) = 1 = \frac{C}{n} \sum_{\text{All } i} \exp[-e_i/kT] = \frac{C}{n} Z.$$

Thus $C = n/Z$ and the result follows.

Unless $\omega \approx \omega_{ij}$, $a(\omega)$ is small, so we can use the approximation

$$(\omega_{ij}^2 - \omega^2) = (\omega_{ij} - \omega)(\omega_{ij} + \omega) \approx 2\omega_{ij}(\omega_{ij} - \omega)$$

to find

$$a(\omega) = \frac{1}{2c\epsilon_0} \frac{f_{ij}/\gamma_{ij}}{X^2(\omega) + 1} \frac{(\exp[-e_i/kT] - \exp[-e_j/kT])}{Z} n, \quad (4.18)$$

where $X(\omega) = 2(\omega_{ij} - \omega)/\gamma_{ij}$. It is important to note that $a(\omega) \propto n$, in fact, so important that we define the absorption cross section, $\sigma_a(\omega)$, such that $a(\omega) \triangleq \sigma_a(\omega)n$, and so

$$\sigma_a(\omega) = \frac{1}{2c\epsilon_0} \frac{f_{ij}/\gamma_{ij}}{X^2(\omega) + 1} \frac{(\exp[-e_i/kT] - \exp[-e_j/kT])}{Z}.$$

Again, note that $\sigma_a(\omega)$ as defined here refers only to the transition from quantum state i to quantum state j , and is a *property of the atom (molecule) alone*, independent of the number of gas atoms (molecules) present.

Consider an absorption cell of length ℓ containing the gas (uniform density). Then the transmittance of a beam of radiation of angular frequency ω (close to ω_{ij}) through the cell is given by

$$T(\omega) = \exp[-a(\omega)\ell] = \exp[-\sigma_a(\omega)n\ell] \triangleq \exp[-\tau_{\text{abs}}(\omega)],$$

where $\tau_{\text{abs}}(\omega)$ is the absorption optical thickness or the absorption optical depth at angular frequency ω . Thus the absorption optical thickness τ_{abs} is the absorption cross section ($\text{m}^2/\text{molecule}$) times $n\ell$ ($\text{molecules}/\text{m}^2$). If the beam has a cross sectional area A , then $n\ell A$ is the the number of molecules traversed by the beam.

As there are many different (and sometimes bewildering) alternate ways of describing the quantities that go into the determination of τ_{abs} , it is worthwhile to digress and examine some of the more common ones. Often the cross section, rather than being specific to one molecule, is specific to the mass, i.e., $\sigma_a^* \triangleq \sigma_a/m$ (m^2/kg), where m is the mass of an individual molecule of the gas. Defining $n^* = mn$ (the actual mass density of the gas in kg/m^3) the optical thickness in this case is $\sigma_a^* n^* \ell$, with ℓ in meters. In determining the absorption properties of gases, spectroscopists usually measure the transmittance of gas in an absorption cell of length ℓ . Treating the gas in the cell as ideal, $n = P/kT$, so $n\ell = P\ell/kT$. It is common for them to measure P in atmospheres and ℓ in meters, in which case a given amount of gas is referred to by $P\ell$ in units of meter-atmospheres, with a reference temperature specified. Then $\sigma_a^{**} \triangleq \sigma_a/kT$ is specified in units meter-atmospheres $^{-1}$, and the absorption optical thickness is $\sigma_a^{**} P\ell$. The number density of an ideal gas at a pressure of 1 atmosphere and a temperature of 273°K is $101325 \text{ N}/\text{m}^2 / (1.38 \times 10^{-23} \text{ J}/\text{molecule}^\circ\text{K} \times 273^\circ\text{K}) = 2.69 \times 10^{25} \text{ molecules}/\text{m}^3$. Thus, measuring ℓ in meters, 1 meter-atmosphere is equivalent to $n\ell = 2.69 \times 10^{25} \ell \text{ molecules}/\text{m}^2$ at atmospheric pressure

and a temperature of 273°K. Also, a beam of cross sectional area A (in m^2) will traverse $n\ell A = 2.69 \times 10^{25} \ell A$ molecules. If all lengths are measured in cm, then $n\ell A = 2.69 \times 10^{19} \ell A$ molecules, which corresponds to an amount of gas equivalent to 1 cm-atmosphere, i.e., $n = 2.69 \times 10^{19}$ molecules/ cm^3 when $P\ell = 1$ cm-atm. One way of specifying gas abundances in the atmosphere is to imagine all of the gas molecules in an atmospheric column to be brought to the Earth's surface at atmospheric pressure and a temperature of 273°K. The resulting volume of gas molecules would have a height H in cm, and the gas abundance would be specified as H cm-atmospheres. A familiar unit of abundance for those who study Ozone in the atmosphere is the Dobson Unit (DU): 1 DU = 10^{-3} cm-atmospheres and is equivalent to 2.69×10^{16} molecules/ cm^2 . The abundance of Ozone in the atmosphere is ~ 200 -400 DU.

It has been assumed in most of this section that the gas has a uniform number density; however, in the atmosphere this is not true. When the density is non-uniform, it is useful to explicitly note that it is a differential quantity, i.e., $n = d\eta/dV$, where $d\eta$ is the number of molecules in a small volume dV . Thus, in an inhomogeneous medium, where the path through the medium is from ℓ_1 to ℓ_2 , the optical depth is

$$\tau_{\text{abs}}(\omega) = \sigma_a(\omega) \int_{\ell_1}^{\ell_2} \frac{d\eta(\ell)}{dV} d\ell = \sigma_a \int_{\ell_1}^{\ell_2} n(\ell) d\ell,$$

or $d\tau_{\text{abs}} = \sigma_a n(\ell) d\ell$. The total number of molecules (N_{Total}) in a vertical column of air with base area A is

$$\int_0^\infty \frac{d\eta(h)}{dV} dh = \frac{1}{A} \int_0^\infty \frac{d\eta(h)}{dh} dh = \frac{1}{A} \int_0^\infty d\eta(h) = \frac{N_{\text{Total}}}{A},$$

where h is the altitude.

We have yet to discuss the line width γ_{ij} . Note from Eq. (4.18) that $a(\omega)$ is reduced to half of its value at the absorption maximum ($\omega = \omega_{ij}$, or $X(\omega) = 0$) when $X(\omega) = \pm 1$, i.e., when $\omega - \omega_{ij} = \gamma_{ij}/2$. Thus, the *full width at half maximum* (FWHM) of the absorption profile is γ_{ij} . This width arises in the classical theory in the following manner. An atom excited by an electric field becomes a dipole oscillating with same frequency as the field. An oscillating dipole loses energy through radiation and thus the oscillations are damped. The line width is related to the damping constant (Chapter 1), which is given by

$$\gamma = \frac{2}{3} \frac{1}{4\pi\epsilon_0} \frac{q^2 \omega_0^2}{mc^3}$$

and for an electron has the value $\gamma \approx 6.3 \times 10^{-24} \times \omega_0^2$, where ω_0 is the natural frequency of the oscillator. Near the center of the visible spectrum (500 nm), $\gamma \approx 8.9 \times 10^7 \text{ s}^{-1}$. The oscillations of the unforced oscillator decay to $\exp[-1]$ of their initial energy in a time

$\gamma\Delta t = 1$. This value of Δt is called the natural lifetime of the oscillation and, again, for an electron near the enter of the visible spectrum, $\Delta t \sim 1.1 \times 10^{-8}$ s. Here, $\gamma_{ij} = \gamma$ and $\omega_{ij} = \omega_0$.

In the quantum theory, the transition from $i \rightarrow j$ takes place over a finite amount of time Δt . By virtue of the Heisenberg uncertainty principle, this time interval leads to a spread in energy difference $e_j - e_i$, $\Delta e \sim h/(2\pi\Delta t)$, which in turn leads to a spread in $\Delta\omega_{ij} \sim 1/\Delta t$, i.e., the transition from i to j is smeared out over an angular frequency range of $\sim 1/\Delta t$. The quantity $\Delta\omega_{ij}$, is also called the *natural line width*. Comparing this with γ_{ij} , we see that, as in the classical theory, $\gamma_{ij} \sim 1/\Delta t$, and is a property of the individual atom (molecule) alone, independent of its environment.

The situation is different when the gas molecules are at a high enough density that there are collisions between individual molecules. Collisions during the process of absorption and/or emission of radiation also cause shifts in the energy levels which in turn leads to a further broadening of the absorption profiles (think of it as a broadening of the energy levels). Typically, the broadening associated with collisions is much larger than the natural line width, γ_{ij} , and of course depends on the environmental conditions. The question of the broadening of absorption features by collisions is a very difficult and not yet totally solved problem of quantum physics. In the following, we try to provide simple classical arguments to suggest how the width of absorption features might depend on environmental conditions.

Consider an atom (molecule) moving with speed v through the medium. Fix the position of all of the other atoms. If each atom is taken to be a sphere with diameter D , then the closest that two atoms can pass one another without colliding is $2D$. Now fix all of the other atoms, except the one in question, in position and shrink them to zero size, while doubling the diameter of the one in question. This atom will collide with the fixed ones at the same rate that it would if they all had the same diameter D . In a time interval δt it will sweep out a volume $V = \pi D^2 v \delta t$. This volume will contain nV atoms. In time δt there will be $nv\pi D^2 \delta t$ collisions. Thus, the average collision rate will be $nv\pi D^2$, or the average time between collisions will be $(nv\pi D^2)^{-1}$. Assume the atom starts to absorb radiation just after one collision has taken place. Then, on the average, there will be another collision within a time interval $\Delta t = (nv\pi D^2)^{-1}$. During collisions, the energies e_i and e_j are changed, and for a given radiation frequency ω the absorption is disrupted. As with the uncertainty principle above, this results in a spectral width $\gamma \sim 1/\Delta t = nv\pi D^2$. According to the equipartition of energy theorem (the average translational kinetic energy of an atom or molecule is $3kT/2$), $\langle v^2 \rangle = 3kT/m$, so the average speed is of the order of $\sqrt{3kT/m}$. Also, $n = P/kT$, so $\gamma \sim P/\sqrt{T}$. If γ is measured at P_0 and T_0 yielding γ_0 , then $\gamma = \gamma_0 \sqrt{T_0/T} P/P_0$. Because γ is proportional to pressure, this type of broadening is usually referred to as “pressure broadening.”

There is another source of line broadening: Doppler broadening. Atoms (molecules) in a gas are in continuous motion. Those moving toward an observer will emit a higher frequency of radiation than those stationary relative to the observer by an amount $\omega_{ij}v/c_\ell$, where v is the component of the atom's velocity in the direction of the observer. This is the *Doppler effect*. Likewise those moving away from the observer will emit a lower frequency. The same happens for absorption: atoms moving toward the light source will have their ω_{ij} 's increased, while the opposite is for those moving away. Since all of the atoms are in random motion, the emission and absorption lines will be broadened by the Doppler effect. Since the Doppler effect depends on the atom's velocity, and the average speed is $\propto \sqrt{T}$, the line width is also proportional to the square root of temperature. This broadening is relatively unimportant in the atmosphere because the temperature is low and the variation in temperature is small.

Finally, in most remote sensing situations, the scene is viewed through an optical filter of some type which passes a band of radiation of frequency $\Delta\omega_{rs}$, that is much greater than FWHM of any gaseous absorption line in the atmosphere. Thus, it is important to know how the transmittance of a path through the gas varies when viewed through such a filter. The mean transmittance can be defined as

$$\langle T \rangle \triangleq \frac{1}{\Delta\omega_{rs}} \int_{\Delta\omega_{rs}} T(\omega) d\omega = \frac{1}{\Delta\omega_{rs}} \int_{\Delta\omega_{rs}} \exp[-\sigma_a(\omega)n\ell] d\omega,$$

with

$$\sigma_a(\omega)n\ell = \frac{1}{2c\epsilon_0} \frac{f_{ij}/\gamma_{ij}}{X^2(\omega) + 1} \frac{(\exp[-e_i/kT] - \exp[-e_j/kT])}{Z} n\ell \triangleq \frac{\tau_c}{X^2(\omega) + 1},$$

where τ_c is the optical thickness at the center ($\omega = \omega_{ij}$) of the absorption line: $\tau_c = \sigma_a(\omega_{ij})n\ell$. Then,

$$\langle T \rangle = \frac{1}{\Delta\omega_{rs}} \int_{\Delta\omega_{rs}} \exp[-\tau_c/(X^2(\omega) + 1)] d\omega = \frac{\gamma_{ij}}{2\Delta\omega_{rs}} \int_{\Delta\omega_{rs}} \exp[-\tau_c/(X^2 + 1)] dX$$

This integral can be evaluated numerically, but it is useful to have some analytical expressions in certain limits. To effect this, one would like to extend the integration limits from $\Delta\omega_{rs}$ to $\pm\infty$; unfortunately, if this is done the integral diverges at both limits (because the integrand $\rightarrow 1$ as $X \rightarrow \pm\infty$). This can be remedied by removing the divergences through computing the average absorptance $\langle A \rangle = \langle 1 - T \rangle$, for which the integral is convergent when the limits are extended:

$$\langle A \rangle = \frac{\gamma_{ij}}{2\Delta\omega_{rs}} \int_{-\infty}^{\infty} \left(1 - \exp \left[-\frac{\tau_c}{(X^2 + 1)} \right] \right) dX. \quad (4.19)$$

We will examine two limits: small τ_c and large τ_c .

For small τ_c the exponential can be expanded and keeping the first nonzero term results in

$$\langle A \rangle \rightarrow \frac{\gamma_{ij}}{2\Delta\omega_{rs}} \int_{-\infty}^{\infty} \frac{\tau_c}{(X^2 + 1)} dX = \frac{\pi\gamma_{ij}}{2\Delta\omega_{rs}} \tau_c. \quad (4.20)$$

Now, we can define an effective optical depth $\langle \tau_{\text{abs}} \rangle$ for the observation passband $\Delta\omega_{rs}$ through

$$\langle A \rangle = 1 - \langle T \rangle = 1 - \exp[-\langle \tau_{\text{abs}} \rangle] \approx \langle \tau_{\text{abs}} \rangle,$$

with the last step following from the fact that if $\langle A \rangle$ is small, $\langle T \rangle$ must be close to unity. Thus, the final result is that when τ_c is small,

$$\langle \tau_{\text{abs}} \rangle = \frac{\pi\gamma_{ij}}{2\Delta\omega_{rs}} \tau_c = \frac{\pi\gamma_{ij}}{2\Delta\omega_{rs}} \sigma(\omega_{ij}) n\ell \quad (4.21)$$

so the effective optical depth $\langle \tau_{\text{abs}} \rangle$ is proportional to $n\ell$. This is very important as it shows that for weak absorption lines, the broad-band transmittance is an exponential function of ℓ or $\langle \tau_{\text{abs}} \rangle$ (see below).

For large τ_c (strong absorption) Eq. (4.19) still holds, but now in $X^2 + 1$ the “1” is of consequence compared to X only over a negligible portion of the integral. This, we can drop the “1” with little error and get

$$\langle A \rangle \rightarrow \frac{\gamma_{ij}}{2\Delta\omega_{rs}} \int_{-\infty}^{\infty} \left(1 - \exp \left[-\frac{\tau_c}{(X^2)} \right] \right) dX = \frac{\gamma_{ij}}{2\Delta\omega_{rs}} \sqrt{\tau_c} \int_0^{\infty} \frac{(1 - \exp[-y])}{y^{(3/2)}} dy,$$

where $y = \tau_c/X^2$. The value of the integral is $2\sqrt{\pi}$, so

$$\langle A \rangle = (\gamma_{ij}\sqrt{\pi}/\Delta\omega_{rs})\sqrt{\tau_c}, \quad \text{and} \quad \langle T \rangle = 1 - \frac{\gamma_{ij}}{\Delta\omega_{rs}}\sqrt{\pi\tau_c}. \quad (4.22)$$

(Note that the unphysical fact that it appears that $\langle T \rangle$ can be negative is an artifact of the extension of the limits to $\pm\infty$.) Notice that now it doesn't even make sense to define an effective absorption optical depth, the transmittance is no longer exponential in τ_c .

The broad-band absorption coefficient would normally be defined as

$$a_{rs} = -\frac{1}{\langle T \rangle} \frac{d\langle T \rangle}{d\ell} = \frac{1}{1 - \langle A \rangle} \frac{d\langle A \rangle}{d\ell}.$$

In the weak absorption limit, this is

$$a_{rs} = \frac{\pi\gamma_{ij}\sigma_a(\omega_{ij})n/(2\Delta\omega_{rs})}{1 - \pi\gamma_{ij}\sigma_a(\omega_{ij})n\ell/(2\Delta\omega_{rs})} \rightarrow \frac{\pi\gamma_{ij}}{2\Delta\omega_{rs}}\sigma_a(\omega_{ij})n,$$

which is constant for a given n . In the strong absorption limit,

$$a_{rs} = \frac{1}{2\ell} \left[\frac{\langle A \rangle}{1 - \langle A \rangle} \right] = \frac{1}{2\ell} [\langle A \rangle + \langle A \rangle^2 \dots],$$

which depends on powers of \sqrt{n} and $\sqrt{\ell}$. Thus, in the weak absorption limit it is possible to define a broad-band absorption coefficient, which depends linearly on n , while in the strong absorption limit this is not possible; however, even in the strong absorption limit one can define a meaningful absorption coefficient if small enough spectral intervals ($\delta\omega \ll \gamma_{ij}$) are chosen, i.e., sufficiently small spectral regions in the absorption line still undergo exponential absorption. In the case of weak absorption, the broadband radiance will still be governed by the radiative transfer equation developed in Chapter 2, with τ_{abs} given by Eq. (4.21). In the strong absorption limit this is no longer the case, and the radiative transfer equation must be solved individually for each increment in frequency $\delta\omega \ll \gamma_{ij}$. This underscores the difficulty encountered in radiative transfer in spectral regions with strong absorption lines. Fortunately, we encounter such situations in ocean remote sensing infrequently.

4.4.2 Vertical Distribution of Principal Absorbing Gases in the Atmosphere

The principal absorbing gases in the atmosphere of importance to ocean color remote sensing are H_2O , NO_2 , O_3 and O_2 . Water vapor is important because it is a strong absorber in the visible and NIR regions and thus must be avoided. The other gases all absorb in spectral regions in which the ocean is viewed and careful account must be taken of their absorption. Figure 4.25 provides typical profiles of these gases as a function of altitude according to the U.S. Standard Atmosphere. Note that the O_2 density is $\sim 21\%$ of the “Total” or “Density” in the top left and right panels, respectively. The “Tropical” and “Subarctic Winter” are provided to suggest a seasonal/geographical range of variation. Note that the water vapor is concentrated in the troposphere, while the Ozone is concentrated in the stratosphere. This reflects the distribution of their sources. In contrast, NO_2 is more uniformly distributed with significant concentrations in both the troposphere and the stratosphere. A principal source of NO_2 in the troposphere is pollution, and this will be the source that is most interesting for our purposes.

4.4.3 Absorption Spectra of Atmospheric Gases

All of the absorbing gases in the visible and NIR are molecular rather than atomic. To lowest order, the individual energy levels (states) of molecules can be divided into rotational states, vibrational states, and electronic states. The separation between adjacent energy states of any of these components Δe_x , where $x = \text{Rot}$ (rotational), Vib (vibrational) and

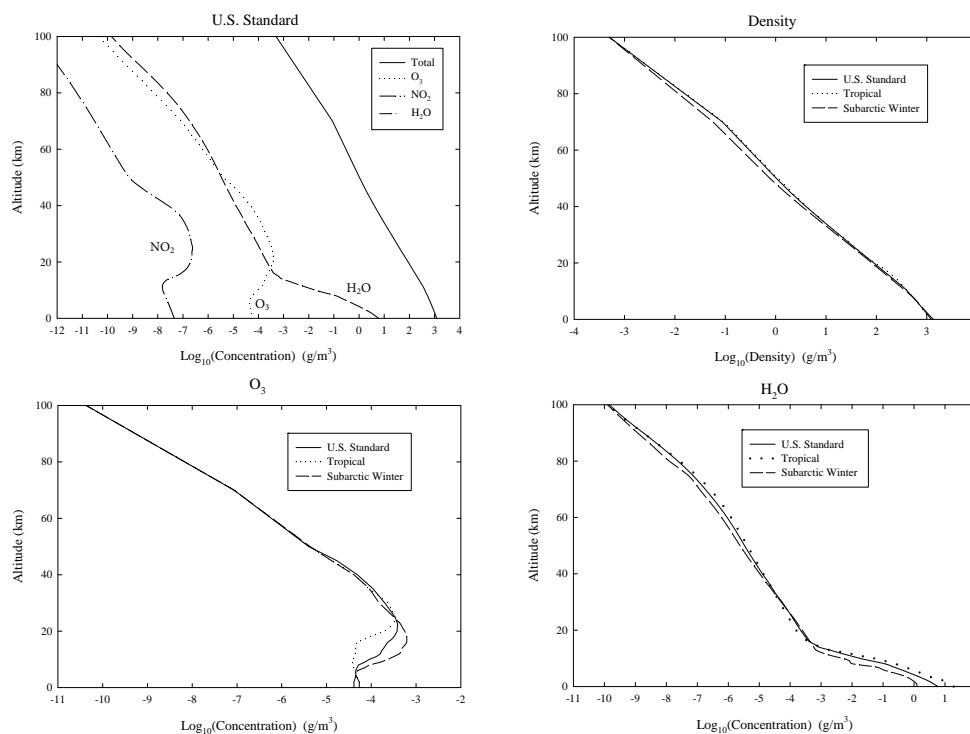


Figure 4.25: Concentration profiles of various gases in the atmosphere (upper left) under average conditions. Variation of the concentration of air (upper right), Ozone (lower left) and water vapor (lower right) typically observed in the atmosphere. For the curves in the lower two figures, the total column concentrations of Ozone (left) are 277, 376 and 343 DU, and H₂O (right) are 4.22, 0.423, and 1.45 g/cm², for the “Tropical,” “Subarctic Winter” and “U.S. Standard,” respectively. In the top left panel, the column concentration of NO₂ is 0.207 DU.

Elect (electronic) typically follows

$$\Delta e_{\text{Rot}} \ll \Delta e_{\text{Vib}} \ll \Delta e_{\text{Elect}}.$$

Transitions that can occur between adjacent energy states are in the far infrared (20 μm to 1000 μm) for adjacent rotational states, in the spectral range 1 μm to 10 μm for the adjacent vibrational states, and in the visible and ultraviolet for adjacent electronic states. (Note, these are coarse wavelength assignments.) If we indicate quantum states by n_{Rot} , n_{Vib} , n_{Elect} , where the n ’s represent the collection of quantum numbers needed to specify the quantum state, then some of the quantum numbers must change during the transition or the oscillator strength associated with the transition will vanish. For rotational transitions only n_{Rot} changes, for vibrational transitions n_{Vib} and n_{Rot} but n_{Elect} does not change, and for electronic transitions n_{Elect} must change while n_{Rot} and n_{Vib} may or may not change.

Other than translation, the possible motions of the atoms in a diatomic molecule are particularly simple. They can rotate (more or less rigidly) about two mutually perpendicular axes through the center of mass, and their interatomic spacing (separation) can vary (vibration). Their energies can be well approximated by

$$e(J, v, n_{\text{Elect}}) = B(v, n_{\text{Elect}})J(J+1) + h\nu \left(v + \frac{1}{2}\right) + e_{\text{Elect}}, \quad (4.23)$$

where J and v are positive integers, B is called the rotational constant, $B = h^2/(8\pi^2 I)$, ν is the natural frequency of vibration of the molecule, h is Planck's constant, and I is the moment of inertia with respect to the center of mass of the molecule in the vibrational state v and electronic state n_{Elect} . The quantum numbers J and v are called the rotational and vibrational quantum numbers, respectively. The angular momentum of rotation is proportional to $\sqrt{J(J+1)}$. Quantum mechanics tells us that unless the molecule has a permanent dipole moment only energy changes with $\Delta J = \pm 1$ and $\Delta v = \pm 1$ are induced by interaction with an electromagnetic field, i.e., $f_{ij} \neq 0$. Consider a situation in which $n_{\text{Elect}} \rightarrow n'_{\text{Elect}}$, J changes by ± 1 , and v increases by 1. In general $B(v, n_{\text{Elect}})$ will not be the same for both values of n_{Elect} and v , but will be nearly independent of v . The energy change is

$$\Delta e = \Delta e_{\text{Elect}} + h\nu \pm \bar{B}m + \Delta Bm^2, \quad m = 1, 2, \dots,$$

where

$$\bar{B} = B(v+1, n'_{\text{Elect}}) + B(v, n_{\text{Elect}}), \quad \text{and} \quad \Delta B = B(v+1, n'_{\text{Elect}}) - B(v, n_{\text{Elect}}),$$

and m is the j -value of the lower state. Radiation of frequencies centered at

$$\omega = \frac{2\pi\Delta e}{h} = \frac{2\pi\Delta e_{\text{Elect}}}{h} + 2\pi\nu \pm \frac{2\pi}{h}\bar{B}m + \frac{2\pi}{h}\Delta Bm^2, \quad m = 1, 2, \dots, \quad (4.24)$$

will be absorbed, resulting in a series of absorption features that are symmetrically spaced about $\omega = 2\pi\Delta e_{\text{Elect}}/h + 2\pi\nu$ when $\Delta B = 0$. When $\Delta B \neq 0$ there will be a maximum or minimum frequency absorbed depending on whether $\Delta B < 0$ (maximum) or $\Delta B > 0$ (minimum). Such spectra are usually referred to a "band spectra" as they had the appearance of absorption bands on photographic spectrograms taken at low spectral resolution.

In molecular spectroscopy the angular frequency ω is not the preferred unit to designate emission or absorption. It is customary to use the wave number, $\tilde{\nu}$, which is the number of wavelengths present in 1 centimeter of length. The wave number is related to ω through

$$\tilde{\nu} = \frac{1}{\lambda} = \frac{c_{\ell}}{\nu} = \frac{2\pi\omega}{c_{\ell}} = \frac{2\pi}{\kappa},$$

where κ is the magnitude of the propagation vector in Chapter 1. In this equation, λ is always in cm, so the units of $\tilde{\nu}$ are cm^{-1} , often referred to as reciprocal centimeters. The

wave number corresponding to radiation with $\lambda = 1 \mu\text{m}$ is $10,000 \text{ cm}^{-1}$. The frequencies in Eq. (4.24) can range from the ultraviolet to the far infrared region of the spectrum. If $\Delta e_{\text{Elect}} = 0$, i.e., the electronic state of the molecule does not change, then the wave number range associated with the transitions is ~ 200 to $10,000 \text{ cm}^{-1}$, and is known as a vibration-rotation spectrum. If $\Delta e_{\text{Elect}} = 0$ and Δe_{Vib} are both zero, which requires $\Delta v = 0$ and can only occur when the molecule has a permanent dipole moment as in CO, HCl, etc., but not O_2 or N_2 , the spectrum is in the far infrared or microwave region with $\tilde{\nu} \sim 1$ to 200 cm^{-1} and is called a pure rotational spectrum. The absorption spectra of interest to for ocean color remote sensing are in the spectral region ~ 0.4 to $\sim 2 \mu\text{m}$ or $\sim 25,000$ to $\sim 5,000 \text{ cm}^{-1}$. These are either vibrational-rotational transitions or vibrational-rotational in combination with electronic transitions. Again, for diatomic molecules they are characterized by a simple absorption spectrum given by Eq. (4.24).

For molecules with more than two atoms the situation is much more complex, as there are more than one mode of vibration. The spectra are further complicated when the molecule is not linear as in the case of three of the gases of interest to us – H_2O , NO_2 and O_3 – there are three axes of rotation rather than two. These molecules have well understood spectra, but there is no simple pattern, as in Eq. (4.24), to the spectrum. In fact, the spectral absorption features in the individual electronic-vibration-rotation transition appear to the eye to be located at random, and for some purposes can be treated as actually occurring at random over a given spectral interval.

4.4.3.1 Molecular Oxygen

A high spectral resolution absorption spectrum of the atmosphere in the region near 760 nm is provided in Figure 4.26. The spectrum is obtained by measuring the atmospheric transmission of direct solar radiation and shows a more or less regular spacing of lines. Most of the lines are due to molecular Oxygen and this particular band is called the “A” band. The $m = 0$ position, where there is no absorption feature is near 762 nm. Note that the lines appear to get farther apart on the long-wave side of 762 nm and closer together on the short-wave side. This happens because of the term proportional to m^2 in the transition energy occurs when $\Delta B \neq 0$. The lines that are crowding together on the short-wave side actually start appearing at longer wavelengths than about 759.5 nm as m increases, i.e., the term proportional to m^2 overtakes the term proportional to m . Since a maximum of Δe or ω occurs, $\Delta B < 0$ for this band. The wavelength 759.5 nm, where features disappear on the short-wave side of the crowding causes enhanced absorption (many transitions occurring near the same wavelength), is called the “band head.” We shall see later, two other Oxygen absorption bands, called the “B” band and the “ γ ” band, are also present in the visible region of the spectrum.

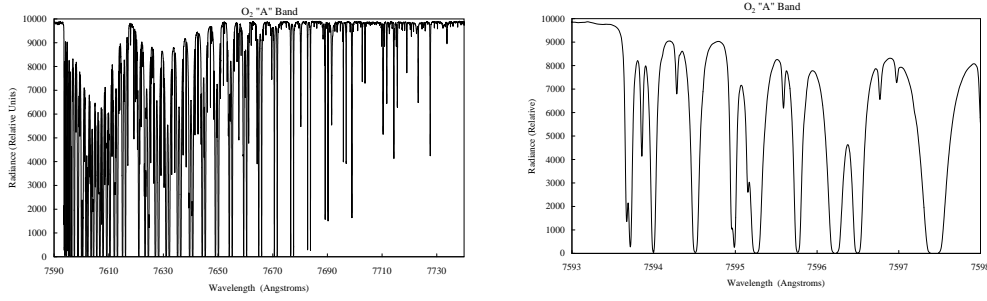


Figure 4.26: The “A” band of Oxygen near 760 nm (1 nm = 10 Angstroms) at very high spectral resolution.

4.4.3.2 Water Vapor

Water vapor is a significant absorber in the visible and NIR. Figure 4.27 provides an example of the absorption spectrum of water vapor. The left panel shows the bands in the visible and NIR, while the right panel is expanded to include the short-wave infrared (SWIR). The concentration of water vapor is that of the U.S. Standard atmosphere (1.45 g/cm²). The spectral resolution of these spectra is 20 cm⁻¹. This is considerably higher than the spectral resolution of most ocean color sensors – normally 10-20 nm in the visible and 20-40 nm in the NIR. (At 500 nm a spectral resolution of 10 nm corresponds to 400 cm⁻¹.) Note the decreasing strength of the absorption with decreasing wavelength. The bands in the visible, although very weak, are important because of the highly variable nature of the water vapor concentration; however, because they are weak, the absorption will be exponential with an optical depth that is proportional to the column concentration of water vapor. In the NIR the bands are much stronger, and the absorption is a more complex function of the concentration;¹³ however, this is of little concern to us because spectral regions of strong atmospheric absorption are to be avoided in remote sensing

¹³If one assumes that there are a large number of lines in $\Delta\omega_{rs}$, and that they all have the same f_{ij} and γ_{ij} but that their spacing are random, then it can be shown that

$$\langle A \rangle = 1 - \exp[-\langle A \rangle_1],$$

where $\langle A \rangle_1$ is the value of $\langle A \rangle$ for a single isolated line. For a weak line $\langle A \rangle_1$ is given by Eq. (4.20), while for a strong line it is given by Eq. (4.22). Thus, for weak and strong lines

$$\langle T \rangle \propto \exp[-c_w \tau_c] \text{ (weak)} \quad \text{and} \quad \langle T \rangle \propto \exp[-c_s \sqrt{\tau_c}] \text{ (strong)}$$

where the c 's are constants. Note that this is *not* the transmittance of a single $\Delta\omega_{rs}$, but it is the transmittance averaged over many $\Delta\omega_{rs}$'s.

applications.

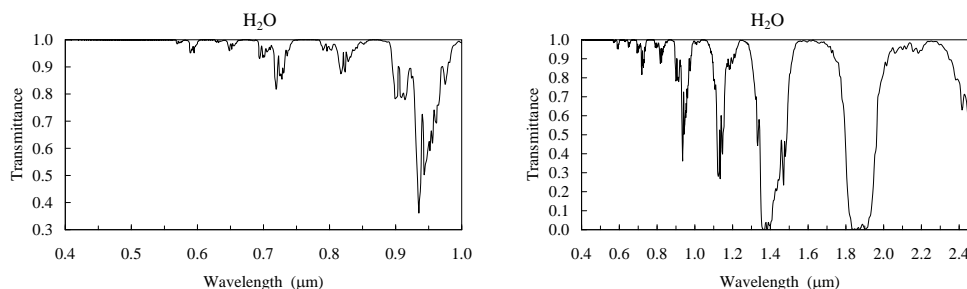


Figure 4.27: The spectrum of water vapor in the visible and NIR (left) and in the visible, NIR, and SWIR (right). The concentration is that of the U.S. Standard Atmosphere.

4.4.3.3 Ozone and NO_2

Figures 4.28 provide the atmospheric transmittance for O_3 and NO_2 . The concentrations are for the U.S. Standard atmosphere. In both cases, the absorption is weak; however, as it is broad band (and as such, unlike H_2O , cannot be avoided) it must be considered in remote sensing. In both cases, the optical thickness is proportional to the gas concentration, so knowing the column amount provides the transmittance. The NO_2 absorption appears to be very low; however, in a highly polluted atmosphere the concentration can reach ten times the U.S. Standard. In that case the transmittance near 400 nm decreases from ~ 0.9965 to ~ 0.9655 , i.e., comparable to O_3 near 600 nm.

4.4.3.4 Total Transmittance of the Atmosphere

The total transmittance of an aerosol-free U.S. Standard atmosphere is provided in Figure 4.29. The O_2 “A”, “B” and “ γ ” bands have been specifically labeled. Most of the other absorption bands are due to H_2O . The regions between H_2O bands and O_2 bands are called “atmospheric windows.” Remote sensing of ocean color of necessity must take place in the atmospheric windows. The atmosphere is essentially opaque in the absorption bands near

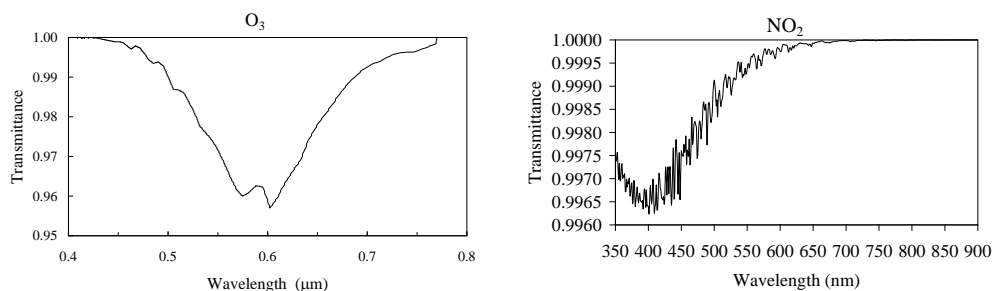


Figure 4.28: Left Panel: The spectrum of Ozone for a concentration equal to that of the U.S. Standard Atmosphere, 343 DU. The spectral resolution is 20 cm^{-1} . The “A” band of Oxygen near 760 nm ($1 \text{ nm} = 10 \text{ Angstroms}$) at very high spectral resolution. Right Panel: The spectrum of NO_2 for a concentration equal to that of the U.S. Standard Atmosphere, 0.207 DU. The spectrum is averaged over 1 nm intervals.

1.4 and $1.85 \mu\text{m}$; however, that does not mean such bands are useless. Because most of the water is in the lower atmosphere (troposphere) observing the earth at these wavelengths allows the observation of cirrus clouds and stratospheric aerosols, i.e., any radiance exiting the atmosphere in these bands must have been scattered by the atmospheric components located *above* the water vapor layer.

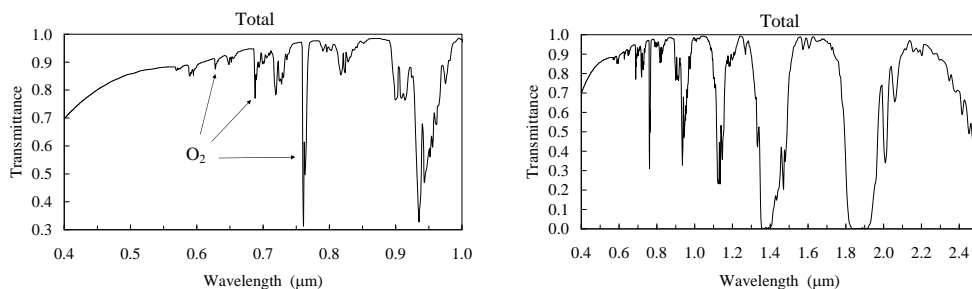


Figure 4.29: The spectrum of aerosol-free transmittance the atmosphere in the visible and NIR (left) and in the visible, NIR, and SWIR (right). The labeled O_2 bands from left to right are called “ γ ”, “B” and “A,” respectively. The gas concentrations are those of the U.S. Standard Atmosphere.

4.5 Simplified Model of the Atmosphere

The facts presented in this chapter suggest a simple atmospheric model as a basis for treating the atmosphere in remote sensing applications. First, because of the strong nature of water vapor absorption and its isolation in well-defined spectral bands (Figure 4.27) it is expedient to just avoid these spectral regions as much as possible. By doing so we can simply ignore H_2O altogether. Second, in most open ocean situations, the aerosol is nearly all in the marine boundary layer, i.e., below most of the atmosphere (Figure 4.23). In this case, a good approximation is to place the aerosol in a layer by itself at the bottom of the atmosphere, or in a layer at the bottom uniformly mixed with air molecules (producing Rayleigh scattering). Next, through most of the troposphere the most important process is Rayleigh scattering by air molecules, suggesting a second layer in which there is no absorption and only Rayleigh scattering. Finally, in the stratosphere the main processes are O_3 absorption (Figure 4.25) and scattering by stratospheric aerosols and thin cirrus clouds, suggesting a third layer for these processes. When the atmosphere is free of cirrus and there have been no recent volcanic eruptions, the main process in the third layer is O_3 absorption. In cases involving dust (Figure 4.24) or urban pollution, the first layer will thicken and have relatively more Rayleigh scattering than when the aerosol is confined in the marine boundary layer. This accounts for the fact that these aerosols in general are mixed into the troposphere and not confined to the marine boundary layer (relatively more Rayleigh scattering in the lowest layer than when the aerosol is confined in the marine boundary layer). When it becomes necessary to add NO_2 , i.e., when the concentration is large, it can be taken into account by the addition of some absorption in the lowest layer; however, it can usually be ignored at background concentrations (U.S. Standard).

Thus, as a starting point for discussion of the role of the atmosphere in ocean remote sensing, our basic model will be a three-layer atmosphere: all the aerosols in the lower layer, with some fraction of the Rayleigh scattering there as well; an aerosol-free, pure Rayleigh scattering, middle layer, and a top layer with only O_3 absorption. More complex situations, e.g., dust, cirrus clouds, etc., will be built on this basic structure.

4.6 Concluding Remarks

Here we have reviewed the optical properties of the atmosphere that are important for the remote sensing of the oceans. Of major importance is the scattering by the air molecules and the scattering and absorption by the aerosol, which we described in detail. Specific details of the absorption of the strongly-absorbing gases are of minor importance as spectral regions displaying such absorption must be avoided — remote sensing is limited to the

“atmospheric windows.” In contrast, some atmospheric windows contain weakly-absorbing lines and they must be considered. Fortunately, the effects of regions of weak absorption are relatively simple to deal with because the absorption is proportional to the total concentration of the gas in the atmosphere. The three-layer model of the atmosphere described above will be used in later chapters to investigate the effect of the atmosphere on the remotely-sensed radiance and to develop a method to remove the aforementioned effects.

4.7 Appendix: Scattering of Spheres and Spheroids – A Brief Comparison

In this chapter, and throughout most of this work, we have tried to understand the IOPs of particles in the environment (aerosols and later, hydrosols) by assuming that they are homogeneous spheres. Here, to provide the reader with a sense of the effect of this assumption, we present a comparison between the scattering by homogeneous spheres and homogeneous spheroids (in random orientation) with the same size distribution.¹⁴ The calculations have been carried out using the T-Matrix method.¹⁵

The particles we examine here are spheroids with an axis of symmetry (the z -axis). In polar coordinates (r, θ, ϕ) , the surface of the particle is given by

$$\frac{1}{r(\theta, \phi)} = \frac{1}{a} \left[\sin^2 \theta + \frac{a^2}{b^2} \cos^2 \theta \right]^{1/2}.$$

A useful parameter for describing the shape of the particle is $\epsilon = b/a$. If the spheroid is prolate (football-like), $\epsilon > 1$; if it is oblate (disk-like) $\epsilon < 1$; and if it is a sphere, $\epsilon = 1$. The size distribution we will use is the log-normal with $R_N = 1.182 \mu\text{m}$ and $\sigma^2 = 0.0953$. These give $R_{\text{eff}} = 1.5 \mu\text{m}$ and $v_{\text{eff}} = 0.1$. The particles within each size range all have the same ϵ . The refractive index is taken to be $\tilde{m} = 1.53 - 0.006i$, which is the mean of the two model indices (in the text) for Saharan dust at 500 nm, the wavelength used in the computation. Three values of ϵ are compared: $\epsilon = 1$ (spheres); $\epsilon = 3/2$ (prolate spheroids); and $\epsilon = 2/3$ (oblate spheroids). The resulting values for the scattering-related cross sections are provided in the table below, and the scattering phase functions are compared in Figure 4.30. The total cross sections and the single scattering albedo for the spheroids are within

¹⁴The radius of a sphere having the same “size” as a spheroid is defined to be that of a sphere that has the same projected area (πr^2) as the spheroid, after the latter has been averaged over all orientations.

¹⁵The T-Matrix method was not discussed in Chapter 3. For smooth particles with a symmetry axis (spheroids, finite cylinders, etc.) it is much more efficient (faster, often by orders of magnitude) and accurate than the DDA, particularly when averaging over particle orientations is required. However, conceptually it is much more complex than, for example the DDA, and is not particularly useful for particles with irregular shapes.

1% of The total cross sections and the single scattering albedo for the spheroids are within

ϵ	σ_c (μm^2)	σ_s (μm^2)	ω_0
2/3	12.21	10.01	0.820
1	12.30	10.14	0.825
3/2	12.21	10.04	0.822

1% of those for the sphere. The greatest effects of non-sphericity here occur in the scattering phase function for $\Theta \gtrsim 45^\circ$ (Figure 4.30). The figure shows that the oblate and prolate spheroids follow similar patterns with respect to their deviation from the sphere: the sphere scattering is larger for $45^\circ \lesssim \Theta \lesssim 100^\circ$; smaller for $100^\circ \lesssim \Theta \lesssim 160^\circ$; and finally, larger for $160^\circ \lesssim \Theta \lesssim 180^\circ$, except close to $\Theta = 180^\circ$. This pattern is generally found when the effective size parameter $x_{\text{eff}} = 2\pi R_{\text{eff}}/\lambda$, is $\gtrsim 5$ -8 (here, $x_{\text{eff}} = 18.8$). It is one of the few regularities in the IOPs of spheroids compared to spheres.

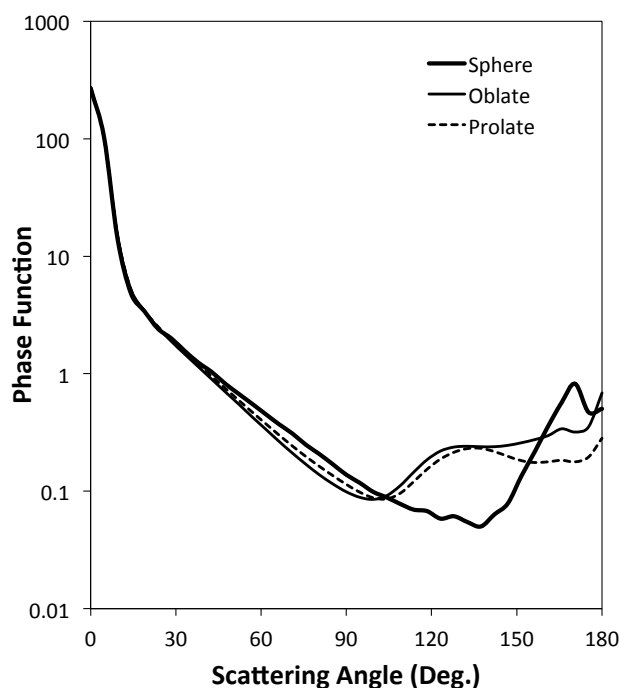


Figure 4.30: $r_g = 1.1819 \mu\text{m}$, $\sigma_g = 0.0953$. $r_{\text{eff}} = 1.5 \mu\text{m}$, $v_{\text{eff}} = 0.10$, $m = 1.53 - 0.006i$, $\lambda = 500 \text{ nm}$.

4.8 Bibliographic Notes

4.2.1 Rayleigh Scattering by the Atmospheric Gas Mixture

The refractivity of atmospheric gases was taken from [Zhang et al. \[2008\]](#), and the King factors for N₂ and O₂ from [Bates \[1984\]](#).

4.2.2 The Rayleigh Optical Depth

Eq. (4.9) is from [Hansen and Travis \[1974\]](#) and (4.10) is from [Bodhaine et al. \[1999\]](#).

4.3.1 Aerosol Particle Size Distributions

The power-law size distribution for aerosols was proposed by [Junge \[1958\]](#). The work referred to in Footnote 4 is [Wang and Gordon \[1993\]](#) and [Zhang and Gordon \[1997\]](#).

The fact that different size distributions can have similar-valued dimensionless optical properties if the distributions have the same values of R_{eff} and v_{eff} (defined in Footnote 6) is shown in [Hansen and Travis \[1974\]](#) and further discussed in [Mishchenko and Travis \[1994\]](#).

4.3.3.1 Haze C Aerosol Model

The Haze C distribution is simply a power-law distribution truncated on both ends. It is called the “Haze C” distribution as that was the term used by [Deirmendjian \[1969\]](#). This distribution was also used by [Fraser \[1976\]](#) in computations of atmospheric radiative transfer.

4.3.3.2 Shettle and Fenn Aerosol Model(s)

These models were developed for LOWTRAN 6 [[Kenizys et al., 1983](#)] and are thoroughly described in [Shettle and Fenn \[1979\]](#). The Coastal model was added by [Gordon and Wang \[1994\]](#).

4.3.3.3 Smirnov et al. Aerosol Model

The Smirnov et al. model is developed in [Smirnov et al. \[2003\]](#) and the influence of wind speed on the large fraction in [Smirnov et al. \[2003\]](#). [Smirnov et al. \[2009\]](#) provides an excellent discussion of the global distribution of aerosols over the ocean obtained from ship and island measurements. The [Smirnov et al. \[2003\]](#) model has been adapted by [Ahmad](#)

et al. [2010] for use in the Gordon and Wang [1994] atmospheric correction algorithm for SeaWiFS and MODIS, replacing the Shettle and Fenn models. The AERONET network is described in Holben et al. [1998].

4.3.3.4 Saharan Dust Models

The Patterson absorption index spectrum is developed in Patterson [1981] and the Moulin et al. absorption index in Moulin et al. [2001]. The particle size distribution for the dust model was developed by Shettle [1984].

4.3.3.5 Stratospheric Aerosol Models

The El Chichon aerosol size distribution was derived in Bandeen and Fraser [1982] and King and Fraser [1983]. The cirrus cloud scattering phase function was computed by Takano and Liou [1989].

4.3.4 Aerosol Vertical Structure

The lidar returns from LITE were taken from the LITE web site.

4.4.1 General Properties of Gaseous Absorption.

Most of this section follows directly from the classical model modified by the introduction of oscillator strengths provided in Chapter 1. The development of the weak- and strong-line average transmittance for an isolated spectrum line followed Goody and Yung [1989].

4.4.2 Vertical Distribution of Principal Absorbing Gases in the Atmosphere

For this section the vertical profiles in the AFGL Atmospheric Constituent Profiles were used [Anderson et al., 1986]. The total concentrations (g/cm^2) were computed by numerical integration.

4.4.3 Absorption Spectra of Atmospheric Gases

Good discussions of molecular spectra can be found in almost any book on quantum chemistry. A classic is *Quantum Chemistry* by Eyring et al. [1944].

4.4.3.1 Molecular Oxygen

The data for the Oxygen “A” band absorption were obtained by Delbouille et al. [1973], and were downloaded from http://bass2000.obspm.fr/solar_spect.php.

4.4.3.2 Water Vapor

The spectra displayed in this section were computed using the LOWTRAN 7 code [Kenizys et al., 1988]. The material in Footnote 13 follows results in Goody and Yung [1989].

4.4.3.3 Ozone and Nitrogen Dioxide

The Ozone spectrum displayed in this section was computed using the LOWTRAN 7 code. The Nitrogen Dioxide spectrum was computed using the absorption coefficients given by Schneider et al. [1987] and Bogumil et al. [2003] and the concentration provided in the U.S. Standard Atmosphere.

4.4.4 Total Transmittance of the Atmosphere

This was computed using LOWTRAN 7.

4.5 Simplified Model of the Atmosphere

This is the atmosphere structure used by Gordon and Wang [1994] in developing the SeaWiFS atmospheric correction algorithm.

4.7 Appendix: Scattering of Spheres and Spheroids – A Brief Comparison

The T-Matrix code used in these computations was developed by M. Mishchenko and downloaded from https://www.giss.nasa.gov/staff/mmishchenko/t_matrix.html. A complete derivation of the T-Matrix method, as well as many comparisons between the scattering between randomly oriented spheroids (and finite-length cylinders) and equal-projected-area spheres, is provided in Mishchenko et al. [2002b].

Chapter 5

Inherent Optical Properties (IOPs) of Natural Waters

5.1 Introduction

In this chapter we discuss the inherent optical properties (IOPs) of natural waters. The IOPs are important because they provide a direct link between constituent concentrations (particulate and dissolved) and radiative transfer within, and out of, the water. The determination of constituent concentrations is the ultimate goal of optical remote sensing of water bodies. Thus, one needs to know what constituents significantly affect the IOPs and to be equipped with quantitative relationships between the constituents and the IOPs. Clearly, considering the range of water bodies one might want to examine, e.g., from the open ocean to turbid estuaries and inland lakes, the scope of this endeavor is enormous. To cut it down to a manageable exercise, we limit the discussion to the open ocean – what is normally called Case 1 waters (see 5.3); however, the methodology of establishing relationships between IOPs and constituents is essentially the same for all waters.

The scope of the chapter is as follows. First, we describe in detail the IOPs of pure water and sea water. Then we provide a survey of the empirical relationships between the IOPs and the principal constituents of most oceanic waters (Case 1): phytoplankton and their immediate detrital material, both particulate and dissolved. Finally, we acquaint the reader with models of the optics of particles, based on scattering theory (Chapter 3), that shed light on some of the empirical relationships. We begin by formally defining the IOPs and describing how they could be measured in the marine environment.

5.1.1 Definitions

The inherent optical properties of natural water, comprising those quantities that enter into the radiative transfer equation, were developed in Chapter 2. These are the elastic volume scattering function $\beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda)$, the inelastic volume scattering function $\beta_{\text{In}}(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda)$, and the absorption coefficient $a(\vec{r}, \lambda)$. Secondary quantities are the elastic and inelastic scattering coefficients

$$b(\vec{r}, \lambda) = \int_{\text{All } \hat{\xi}} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) d\Omega(\hat{\xi})$$

and

$$b_{\text{In}}(\vec{r}, \lambda_E \rightarrow \lambda) = \int_{\text{All } \hat{\xi}} \beta_{\text{In}}(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda) d\Omega(\hat{\xi}),$$

respectively. The total or beam attenuation coefficient $c(\vec{r}, \lambda)$ is given by

$$c(\vec{r}, \lambda) = a(\vec{r}, \lambda) + b(\vec{r}, \lambda) + \int b_{\text{In}}(\vec{r}, \lambda \rightarrow \lambda_x) d\lambda_x, \quad (5.1)$$

where λ_x stands for all wavelengths to which photons at λ inelastically scatter and the integral is over all λ_x .¹

A fundamental property of the IOPs is that they are summable over the constituents. By this we mean if, for example, the beam attenuation coefficient for pure water is denoted by c_w , and that of the i^{th} constituent ($i = 1 \rightarrow N$) is c_i , then

$$c(\vec{r}, \lambda) = c_w(\lambda) + \sum_{i=1}^N c_i(\vec{r}, \lambda),$$

etc. The goal of this chapter is to describe the individual IOPs of the water itself and of its constituents, and to relate the constituent concentrations (where possible) to their IOPs. This will enable us to address a principle goal of water color remote sensing – relating the constituent concentrations to the spectral radiance exiting the water and propagating to the top of the atmosphere.

¹The integral term here is a little tricky experimentally if one directly measures the loss of photons from a beam after accounting for elastic scattering. If both the source and receiver are equipped with narrow band spectral filters passing the *same* band of wavelengths, then any inelastic process represents a loss of photons which is usually interpreted as absorption and therefore the measurement automatically includes the last term. If the source has a broad spectral band and the receiver is equipped with a narrow band filter, then some inelastically scattered light from shorter wavelengths in the source could be collected by the receiver and the measurement interpreted as a smaller absorption coefficient. Finally, if the source is narrow band but the receiver is broad band, the receiver could detect some inelastically scattered light at a longer wavelength and, again the experimental result is interpreted as a decrease in absorption. Here, we consider as “true” absorption, the “ a ” in Eq. (5.1), that in which the energy is lost in the optical sense, e.g., it goes directly to produce heating or to stimulate photosynthesis, rather than reappearing somewhere as “new” photons.

5.1.2 Measurement Principles

There are various methods for measuring the absorption coefficient and the scattering properties of natural waters. Here we describe the basic principles of all measurements based on their operational definitions. Figure 5.1 is a schematic of what might be called a c - β meter. In the figure photons are emitted from a “point” source P and pass through a filter F1 to select a band of wavelengths $\Delta\lambda$ around a central wavelength λ . The light is then rendered into a collimated (parallel) beam by lens L1. In the absence of scattering and absorption in the path between lenses L1 and L2, the beam is brought to a focus at the position of S1 (“S” stands for “stop”), passes through the small hole in the stop and the filter F2 and is measured at the radiation detector D1. The filter F2 is identical to F1. In this configuration, essentially all of the power passing through lens L1 will be recorded by the detector D1. Now, let the space between the two lenses be occupied by an absorbing-scattering medium. Clearly, photons from P that are absorbed in the medium will not reach the detector, so the power will fall when the medium is added. Photons that are scattered will also cause the power reaching the detector to fall. Consider the dashed ray from the source. If a photon along this ray is scattered through a small angle but still reaches L2 (as shown), it will not pass through the hole in the stop S1 and is not recorded. In the limit that the size of the hole in stop S1 goes to zero, any photon that is scattered (or absorbed) between the two lenses will *not* reach D1. Thus, by recording the power detected in the absence and presence of the medium we can measure the beam attenuation coefficient of the medium, c at the wavelength λ (Eq. (5.1)). You might wonder why both F1 and F2 are needed. They are both needed because of the possibility of inelastic processes in the medium. For example, if F1 were omitted the some wavelengths from the source might cause the medium to fluoresce, which would *create* photons within the space between the two lenses, and some of these could be detected by D1. Conversely, if F2 were omitted, photons fluoresced from excitation at λ could reach the detector. Both of these processes would cause the measurement of $c(\lambda)$ to be too small.

The optical arrangement in the upper left of the figure is a radiometer aimed to measure the light that is scattered from the beam through an angle Θ . Again the lens-stop arrangement (L3-S2) restricts the field of view of the detector D2, which views the hole in the stop through a filter F3. The filter F3 is identical to F1 and F2, and its purpose is also to discriminate against inelastic scattering. The power recorded by the detector D2 when the sample is in place is proportional to $\beta(\Theta)$, which can be obtained by combining it with the power recorded in D1 without the scattering medium, and with the geometry of the apparatus allowing one to determine the scattering volume, etc. (Chapter 2). Clearly, measurement of $\beta(\Theta)$ at small Θ (near 0) and large Θ (near 180°) presents experimental challenges.

How do we account for the last term in Eq. (5.1), i.e., the inelastic scattering? The obvious answer is to replace filter F_3 by a series of filters with transmission wavelengths λ_x greater than λ . Measurement with these filters in place would provide $\beta(\Theta, \lambda \rightarrow \lambda_x)$. Fortunately, the last term in Eq. (5.1) is much smaller than the other two and is usually ignored or considered part of $a(\lambda)$; however, when one desired to study inelastic processes, this is how β_{In} has to be determined.

5.2 The IOPs of Pure Water and Sea Water

In this section we describe the IOPs of pure water and sea water in sufficient detail to be useful in water color remote sensing. We begin by presenting the absorption coefficient, then we describe the scattering coefficient, and finally we consider Raman scattering by water molecules.

5.2.1 The Absorption Coefficient of Pure Water and Sea Water

Water absorbs strongly throughout much of the electromagnetic spectrum. Interestingly, the least absorption is found in the visible (the blue actually), with increasing absorption as one moves from the visible into the ultraviolet and the infrared. Measuring the absorption coefficient of pure water is, in principle, simple: use an apparatus as shown in Figure 5.1 to measure c and β ; compute the scattering coefficient b (Section 5.2.2) and subtract it from c along with the loss due to Raman scattering (Section 5.2.3). Figure 5.2 provides measurements of the absorption coefficient of pure water in the visible and Figure 5.3 in the near infrared. The minimum of absorption is seen to be approximately 420 nm where the value is approximately 0.0044 m^{-1} . To appreciate this number note that the transmission of 1 m of pure water (ignoring scattering) at 420 nm would be 99.56%, 10 m – 95.7%, and 100 m – 64.4%. Clearly, measuring such small absorption with high precision presents a significant challenge, requiring very large path lengths.² As such, there is some uncertainty in the literature as to the actual absorption coefficients near the minimum. Another challenge is to purify the water to remove all constituents. This is particularly

²For the measurements provided here, the long path lengths are achieved by placing the water inside a cavity with diffusely reflecting walls having very high reflectivity ($> 99\%$) – an integrating cavity absorption meter. The measurement method is as follows. Integrate the three-dimensional radiative transfer equation over all solid angles and ignore inelastic process. This yields $\nabla \bullet \vec{F} = -a \int L d\Omega$, where $\vec{F} = \int \hat{\xi} L d\Omega$. Integrating this over the volume of the medium and using the divergence theorem, we find $\int_S \vec{F} \bullet \hat{n} = aV \int L d\Omega$, where S is the surface of the cavity, V the volume, and \hat{n} is the *outward* normal to the cavity. The left hand side is the net power *out* of the cavity: $P_{\text{Out}} - P_{\text{In}}$. If L is totally diffuse, the integral is $4\pi L$, but πL is just the irradiance (power out per unit surface area) *out* of the cavity, so $\pi L = P_{\text{Out}}/A$, where A

important in the UV where even a trace of dissolved organic material can cause a significant absorption. Thus, one should consider the values provided in Figure 5.2 in the blue and UV as upper limits, while for wavelengths greater than about 450 nm the results are generally accepted as being quite accurate. In the near infrared there is a small linear dependence on temperature ($\sim 0.01 \text{ m}^{-1}/^{\circ}\text{K}$ at 750 nm) as well as salinity; however, these would have negligible effects in remote sensing. In the visible, the temperature dependence is $\sim 0.0001 \text{ m}^{-1}/^{\circ}\text{K}$ or less. The high absorption coefficients near 750 and 850 nm are important because these wavelengths are in atmospheric “windows” (Chapter 4) and the strong absorption by liquid water ensures that negligible radiance will exit the water at these wavelengths (except at high concentrations of suspended material). This is important in remote sensing for removal of the effects of the intervening atmosphere. In atmospheric windows further in the infrared (the short waver infrared or SWIR) the water absorption is even higher: $\sim 40 \text{ m}^{-1}$ near $1.06 \mu\text{m}$; $\sim 70 \text{ m}^{-1}$ near $1.6 \mu\text{m}$; and $> 1000 \text{ m}^{-1}$ near $2.2 \mu\text{m}$. In the SWIR there will be negligible radiance exiting the water except at extraordinarily high concentrations of suspended material.

5.2.2 The Scattering of Pure Water and Sea Water

The volume scattering function for molecular scattering by a liquid was developed in Chapter 3:

$$\beta(\Theta) = \frac{\pi^2}{2\lambda^4} C \left(\frac{6 + 3\delta}{6 - 7\delta} \right) \left(\frac{(1 + \delta) + (1 - \delta) \cos^2 \Theta}{1 + \delta/2} \right), \quad (5.2)$$

where Θ is the scattering angle, δ is the depolarization factor,

$$C \triangleq \left(\frac{\partial K_e}{\partial \rho} \right)_T^2 \rho^2 T \beta_T k_B + \left(\frac{\partial K_e}{\partial T} \right)_\rho^2 \frac{k_B T}{\rho c_v},$$

k_B is the Boltzmann constant, β_T is the isothermal compressibility of water, K_e is the dielectric constant, ρ is the density, T is the absolute temperature and c_v is the specific heat at constant volume. Simple integration of β over solid angle yields the total scattering

is the surface area of the cavity. The final result is

$$a = \frac{P_{\text{In}} - P_{\text{Out}}}{P_{\text{Out}}} \frac{A}{4V}.$$

Thus measuring the P 's and the shape of the cavity allows determination of a . Reiterating, the requirements for this are that the radiance within the cavity is totally diffuse and independent of position. These requirements can be met approximately if the absorption is not too large and the walls are lambertian with a high reflectivity. An attractive feature of the integrating cavity absorption meter is the fact that it is insensitive to scattering in the medium. Accurate measurements of the absorption coefficient are possible even when the scattering coefficient exceeds the absorption coefficient by an order of magnitude or more.

coefficient (b):

$$b = \frac{8\pi}{3} \left(\frac{2 + \delta}{1 + \delta} \right) \beta(90^\circ). \quad (5.3)$$

The full volume scattering Mueller matrix for water is

$$\begin{aligned} \beta(\Theta) = \frac{3b}{8\pi} \left[\frac{1 - \delta}{1 + \delta/2} \right] & \begin{pmatrix} \frac{1}{2}(1 + \cos^2 \Theta) & -\frac{1}{2}\sin^2 \Theta & 0 & 0 \\ -\frac{1}{2}\sin^2 \Theta & \frac{1}{2}(1 + \cos^2 \Theta) & 0 & 0 \\ 0 & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & \cos \Theta \end{pmatrix} \\ + \frac{3b}{8\pi} \left[\frac{\delta}{1 + \delta/2} \right] & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\cos \Theta \end{pmatrix} \end{aligned}$$

Most measurements of scattering by liquids involve measuring the scattering at $\Theta = 90^\circ$ with some combination of polarizers in the incident and scattered beams (see Chapter 3, Appendix 3). Direct measurements of the scattering at 90° by pure water are provided in Table 5.1. There is a wide range in the laboratory measurements of δ : 0.039 to 0.09. Some

Table 5.1: Measured values of $\beta(90^\circ)$ for pure water as a function of the wavelength λ at $\sim 20^\circ\text{C}$. The units of λ are nm, and β and b are m^{-1} . The values of the latter two have been multiplied by 10^4 . The value of b has been computed from $\beta(90^\circ)$ using Eq. (5.3) with $\delta = 0.039$.

λ	366	405	436	546	578
$\beta(90^\circ)$	4.53	2.90	2.12	0.835	0.660
b	74.5	47.7	34.9	13.7	10.9

of this discrepancy is attributed to the difficulty in removing all of the suspended particles the samples. The presence of particles will influence the polarization of the scattered light – usually increasing the depolarization. In addition, internally scattered light within the measuring instrument will also lead to an apparent increase in the measured δ . Thus, the smallest of the measurements (0.039) is considered to be the most reliable.

If the spectral measurements of $\beta(90^\circ)$ in Table 5.1 are fit to a power law function of wavelength, it is found that

$$\frac{\beta(90^\circ, \lambda)}{\beta(90^\circ, 546 \text{ nm})} = 0.9998 \left(\frac{546}{\lambda} \right)^{4.2},$$

where λ is in nm. The deviation from the λ^{-4} dependence in Eq. (5.2) is due to the variation of the dielectric constant (refractive index) of water with wavelength.

The scattering by sea water is influenced by the dissolved ions in the water. In each elementary volume of water the concentration of ions undergoes fluctuation just as the concentration of water molecules. This in turn leads to a fluctuation in the dielectric constant of the volume and hence additional scattering. Measurements of the scattering by sea water are provided in Table 5.2. It is seen that sea water at a salinity of 38.5 parts per thousand ($^{\circ}/_{00}$) scatters about 30% more than pure water. Experiments on prepared solutions of salts suggest that, for the range of salinities mostly encountered in natural waters, $\beta(90^{\circ})$ is approximately a linear function of the salt concentration. This suggests that, as long as the relative concentration of the individual salts in sea water is nearly constant, the scattering should be a linear function of the salinity S . In fact,

$$\beta(90^{\circ})_{\text{Sea Water}} = \beta(90^{\circ})_{\text{Pure Water}} \left(1 + 0.3 \frac{S}{38.5} \right), \quad (5.4)$$

where S is in $^{\circ}/_{00}$.

Table 5.2: Measured values of $\beta(90^{\circ})$ for sea water (salinity = 38.5 $^{\circ}/_{00}$) as a function of the wavelength λ at $\sim 20^{\circ}\text{C}$. The units of λ are nm, and β and b are m^{-1} . The values of the latter two have been multiplied by 10^4 . The value of b has been computed from $\beta(90^{\circ})$ using Eq. (5.3) with $\delta = 0.039$.

λ	366	405	436	546	578
$\beta(90^{\circ})$	5.87	3.84	2.78	1.08	0.86
b	96.5	63.1	45.7	17.8	14.1

Of particular interest in remote sensing is the backscattering coefficient b_b defined through

$$b_b = 2\pi \int_{90^{\circ}}^{180^{\circ}} \beta(\Theta) \sin \Theta d\Theta,$$

because the upward propagating radiance near the surface is proportional to $b_b/(a + b_b)$, where a is the absorption coefficient. The backscattering coefficient accounts for all photons scattered from a beam with $\hat{\xi} \bullet \hat{\xi}' < 0$. Although suspended particles usually scatter much more strongly than water itself, most of the particle scattering is in the near forward direction, and at scattering angles $\Theta > 90^{\circ}$ the particle and water contributions to b_b can be similar. In fact, water can even be dominate for $\Theta > 90^{\circ}$: for the data presented in Figure 6.1 in Chapter 6 from near Hawaii, water contributes 80% of the total b_b at 400 nm and more than 70% at 450 nm. Direct integration of Eq. (5.2) shows that $b_b = b/2$.

Theoretical computation of $\beta(90^{\circ})$ using Eq. (5.2) is possible if the measured value of δ is used. In the case of liquid water, the second term in C is only about 1% of the first and

can be ignored so

$$C = \left(\frac{\partial m^2}{\partial \rho} \right)_T^2 \rho^2 T \beta_T k_B = 4m^2 \left(\frac{\partial m}{\partial \rho} \right)_T^2 \rho^2 T \beta_T k_B,$$

where the usual substitution $K_e = m^2$ has been used. The indicated derivative is apparently difficult to measure in practice, so the following replacement is made,

$$\left(\frac{\partial m}{\partial \rho} \right)_T = \left(\frac{\partial m}{\partial P} \right)_T \left(\frac{\partial P}{\partial \rho} \right)_T = \left(\frac{\partial m}{\partial P} \right)_T \frac{1}{\rho \beta_T},$$

resulting in an easier derivative to measure. Collecting all of the terms we have,

$$\beta(90^\circ) = \frac{2\pi^2 m^2}{\lambda^4 \beta_T} k_B T \left(\frac{\partial m}{\partial P} \right)_T^2 \left(\frac{6 + 6\delta}{6 - 7\delta} \right), \quad (5.5)$$

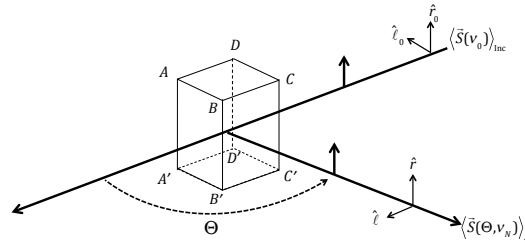
as the final relationship between the the scattering by liquid water and thermodynamic properties. This equation along with the measured thermodynamic properties provide theoretical values of $\beta(90^\circ)$ that agree with those in Table 5.1 with a difference of $< 2\%$. Although $\beta(90^\circ)$ appears to increase in direct proportion to temperature (Eq. (5.5)), n , β_T , and the derivative all depend on T , with the result that $\beta(90^\circ)$ actually has a minimum at $\sim 26^\circ\text{C}$ with a variation of 3.7% from 0 to 26°C . Similarly, consideration of the thermodynamics of fluctuations in salinity results in agreement between theory and Eq. (5.4).

5.2.3 The Raman Scattering of Pure Water

As described in Chapter 3, Raman scattering involves a *frequency* change from the incident to the scattered radiation. If the angular frequency of the incident wave is ω_0 , the Raman-scattered waves will have frequencies $\omega_0 \pm \omega_v$, where ω_v is a vibrational (angular) frequency of the molecule (here the water molecule). It is traditional in molecular spectroscopy to use the wave number ν rather than the frequency ω . The wave number is the inverse of the wavelength: $\nu = 1/\lambda$ with λ measured in cm. Thus, the wave number corresponding to a frequency ω is $\nu = \omega/2\pi c_\ell$, where c_ℓ is the speed of light (here in cm/sec), and the wave number of the Raman-scattered light (ν_N , where “N” stands for “new”) is $\nu_N = \nu_0 \pm \nu_v$. As discussed in Chapter 3, the scattered wave number with the “+” sign (the “anti Stokes” line) is much weaker than that with the “−” sign (the “Stokes” line) and will be ignored in what follows. Further, we note that although classically (i.e., as in Appendix II in Chapter 3) there is only a single value of ν_v , quantum effects in Raman scattering are manifest in many, closely-spaced values of ν_v replacing the single value. The result of this is that the Raman scattering, rather than producing radiation at a single wave number ν_N , produces

a continuum of radiation. (However, as we shall see, there is significant power in the continuum only over a relatively narrow spectral band.)

The definitive experiment to determine the magnitude of Raman scattering by pure water consisted of measuring the Raman scattering by a beam at a wavelength 488 nm (the excitation wavelength). Then, using the relationships developed in Appendix III of Chapter 3, and measuring the scattering at various states of polarization at a given scattering angle (in this case 90°), allows estimation of the Raman scattering coefficient for scattering from 488 nm to longer wavelengths. We now describe the results of this experiment.



The figure above shows the experimental setup. Light, linearly polarized in the \hat{r}_0 direction, is incident on a cuvette through face $CC'DD'$, and is scattered at 90° exiting face $BB'CC'$. The incident light has a wavenumber ν_0 and has a spectral spread $\Delta\nu_0$. The scattered power has a wave number ν_N . Its component in the \hat{r} direction in a wave number band $\Delta\nu_N$ is measured using a monochromator equipped with a polarizer with pass direction parallel to \hat{r} . The Raman volume scattering function in this geometry is given by³

$$\beta_R(\nu_0 \rightarrow \nu_N) = \frac{\Delta^2 P'(\text{in } \Delta\nu_N)/\Delta\Omega}{P'(\nu_0)\Delta\nu_0\Delta\ell},$$

where $\Delta\Omega$ is the solid angle the detector subtends at the scattering center, i.e., the solid angle in which the scattered light is detected, and $\Delta^2 P'$ is the radiant power per wave number in $\Delta\nu_N$ scattered into $\Delta\Omega$. Note that the units of $\beta_R(\nu_0 \rightarrow \nu_N)$ are $\text{m}^{-1}\text{Sr}^{-1}/\text{cm}^{-1}$, where the cm^{-1} comes from the $\Delta\nu_N$ in $\Delta^2 P'$. With the particular experimental setup, β_R is what we referred to as $\beta_{r \rightarrow r}$ in Appendix III of Chapter 3, i.e., $\beta_R(\nu_0 \rightarrow \nu_N) = \beta_{r \rightarrow r}(90^\circ)$.

³Unlike the definition in Chapter 2, where $P(\lambda_0) = \Delta\mathcal{P}(\lambda_0)/\Delta\lambda_0$, with $\Delta\mathcal{P}(\lambda_0)$ being the sum of all the time-averaged Poynting vectors in a beam of radiation (times the cross sectional area of the beam) having wavelengths between $\lambda_0 - \Delta\lambda_0/2$ and $\lambda_0 + \Delta\lambda_0/2$, and with units Watts/ μm , here $P'(\nu_0) = \Delta\mathcal{P}(\nu_0)/\Delta\nu_0$, with $\Delta\mathcal{P}(\nu_0)$ the sum of all the time-averaged Poynting vectors (times the cross sectional area of the beam) with wave numbers ν between $\nu_0 - \Delta\nu_0/2$ and $\nu_0 + \Delta\nu_0/2$, and units of Watts/ cm^{-1} . In this case, $P'(\nu_0)\Delta\nu_0$ is the radiant power on a plane normal to the beam's propagation direction, i.e., face $CC'DD'$.

Using the relationships derived in Appendix III, i.e.,

$$\beta_{r \rightarrow r}(\Theta) = \frac{3b}{8\pi} \left(\frac{1 - \delta/2}{1 + \delta/2} \right) = \frac{3b}{8\pi} \left(\frac{1}{1 + 2\rho} \right)$$

allows determination of the total Raman scattering coefficient from ν_0 to ν_N (in $\Delta\nu_N$), $b_R(\nu_0 \rightarrow \nu_N)$

$$b_R(\nu_0 \rightarrow \nu_N) = \frac{8\pi}{3} \beta_R(\nu_0 \rightarrow \nu_N) (1 + 2\rho(\nu_N)),$$

once the Raman depolarization ratio $\rho = \beta_{\ell \rightarrow r} / \beta_{r \rightarrow r}$ is known.⁴ The Raman depolarization ratio was directly measured in a different experiment, in which scattered intensities proportional to $\beta_{\ell \rightarrow r}$ and $\beta_{r \rightarrow r}$ were individually measured at $\Theta = 180^\circ$, but at $\lambda_0 = 460$ nm. The results of that measurement as a function of the Raman shift $\nu_s = \nu_0 - \nu_N$ are provided in Figure 5.4. The value of ρ averaged over the entire Raman band is 0.161. The value near the maximum of $b_R(\nu_0 \rightarrow \nu_0 - \nu_s)$ is ~ 0.175 . Using ρ determined in this manner, $b_R(\nu_0 \rightarrow \nu_N) = b_R(\nu_s)$, is shown in Figure 5.5. When $b_R(\nu_0 \rightarrow \nu_N)$ is multiplied by $\Delta\nu_N$ and summed over all ν_N we get

$$\sum_{\Delta\nu_N} b_R(\nu_0 \rightarrow \nu_N) \Delta\nu_N = \sum_{\Delta\nu_s} b_R(\nu_0 \rightarrow \nu_N) \Delta\nu_s \triangleq b_R(\nu_0 \rightarrow \nu_N \text{ Band}),$$

where $b_R(\nu_0 \rightarrow \nu_N \text{ Band})$ is the Raman scattering coefficient for scattering from ν_0 into the entire Raman band (with units m^{-1}). For $\lambda_0 = 488$ nm, the resulting $b_R(\nu_0 \rightarrow \text{Band})$ is $2.61 \times 10^{-4} \text{ m}^{-1}$. This is the contribution of Raman scattering to the third term in Eq. (5.1) at 488 nm, i.e., the contribution of Raman scattering to the extinction coefficient of water at 488 nm.⁵ It is important to note that the normalized spectrum of the Raman shift,

$$[b_R(\nu_s)]_{\text{Norm}} \triangleq \frac{b_R(\nu_s)}{b_R(\nu_0 \rightarrow \nu_N \text{ Band})}$$

is independent of ν_0 , i.e., the *spectral shape* of the Raman emission is independent of the wavelength of the excitation.⁶ Figure 5.6 shows the value of the above sum (or $\int b_R(\nu_s) d\nu_s$) as a function of the upper limit. It shows that 90% of the Raman scattering is between $\nu_s \approx 3052$ and 3625 cm^{-1} , with a median at 3353 cm^{-1} . There is a slight variation in the spectral *shape* of the Raman spectrum with temperature: the shoulder near 3250 cm^{-1} decreases slightly with increasing temperature, while the peak at 3425 cm^{-1} increases

⁴The reason the b can be determined from $\beta_{r \rightarrow r}$ at a single scattering angle (given ρ) is, of course, due to the fact that the full Raman (and Rayleigh) Mueller scattering matrix is (are) known.

⁵For pure water at 488 nm, the coefficients associated with loss from a beam by scattering and absorption are: the absorption $139 \times 10^{-4} \text{ m}^{-1}$; the elastic scattering (Rayleigh scattering), $21.9 \times 10^{-4} \text{ m}^{-1}$; Raman scattering, $2.61 \times 10^{-4} \text{ m}^{-1}$.

⁶This is suggested by the simple classical theory of Raman scattering provided in Appendix II of Chapter 3: the Raman emission depends on the excitation only through the κ^4 term in Eq. (3.61).

slightly. These changes tend to cancel in the above integral reducing the overall temperature dependence. For our purposes it is safe to assume that b_R is independent of temperature. (Salinity appears to have little or no affect on b_R .)

In general, what we desire to know at a given wave number or wavelength is not only the *loss* of power due to Raman scattering *from* a beam, but also the *gain* of power resulting from Raman scattering by other wavelengths *into* a beam. For example, suppose we are interested in the effect of Raman scattering on radiative transfer at a wave number ν_I , where the subscript “ I ” stands for “interest.” We might ask the question for example, what is the contribution of Raman scattering from higher wave numbers to the radiance observed at ν_I ? If we ignore polarization, the relevant term in the radiative transfer equation is

$$Q(\vec{r}, \hat{\xi}, \nu_I) = \int_{\nu_0} \int_{\text{All } \hat{\xi}'} \beta_R(\hat{\xi}' \rightarrow \hat{\xi}; \nu_0 \rightarrow \nu_I) L(\vec{r}, \hat{\xi}', \nu_0) d\Omega(\hat{\xi}') d\nu_0,$$

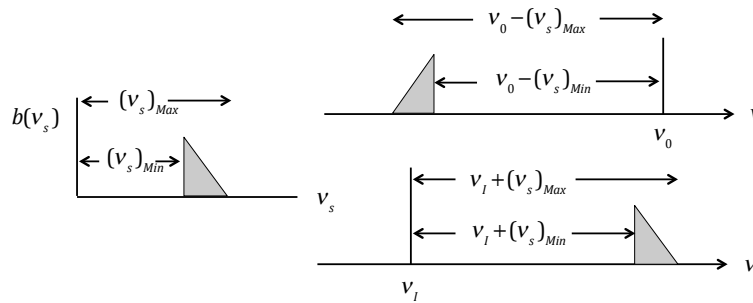
or, since $\nu_0 = \nu_I + \nu_s$ and $\beta_R(\nu_0 \rightarrow \nu_I; \hat{\xi}' \rightarrow \hat{\xi}) = 4\pi P_R(\hat{\xi}' \rightarrow \hat{\xi}) b_R(\nu_I + \nu_s \rightarrow \nu_I)$,

$$Q(\vec{r}, \hat{\xi}, \nu_I) = \int_{\nu_0} \int_{\text{All } \hat{\xi}'} 4\pi P_R(\hat{\xi}' \rightarrow \hat{\xi}) b_R(\nu_I + \nu_s \rightarrow \nu_I) L(\vec{r}, \hat{\xi}', \nu_0) d\Omega(\hat{\xi}') d\nu_s, \quad (5.6)$$

where in these equations, β_R and P_R are the 11 elements of the associated scattering and phase Mueller matrices. In addition, if $L(\vec{r}, \hat{\xi}', \nu_0)$ a slowly varying function of ν_0 , as it usually is, then it can be replaced by $L(\vec{r}, \hat{\xi}', \bar{\nu}_0)$, where $\bar{\nu}_0$ is a representative wave number within the Raman excitation band, and taken outside the ν_s integration, i.e.,

$$Q(\vec{r}, \hat{\xi}, \nu_I) = \int_{\text{All } \hat{\xi}'} 4\pi P_R(\hat{\xi}' \rightarrow \hat{\xi}) L(\vec{r}, \hat{\xi}', \bar{\nu}_0) d\Omega(\hat{\xi}') \int_{\nu_s} b_R(\nu_I + \nu_s \rightarrow \nu_I) d\nu_s$$

So, to answer questions such as this, we need $b_R(\nu_I + \nu_s \rightarrow \nu_I)$; however, we only have $b_R(\nu_0 \rightarrow \nu_0 - \nu_s)$, the scattering coefficient to go from a wave number ν_0 to a smaller wave number. This situation is shown schematically in the drawing below. On the left we have $b(\nu_s)$ from the experiment described earlier in the text and determined according to the



drawing on the top right, where the shaded right triangle is the Raman emission band.

The bottom right drawing depicts a band of higher wave numbers contributing Raman emission into wave number ν_I . In the two drawings on the right, the effect of $b_R(\nu_s)$ in emission *to a single* ν_I is simply the mirror image of its effect in emission *from a single* ν_0 , assuming that the excitation irradiance at $\nu_I + \nu_s$ is independent of ν_s . Since the integrals over ν_s of $b_R(\nu_s)$ and its mirror image are identical, we have that the Raman scattering coefficient from a single wave number at ν_0 to a band of wave numbers near ν_N , is the same as the Raman scattering coefficient *into* a single wave number at ν_N from a band of wave numbers near ν_0 , i.e.,

$$b_R(\nu_0 \text{ Band} \rightarrow \nu_N) = b_R(\nu_0 \rightarrow \nu_N \text{ Band}),$$

For $\nu_0 = 20,492 \text{ cm}^{-1}$ ($\lambda_0 = 488 \text{ nm}$) these are both equal to $2.61 \times 10^{-4} \text{ m}^{-1}$.

It is customary, when focus is on the visible spectrum, to use wavelength rather than wave number. If we want to know, for example, the range of excitation wavelengths contributing to the Raman emission at ν_I , we have $\nu_0 = \nu_s + \nu_I$, or $1/\lambda_0 = \nu_s + 1/\lambda_I$, with 90% of the excitation between $\nu_s = 3052$ and 3625 cm^{-1} . In this manner, Figure 5.7 provides the excitation wavelength (and the 90% band) as a function of the emission wavelength across the visible spectrum. Clearly, in the visible the excitation is confined to a relatively narrow spectral band. This supports the simplification that $L(\vec{r}, \hat{\xi}', \nu_0)$ is a slowly varying function of ν_0 .

The total Raman scattering coefficient into the wavelength λ_I depends on λ_I . Experimentally, this dependence is found to be approximately

$$b_R(\lambda_0 \text{ Band} \rightarrow \lambda_I) = 2.61 \times 10^{-4} \text{ m}^{-1} \left(\frac{589}{\lambda_I} \right)^{4.8 \pm 0.3}.$$

This result demonstrates that the simple model presented in Chapter 3 has more shortcomings in explaining some aspects of Raman scattering. It predicts that the wavelength dependence of b_R should be $\propto \lambda_0^{-4}$. This is remedied by the quantum mechanical theory of Raman scattering.

Raman and Rayleigh scattering share a common functional form for the volume scattering function and the Mueller phase matrix. In Chapter 3 we found that the volume scattering function for unpolarized radiation (without regard for polarization of the scattered radiation, i.e., β_{11} or $\beta_{n \rightarrow n}$) is

$$\begin{aligned} \beta_R(\Theta, \nu_I + \nu_s \rightarrow \nu_I) &= \frac{3}{16\pi} b_R(\nu_I + \nu_s \rightarrow \nu_I) \left[\frac{(1 + \delta(\nu_s)) + (1 - \delta(\nu_s)) \cos^2 \Theta}{1 + \delta(\nu_s)/2} \right] \\ &= \frac{3}{16\pi} b_R(\nu_I + \nu_s \rightarrow \nu_I) \left(\frac{1 + 3\rho(\nu_s)}{1 + 2\rho(\nu_s)} \right) \left[1 + \frac{1 - \rho(\nu_s)}{1 + 3\rho(\nu_s)} \cos^2 \Theta \right], \end{aligned}$$

where δ is the alternate form of the depolarization factor: $\delta = 2\rho/(1 + \rho)$ and Θ is the scattering angle. If $L(\vec{r}, \hat{\xi}, \nu_0)$ varies rapidly with ν_0 , then these detailed expressions must be used in Eq. (5.6), with $\nu_0 = \nu_I + \nu_s$. Figure 5.8 provides data from which the variation of β_R with Θ can be determined. The figure shows

$$\frac{3}{16\pi} b_R(\nu_I + \nu_s \rightarrow \nu_I) \left(\frac{1 + 3\rho(\nu_s)}{1 + 2\rho(\nu_s)} \right) \quad \text{and} \quad \frac{1 - \rho(\nu_s)}{1 + 3\rho(\nu_s)}$$

as a function of ν_s , and indicates how the angular distribution varies across the Raman band. However, if $L(\vec{r}, \hat{\xi}', \nu_I + \nu_s)$ varies weakly with ν_s the variation of β_R with ρ_R can be replaced by an average as above. Data on Figures 5.4 and 5.5 provide the b_R -weighted average ρ , i.e., $\int b_R(\nu_s) \rho(\nu_s) d\nu_s / \int b_R(\nu_s) d\nu_s$, which is ≈ 0.17 , so $(1 - \rho(\nu_s))/(1 + 3\rho(\nu_s)) \approx 0.55$, and

$$\beta_R(\Theta, \nu_I + \nu_s \rightarrow \nu_I) = \frac{1}{4\pi \times 1.1835} b_R(\nu_I + \nu_s \rightarrow \nu_s) \left[1 + 0.55 \cos^2 \Theta \right].$$

This equation should suffice to represent the volume scattering function for Raman scattering except when the excitation band includes, for example, a strong solar Fraunhofer absorption line, in which case the argument of the integrand in Eq. (5.6) will depend strongly on ν_s . Thus, when the excitation is slowly varying with $\nu_I + \nu_s$ or with λ_0 , we will take⁷

$$b_R(\lambda_0 \text{ Band} \rightarrow \lambda_I) = 2.61 \times 10^{-4} \left[\frac{589}{\lambda_I} \right]^{4.8}$$

and

$$\beta_R(\Theta, \lambda_0 \text{ Band} \rightarrow \lambda_I) = \frac{1}{4\pi \times 1.1835} b_R(\lambda_0 \text{ Band} \rightarrow \lambda_I) \left[1 + 0.55 \cos^2 \Theta \right].$$

5.3 The IOPs of the Constituents of Natural waters

For our purposes, natural waters can be conveniently separated into two classes: Case 1 waters in which the IOPs are dominated by phytoplankton and their immediate detrital material; and Case 2 waters in which other constituents play an important role. Specifically, Case 1 waters include (1) living phytoplankton, (2) particulate detritus associated with phytoplankton, originating from decay of nonliving phytoplankton as well as debris

⁷Note also that $b_R(\lambda_0 \text{ Band} \rightarrow \lambda_I) = b_R(\lambda_0 \rightarrow \lambda_I \text{ Band})$, where the right hand side of this equation the scattering coefficient for Raman scattering from λ_0 into the entire Raman band. Thus, the attenuation of radiation at λ_0 due to Raman scattering from λ_0 into the full Raman band around λ_I is equal to $2.61 \times 10^{-4} [589/\lambda_I]^{4.8}$ as well.

originating from grazing by zooplankton, and (3) (chromophoric) dissolved organic material (CDOM), associated with and originating from, phytoplankton. However, these are not the only constituents one expects to find in natural waters. There can also be sediment particles (4) resuspended from the bottom in shallow water or (5) introduced into the water by natural runoff from land, (6) dissolved organic material also derived from natural runoff from land, and (7) other particulate and dissolved materials introduced by the activities of man (pollution). Case 2 waters include all seven of these sources of particulate and dissolve material, i.e., they include the constituents of Case 1 waters as well as the four other sources. Case 1 waters are particularly simple in that the IOPs can usually be related to a single measurable parameter: the concentration of the photosynthetic pigment contained in phytoplankton – Chlorophyll a – in the water.⁸ Here, we will limit our discussion mostly to Case 1 waters because (1) they comprise most of the worlds oceans and (2) the IOPs of Case 2 waters are usually site specific, i.e., the scattering by mineral particles introduced by a river flow will depend on the composition and size distribution of the particular particles at that location.

5.3.1 Measurement of the Absorption of the Constituents of Case 1 waters

There are several techniques for measuring the absorption coefficient of the constituents. First, for Case 1 waters, the absorption coefficient can be written

$$a = a_w + a_\phi + a_y + a_{NAP},$$

where a_w is the absorption coefficient of pure water, a_ϕ is that due to phytoplankton, a_y is that due to CDOM, and a_{NAP} is that due to non-algal particles. The subscript “y” is used with CDOM because in the older literature CDOM was often referred to as gelbstoff (in German) or “yellow substance,” due to its ever-increasing absorption into the blue. The absorption due to particles alone is denoted a_p and

$$a_p = a_\phi + a_{NAP}.$$

Measurement of the absorption coefficient of CDOM is straightforward. First, the sample is filtered through a filter with a $0.22\ \mu\text{m}$ pore size (traps all particles with “diameter” >

⁸Exception to this simplification is presented by certain phytoplankton possessing mineral shells. For example the coccolithophore *E. Huxleyi* has a shell consisting of small plates of calcite. Both the intact cells (comprising compartment #1) and detached plates (coccoliths comprising compartment #2) scatter light much more strongly than most phytoplankton and detrital particles. In this case, the concentration of detached coccoliths in the water is also required to determine the scattering properties, and this concentration is *not* in general related to the Chlorophyll a concentration in any direct manner.

0.22 μm). The few small particles that pass through the filter are defined to be part of the dissolved component. Then it is placed in an instrument similar to that in Figure 5.1 and c is measured. Since the scattering by CDOM is negligible, this provides $a_w + a_y$, from which a_y can be determined.

There are two common techniques for measuring a_p .⁹ Obviously if the total a is measured, one simply subtracts a_y and a_w . Can one measure a without having to measure c and β ? The answer is yes and the technique is not difficult. Figure 5.9 provides a schematic of the apparatus. It is similar to the “ c ” of the c - β meter in Figure 5.1, except that the cylindrical beam from P is surrounded by a shiny tube, the lens-stop combination (L2, S1) are removed, and the detector and filter are moved to the edge of the tube. In this apparatus, any photons scattered through small angles will still strike the detector and their power will not be lost. In addition, photons that are scattered through larger angles will strike the shiny tube and still reach the detector (dashed path in Figure 5.9). Only photons that are scattered at large angles (or scattered with $\Theta > 90^\circ$) will be lost from the beam. When the particles are strongly forward scattering (as marine particles are – see below) the loss from the beam is small (and easily estimated). In addition, there is an increase in the path length through the medium for the scattered photons (the average path length is no longer the distance between the lens and the F2-D2 combination but is a little larger); however, this is also easy to assess through modeling the particle scattering or by calibration using non-absorbing particles of known scattering characteristics in a solution having known absorption. Thus, this instrument (with an estimate of the small scattering loss and path-length stretching) allow measurement of $a_w + a_y + a_p$ directly.

The second technique involves filtering the water through a 0.22 μm filter pad (as in the measurement of a_y above), placing the filter pad in front of F2 in Figure 5.9 and measuring the light absorbed by the filter pad. (Note, in this case there is no medium between L1 and F2 except for the filter pad containing the particles.) If τ_f is the optical thickness of the filter pad containing the particles, i.e., $\tau_f = -\ln(\mathcal{T}_f)$, where \mathcal{T}_f is the transmittance of the filter pad-particle combination, and τ_b is the optical thickness of an identical blank (without particles) filter pad, then the additional optical thickness due to the particles is $\tau_p = \tau_f - \tau_b$. The absorption coefficient is the optical thickness times the path length. What is the path length? Imagine the particles in the water in a cylindrical tube of cross sectional area A_f and length ℓ with the light entering one end and exiting the other. The path length would be ℓ . Now a volume of fluid V_f was filtered through a filter of area A_f , so the effective path length is just $\ell = V_f/A_f$, i.e., this is the path the light would have take if the particles were suspended in an absorption cell between L1 and L2 in Figure 5.1. Thus, one would expect that $a_p = (\tau_f - \tau_b)/\ell$; however, the particles are embedded in the filter material (usually glass fiber filters) so the light propagating through

⁹The integrating cavity absorption meter described in Footnote 2, although very attractive because of its insensitivity to scattering, is not in general use.

the filter is diffuse rather than collimated. The extinction for diffuse light k_d is related to the extinction for collimated light k_c by $k_d = k_c / \langle \mu \rangle$, where $\langle \mu \rangle$ is the average cosine of the radiance distribution *within* the filter. (This is similar to the decay coefficient, K_V for the “vector” irradiance, $E_d - E_u$ described in Chapter 6, i.e., $K_V = a / \langle \mu \rangle$). If the light is totally diffuse, $\langle \mu \rangle = 0.5$, and assuming this is the case in the filter pad,¹⁰

$$a_p = \langle \mu \rangle \frac{\tau_f - \tau_b}{\ell} \approx 0.5 \frac{\tau_f - \tau_b}{\ell}.$$

In order to decompose the a_p into its component parts a_ϕ and a_{NAP} , it is necessary to remove one or the other of the components. Fortunately, it is possible to remove nearly all of the absorbing pigments within phytoplankton by soaking the filter pad in hot methanol, i.e., remove a_ϕ .¹¹ Measurement of τ_f for the Methanol-treated filter pad yields a_{NAP} and $a_\phi = a_p - a_{NAP}$.

Through the methods described in this section, it is possible to measure the absorption coefficient of each of the individual absorbing components (particulate and dissolved) present in Case 1 waters. In Case 2 waters, the same techniques allow a partial separation of the absorption into its components; however, in this case the non-algal components includes materials over and above detritus, e.g., resuspended minerals etc., and the dissolved component includes both phytoplankton-derived CDOM and terrestrial-originated CDOM.

5.3.2 Absorption of the Constituents of Case 1 waters

We now examine the absorption properties of constituents in Case 1 waters. Particle absorption is complex and we will only discuss it to the extent required for our purposes (however, we will discuss some theoretical aspects of the subject in a later section). In most studies the absorption coefficient is measured along with the concentration (weight per volume of water) of the photosynthetic pigment Chlorophyll *a* or the concentration (dry weight per unit volume of water) of particulate organic carbon (POC) at a number of stations and depth in the water column. Then statistical relationships are developed among the various quantities. The apparent “noise” in such relationships represents the natural variability in the particle population, i.e., different species of phytoplankton, etc.

¹⁰The factor $\langle \mu \rangle$ in the biological literature is usually denoted by $1/\beta$, where β is called the *path length amplification factor* (not to be confused with the volume scattering function).

¹¹This technique is usually referred to as the *Kishino* method..

5.3.2.1 Absorption of Particles of Case 1 waters

A large data set from representative waters around the globe in which a_ϕ , a_{NAP} , and a_p were measured, along with the concentration of Chlorophyll a , has been accumulated. Analysis shows that the particle absorption coefficient at 440 nm is statistically related to the concentration of Chlorophyll a (C) through

$$a_p(440) = A_p C^{0.635}, \quad (5.7)$$

where $A_p = 0.0520$, C is in mg/m^3 , and a_p is in m^{-1} . There is considerable noise in this relationship with A_p values falling between 0.03 and 0.08 with roughly uniform probability for $C < 5.0 \text{ mg}/\text{m}^3$. This absorption is apportioned between algae and non-algal particles (detritus) according to

$$a_\phi(440) = A_\phi C^{0.627} \quad \text{and} \quad a_{NAP}(440) = 0.0124 C^{0.724}, \quad (5.8)$$

with $A_\phi = 0.0378$ and ranging from ~ 0.2 to 0.7 for $C < 5 \text{ mg}/\text{m}^3$. In these expressions, the exponents of C are statistically indistinguishable. The spectral variations of a_ϕ and a_p can be reproduced through the statistical products

$$a_\phi(\lambda) = A_\phi(\lambda) C^{E_\phi(\lambda)} \quad \text{and} \quad a_p(\lambda) = A_p(\lambda) C^{E_p(\lambda)} \quad (5.9)$$

The various quantities in these equations are provided in Figure 5.10. The curves for $A_\phi(\lambda)$ and $A_p(\lambda)$ are, of course, $a_\phi(\lambda)$ and $a_p(\lambda)$ for $C = 1 \text{ mg}/\text{m}^3$.

The spectrum of a_{NAP} is totally different from $a_\phi(\lambda)$ which allows their separation. Figure 5.11 shows spectra of a_p , a_ϕ and a_{NAP} for a concentration of $1 \text{ mg}/\text{m}^3$. The spectrum of a_{NAP} is featureless, with absorption increasing exponentially toward the blue:

$$a_{NAP}(\lambda) = a_{NAP}(440) \exp[-S_{NAP} \times (\lambda - 440)], \quad (5.10)$$

where $S_{NAP} = 0.0011 \pm 0.002 \text{ nm}^{-1}$ and λ is in nm.

5.3.2.2 Absorption by CDOM in Case 1 waters

CDOM absorption is similar to the absorption by NAP, i.e., it is also featureless and increases exponentially into the blue with absorption given by an equation similar to Eq. (5.10), i.e.,

$$a_y(\lambda) = a_y(440) \exp[-S_y \times (\lambda - 440)], \quad (5.11)$$

but with a larger value of S , typically $S_y \sim 0.011 - 0.017 \text{ nm}^{-1}$. For Case 1 waters, $a_y(440) \lesssim 0.1 \text{ m}^{-1}$.

5.3.3 Scattering by Particles in Case 1 Waters

We now turn to the scattering properties deduced for particles in Case 1 waters. As with absorption, here we describe the results of measurements and leave the theoretical discussion of the measurements for a later section.

5.3.3.1 Total Scattering Coefficient of Particles in Case 1 Waters

Instruments have been developed to measure the total scattering coefficient (water plus particles). Early instruments attempted to measure b through β . An example is the $c - \beta$ meter shown schematically in Figure 5.1 which in principle allows measurement of b via measurement of β . Other instruments make use of the fact that most of the particle scattering is confined to small scattering angles, e.g., $\Theta \lesssim 20^\circ$. Thus, in Figure 5.12, if the detector (D1) is a photographic plate or a CCD, the volume scattering function at small angles can be retrieved (and thus b), while if it is simply a photocell, a signal related to b (in some complex manner) is detected. Another approach is provided by scattering theory, which suggests β at some small scattering angle ($4\text{--}10^\circ$) is proportional to b , i.e., the scattering phase function for particles in Case 1 waters is approximately invariant at some small angle.

More recently, however, with the development of accurate instruments for routine measurement of beam attenuation coefficient c and absorption coefficient a , most measurements of b have been carried out by measuring the beam attenuation coefficient and subtracting measurements of the absorption coefficient, i.e., $b = c - a$, or when the IOPs of the water itself are subtracted, $b_p = c_p - a_p - a_y$.¹² Such measurements can be carried out at any wavelength; however, often measurement is made only in the red, where a_p and a_y are small (Figure 5.11). Then, considering that the magnitude of c_p at 660 nm is $\sim 0.2 - 0.8 \text{ m}^{-1}$ for $C = 1 \text{ mg/m}^3$, while $a_p \sim 0.012 \text{ m}^{-1}$ and a_y is negligible, to a good approximation $c_p(660) \approx b_p(660)$. This latter relationship has been used often to infer particle scattering measurements from those of beam attenuation in the red without having to measure the absorption coefficient.

Figure 5.13 provides a summary of measurements in various waters (Case 1 and Case 2) carried out in the early to mid 1970's with b_p determined by the “earlier” methods

¹²One needs, however, to take care in assessing such measurements, as the value of c obtained has been shown to be significantly dependent on the acceptance angle of the c -meter (governed by the size of the aperture in the stop S_1 in Figure 5.1). The error is larger when large particles are present: much of their strongly forward-scattered light can reach the detector rendering a too-small value for c . Thus, a correction dependent on the particle size distribution and acceptance angle of the c -meter is required. The reader is encouraged to be skeptical of c measurements reported *without* addressing this effect.

above. $C_{C\&P}$ here refers to the concentrations of Chlorophyll *a* and a degradation product Phaeophytin *a* (both determined fluorometrically with the Phaeophytin contribution probably over estimated because of the interfering presence of Chlorophyll *b*).¹³ The most interesting aspect of this data is the fact that b_p at 550 nm is related to $C_{C\&P}$ nonlinearly. Regression analysis yields

$$b_p(550) = b_p^0 C_{C\&P}^{0.62}, \quad (5.12)$$

where $C_{C\&P}$ is in mg/m^3 , and for Case 1 waters $b_p^0 = 0.30 \text{ m}^{-1}$ but ranges from 0.12 to 0.45 m^{-1} with smaller values being appropriate to deeper samples. Clearly, even in Case 1 waters, chlorophyll-containing phytoplankton are not the only particulate component: there are living phytoplankton containing Chlorophyll *a*, detrital particles derived from phytoplankton (fragments of cells), zooplankton, etc. (In discussing absorption, these other particles were all lumped together as non-algal particles.) The concentration of these particles is measured by filtering a known volume of water. It is usually referred to as “seston” — S , the total mass per unit volume of suspended material as collected on pre-weighed filters after they are dried. Similar analysis relating b_p to S is provided in Figure 5.14, and shows the remarkably simple relationship

$$b_p(550) \approx S, \quad (5.13)$$

where $b_p(550)$ is in m^{-1} and S is in g/m^3 . Consistency between these requires that

$$S \propto C_{C\&P}^{0.62} \quad \text{or} \quad \frac{C_{C\&P}}{S} \propto C_{C\&P}^{0.38}, \quad (5.14)$$

i.e., the chlorophyll concentration increases faster than the seston concentration: more and more of the seston is related to the component containing chlorophyll as $C_{C\&P}$ increases.

The scatter of points in Figure 5.13 even in the Case 1 waters (the region between the two lines) is roughly a factor of 3 (+50% to −100%). With the newer techniques it was believed that this “noise” in the $b_p - C_{C\&P}$ relationship would be considerably reduced; however, this is not the case. Using measurements of c and a in Case 1 waters at 660 nm, the new relationship is¹⁴

$$b_p(660) = 0.347 C^{0.766} \quad \text{or} \quad b_p(550) \approx 0.416 C^{0.766}, \quad (5.15)$$

where C is the concentration of Chlorophyll *a* determined by HPLC.¹⁵ To derive this relationship for 550 nm it was assumed that there is a λ^{-1} variation in $b_p(\lambda)$. The “noise”

¹³The quantity $C_{C\&P}$ is usually referred to as the “pigment concentration.” In later chapters it will be indicated by C_P .

¹⁴But, note the caveat discussed in Footnote 11.

¹⁵This is an acronym for *high-performance liquid chromatography* also called *high-pressure liquid chromatography*. It is a method of analytical chemistry for analyzing the components of a mixture. Here it is used to analyze the concentrations of the various phytoplankton pigments present in a water sample, particularly in this case, Chlorophyll *a*. Before HPLC, fluorescence techniques were used to measure Chlorophyll *a*, but these could not separate Chlorophyll *a* from its degradation product Phaeophytin *a*, and hence the term *pigment concentration*, $C_{C\&P}$, was reported rather than Chlorophyll *a*.

in this relationship is only somewhat reduced from the earlier $b_p - C_{C\&P}$ relationship: roughly by a factor of 2 (+50% to -60%).

Biological processes in the ocean render it a sink for CO_2 . In recognition of the importance of phytoplankton photosynthesis in the global carbon cycle, the interest in seston has been replaced by interest in particulate organic carbon (POC). POC is the mass of particulate carbon in a given mass or volume of water. It is this form of carbon, produced through photosynthesis, that sinks through the upper layers of the oceans to the deep ocean where it is sequestered for 500 – 1000 years. To measure POC, particles are filtered from water and the carbon concentration determined by chemical means. Typical units of POC are mg C/m^3 ; however another common unit is $\mu\text{mole C/liter}$ ($1 \text{ mg C/m}^3 = 0.0833 \mu\text{mole C/liter}$).

If we assume that the particulate matter in Case 1 water is almost all organic (as it is derived from phytoplankton), then there should be a relationship between POC and seston.¹⁶ If the POC is in the form of carbohydrates, i.e., $\text{C}_n(\text{H}_2\text{O})_n$, then the mass of POC is 0.3 times the mass of seston, or the mass of seston is 3.3 times the mass of Carbon.¹⁷ If POC is measured in mg C/m^3 (*note, mg C not g C*), Eq. (5.13) suggests

$$b_p(550) = 0.0033 \times \text{POC}, \quad \text{or} \quad b_p(660) = 0.0027 \times \text{POC},$$

where b_p is in m^{-1} , and a λ^{-1} dependence is assumed in converting from 550 nm to 660 nm. This agrees well with direct measurements that yield (in the same units)

$$c_p(660) \approx C_{\text{POC}} \times \text{POC},$$

where $0.0007 \lesssim C_{\text{POC}} \lesssim 0.0045$, with a mean near 0.003.

The spectral variation of b_p is usually (as above) assumed to be proportional to λ^{-1} . As we shall see later in this chapter, theory suggests, and experiments verify, that c_p for the residue of marine particles is rather featureless and varies as λ^{-p} , where p is determined mostly by the size distribution of the particles. If the particles have a featureless absorption spectrum, e.g., detrital particles with a spectrum given by $a_{\text{NAP}}(\lambda)$ (Eq. (5.10)), then the spectrum of b_p will also be featureless. However, phytoplankton have a spectrum similar to that shown in Figure 5.10, so if c_ϕ is featureless for phytoplankton (as it usually is), then b_ϕ (i.e., $c_\phi - a_\phi$) for phytoplankton will decrease in spectral regions where absorption is significant, i.e., in the blue and the red. This effect is actually easily observed. When a

¹⁶An exception to this assumption is a coccolithophore bloom where a significant amount of Carbon resides in the form of CaCO_3 — the composition of the coccolith shell surrounding the cell itself. This is particulate *inorganic* carbon (PIC) .

¹⁷A generic chemical composition for organic matter in the oceans is $(\text{CH}_2\text{O})_{106}(\text{NH}_3)_{16}\text{H}_3\text{PO}_4$ for which the fraction of mass of Carbon is 0.276. The ratios C:N:P are 106:16:1, which are known as the Redfield ratios, are similar to the C:N:P ratios of dissolved matter in the ocean.

spectrum of c for a culture of phytoplankton is observed, the absorption features can only weakly be seen; however, when some of the scattered light from the beam is collected, the absorption features are more vivid, implying that the scattering is increased in the pigment absorption bands. We shall explore this spectral dependence in a later section.

5.3.3.2 Volume Scattering Function of Particles in Case 1 Waters

Prior to the late 1990's there were far fewer measurements of the volume scattering function $\beta_p(\Theta)$, or the scattering phase function $P_p(\Theta) = 4\pi\beta(\Theta)/b_p$, than the total scattering coefficient. The simple reason for this is the complexity of the measurement (Figure 5.1), which in the past required moving parts of the instrument at large angles, and a separate instrument for small scattering angles (Figure 5.12). In addition, most early measurements were made before the Case 1 – Case 2 distinction was developed, and many of the measurements were not in Case 1 waters. Even now there are virtually no measurements that separate the scattering by living and detrital particles.

One of the earliest measurements of the phase function covering the full range of Θ was carried out in the early 70's by Petzold and referred to here as “Phase-T,” “Petzold Phase-T” or just “Petzold.” Phase-T is provided in Figure 5.15. This was not actually measured in Case 1 waters (in fact the water was quit turbid, $b_p \sim 1.6 \text{ m}^{-1}$, which actually made the measurement more accurate); however, because of the paucity of VSF measurements, it has frequently been used to characterize the scattering by particles *even in* Case 1 waters. For Phase-T the scattering at small angles ($\Theta \lesssim 10^\circ$) follows a power law:

$$P_p(\Theta) \propto \frac{1}{\Theta^m},$$

with $m \approx 1.4$. Thus, the value of the phase function at 0.1° is roughly 3000 times that at 10° . Inadequate as they may be, these measurements clearly demonstrate that marine particles scatter very strongly in the forward direction with only a small probability ($< 2\%$) for scattering at angles between 90° and 180° . Recently (2009) Sullivan and Twardowski (S&T) reported the results of several thousand measurements of the VSF over the angular range 10° to 170° , for a variety of waters. Analysis of the measurements showed that the particulate component of most of the measured VSFs have a similar variation with scattering angle for $\Theta > 90^\circ$, i.e., the “shape” of $\beta_p(\Theta)$ for $\Theta > 90^\circ$ was similar over most of the measurements, even though the numerical value of $\beta_p(90^\circ)$ showed a variation of over 2 orders of magnitude. Here, we refer to the normalized shape of their β_p as the “WetLab Average” (abbreviated WetLab Avg.). This analysis suggests that, at least for a first order attempt to model marine particle scattering, a universal shape (the WetLab Avg.) should be assumed for the particle phase function $P(\Theta)$ for $\Theta > 90^\circ$. As the Petzold measurements are the only ones available for really small scattering angles, it is suggested

that a new phase function should be adopted for modeling marine particles. This phase function consists of Phase-T for $\Theta < 90^\circ$, and the WetLab Avg. shape for $\Theta > 90^\circ$. We refer to this phase function as “Petzmas.”¹⁸

Figure 5.16 compares Phase-T with some of the other phase functions used in this work: the “O99” phase function is that of the “Oceanic” model for aerosols with relative humidity of 99% (Chapter 4); the “HG” curve, a one-parameter analytic approximation to the phase function given by

$$P_p(\Theta) = \frac{(1 - g^2)}{(1 + g^2 - 2g \cos \Theta)^{3/2}}$$

with $g = 0.9$ here (HG stands for Henyey-Greenstein); and “WetLab Avg.,” here normalized to have the same value as Phase-T at $\Theta = 90^\circ$. The main messages to take for these comparisons are (1) the similarities of the particle and aerosol phase functions suggest that the particles occupy similar size ranges, (2) in contrast to the aerosol, the marine particles do not show any structure at large scattering angles which indicates the particles are likely to be absorbing (as we already know) and non-spherical, and (3) the HG phase function is a good approximation to the particle phase function *except* for $\Theta \lesssim 10 - 20^\circ$. A more detailed comparison between Phase-T and the WetLab Avg. at $\Theta \geq 90^\circ$ is provided in Figure 5.17. Note that WetLab Avg. has less variation with Θ than Phase-T for $\Theta > 90^\circ$. As described above, when the WetLab Avg. is extended to small angles by requiring that it be equivalent to Phase-T, the result we denote by Petzmas. The two particle phase functions, Phase-T and Petzmas, are used extensively later in developing relationships between the IOPs and the apparent optical properties (AOPs) of water bodies (Chapter 6); however, we shall see below that a further modification of the phase functions is required to model scattering by marine particles.

Finally, another two-parameter phase function that is capable of providing an analytical approximation to both Phase-T and Petzmas is the Fournier-Forand (FF) phase function. It is the default phase function used in the HydroLight radiative transfer code. It is described in some detail in Appendix 1.

5.3.3.3 The Backscattering Coefficient of Particles in Case 1 Waters

The backscattering coefficient for particles is defined by

$$b_{bp} \triangleq 2\pi \int_{\pi/2}^{\pi} \beta_p(\Theta) \sin \Theta d\Theta,$$

¹⁸This term originates from combining Petzold and MASCOT (MultiAngle SCattering Optical Tool), the instrument used to carry out the S&T measurements.

where it has been assumed that, for a given Θ , the volume scattering function is independent of the azimuth angle around the incident beam.¹⁹ The total backscattering coefficient of the medium is $b_b = b_{bw} + b_{bp}$, where b_{bw} is the backscattering coefficient of water ($b_w/2$). The backscattering coefficient is important in remote sensing because the radiance from the Sun and sky backscattered out of the water is proportional to $b_b/(a + b_b)$, where a is the absorption coefficient.

The most direct method of measuring the backscattering coefficient is to measure $\beta(\Theta)$ and perform the above integration. This is the technique used in the earliest measurements. A second is to measure $\beta(\Theta)$ at a few scattering angles and perform a fit to an analytic formula in Θ , and then integrate the formula. A good approximation to the shape of the scattering phase function can be obtained using the formula

$$\beta_{BZ}(\Theta) = \frac{\beta(90^\circ)}{(1 - e_f \cos \Theta)^4 (1 + e_b \cos \Theta)^4}.$$

Measuring $\beta(45^\circ)$, $\beta(90^\circ)$ and $\beta(135^\circ)$, for which commercial laboratory instruments were available, allowed estimation of e_f and e_b , and integration of β_{BZ} over the backward hemisphere provided b_b . An example of the quality of the fit of β_{BZ} to the Petzmas phase function is presented in Figure 5.18. Finally, newer instruments estimate b_{bp} by measuring $\beta(\Theta)$ at a single scattering angle ($110^\circ \lesssim \Theta \lesssim 120^\circ$), where the $b_{bp} - \beta_p$ relationship displays the least variability. This angular range was first developed from theoretical computations and then confirmed by experimental measurements (e.g., those measurements leading to WetLab Avg.). An alternative to actually measuring $b_{bp}(\lambda)$ is to try to estimate it from measurements of the irradiance reflectance R (also proportional to $b_b/(a + b_b)$). Methods for effecting this are described in Chapter 6.

Oceanic measurements in oligotrophic surface waters suggest that $b_{bp} \propto b_p$ with a proportionality constant of ~ 0.01 , i.e., $B_p \triangleq b_{bp}/b_p \sim 0.01$, but with a significant range of variation 0.006 to 0.015. For reference, the backscattering probability (B_p) for the Petzmas phase function is 0.0181. For higher concentrations of Chlorophyll *a*, B_p appears to decrease because an increase in C usually means an increase in phytoplankton concentration relative to detritus (Eq. (5.14)), and the larger phytoplankton particles (compared to detrital particles) are less efficient backscatterers. Indeed, the examination of the spectrum of irradiance reflectance (R) in a variety of waters (Case 1) and a large range in C suggests that

$$B_p \approx 0.002 + 0.01 \left(\frac{1}{2} - \frac{1}{4} \log_{10} C \right), \quad (5.16)$$

approximately independent of wavelength. This gives $B_p = 0.012$ for $C = 0.01$ mg/m³ and 0.002 for $C = 100$ mg/m³. Combining this with Eq. (5.15) and fitting the result to a

¹⁹This assumption is valid for spherical particles and for a suspension of non-spherical particles, e.g., cylinders, spheroids, etc., in *random orientation*.

power-law in C yields

$$b_{bp}(550) = 0.00279 C^{0.624}.$$

Note that the closeness of the exponents of C for b_p and b_{bp} suggest that if one is proportional to POC, the other is also. Recognizing that larger particles like phytoplankton tend to scatter with a smaller spectral variation, and that the high- C waters have more phytoplankton (larger particles) compared to detritus (smaller particles) than the low- C waters, improved agreement between measured and reconstructed spectra of R is obtained if the spectral variation of b_p , rather than assumed to be simply proportional to λ^{-1} , is *guessed*, but reasonably though, to vary with C according to

$$b_p(\lambda, C) = b_p(550, C) \left(\frac{\lambda}{550} \right)^v,$$

with $v = 0.5(\log_{10} C - 0.3)$ for $C < 2 \text{ mg/m}^3$, and $v = 0$ otherwise. Accordingly, the particle scattering coefficient varies as $\lambda^{-1.15}$ for $C = 0.01 \text{ mg/m}^3$ and λ^0 for $C \geq 2 \text{ mg/m}^3$. Thus,

$$b_{bp}(\lambda, C) = \{0.002 + 0.01 (\frac{1}{2} - \frac{1}{4} \log_{10} C)\} b_p(550, C) \left(\frac{\lambda}{550} \right)^v.$$

5.3.3.4 Polarization of Scattering in Case 1 Waters

We complete our discussion of the observed scattering properties of Case 1 waters with examples of the polarization properties of the scattered light as manifest in the scattering Mueller matrix (Chapter 3). Write the volume scattering Mueller matrix $\beta(\Theta)$ as

$$\beta(\Theta) = \begin{pmatrix} \beta_{11}(\Theta) & \beta_{12}(\Theta) & \beta_{13}(\Theta) & \beta_{14}(\Theta) \\ \beta_{21}(\Theta) & \beta_{22}(\Theta) & \beta_{23}(\Theta) & \beta_{24}(\Theta) \\ \beta_{31}(\Theta) & \beta_{32}(\Theta) & \beta_{33}(\Theta) & \beta_{34}(\Theta) \\ \beta_{41}(\Theta) & \beta_{42}(\Theta) & \beta_{43}(\Theta) & \beta_{44}(\Theta) \end{pmatrix},$$

and normalize it by dividing every element by $\beta_{11}(\Theta)$ to form

$$\frac{\beta(\Theta)}{\beta_{11}(\Theta)} = \begin{pmatrix} 1 & N_{12}(\Theta) & N_{13}(\Theta) & N_{14}(\Theta) \\ N_{21}(\Theta) & N_{22}(\Theta) & N_{23}(\Theta) & N_{24}(\Theta) \\ N_{31}(\Theta) & N_{32}(\Theta) & N_{33}(\Theta) & N_{34}(\Theta) \\ N_{41}(\Theta) & N_{42}(\Theta) & N_{43}(\Theta) & N_{44}(\Theta) \end{pmatrix} = \frac{\mathbf{P}(\Theta)}{P_{11}(\Theta)},$$

where $N_{ij} = \beta_{ij}/\beta_{11}$ and \mathbf{P} and P_{11} are, respectively, the Mueller phase matrix and its 11 element (also called the phase function). Measurements of the Mueller matrix in the

open ocean show that, within the accuracy of the measurements, the matrix has a block diagonal structure, i.e.,

$$\frac{\mathbf{P}(\Theta)}{\beta_{11}(\Theta)} = \begin{pmatrix} 1 & N_{12}(\Theta) & 0 & 0 \\ N_{21}(\Theta) & N_{22}(\Theta) & 0 & 0 \\ 0 & 0 & N_{33}(\Theta) & 0 \\ 0 & 0 & 0 & N_{44}(\Theta) \end{pmatrix} = \frac{\mathbf{P}(\Theta)}{P_{11}(\Theta)}. \quad (5.17)$$

This overall structure of the scattering matrix is similar to that of Rayleigh scattering, as well as to scattering by larger particles in the Rayleigh-Gans approximation.²⁰ In fact, the matrix elements N_{ij} are quantitatively similar to those in the Rayleigh-Gans approximation for anisotropic particles in random orientation.

Figure 5.19 shows the average N_{12} and N_{21} for sixty samples from Atlantic and Pacific waters. Included for comparison are the corresponding Rayleigh-Gans elements for an isotropic ($\delta = 0$, dashed curve) and an anisotropic ($\delta = 0.2$, indicated by solid triangles) particle of arbitrary shape and random orientation. The most obvious feature of the phase matrix elements shown in Figure 5.19 is the fact that $N_{12} \approx N_{21}$. (Note that $N_{21} = N_{12}$ for homogeneous spherical particles.) In addition, we see that N_{12} and N_{21} can be approximated by anisotropic Rayleigh-Gans scattering with a depolarization ratio of approximately 0.2.

Figure 5.20 shows the average N_{33} and N_{44} along with that for Rayleigh-Gans scattering. Again, we see $N_{33} \approx N_{44}$. Both N_{33} and N_{44} follow their isotropic Rayleigh-Gans counterpart reasonably well in forward directions, while both are larger than the corresponding element in backward directions. Unlike N_{33} , N_{44} for anisotropic Rayleigh-Gans scattering depends on the depolarization factor δ . In the case of $\delta = 0.2$, for which the Rayleigh-Gans scattering element fit the oceanic averaged N_{12} and N_{21} quite well (Figure 5.19), the agreement with the measured N_{44} is not as good, and is especially poor for $\Theta < 90^\circ$. However, it should be noted that N_{44} is of little importance in radiative transfer in the ocean-atmosphere system, as it governs the scattering of elliptically polarized light into elliptically polarized light — a process that is of little importance in environmental optics.

The oceanic-averaged element N_{22} is provided in Figure 5.21. For isotropic Rayleigh, isotropic Rayleigh-Gans, and scattering by homogeneous isotropic spherical particles of any size, $N_{22} = N_{11} = 1$; however, in Figure 5.21 one sees a departure of N_{22} from

²⁰The elements $N_{ij}(\Theta)$ for randomly oriented anisotropic Rayleigh scattering particles, and for scattering by a collection of randomly oriented anisotropic homogeneous particles of arbitrary shape, but constant values of α_x , α_y , and α_z , in the Rayleigh-Gans approximation, are identical. Note that this is not to say that the full scattering matrices are identical: only the scattering matrices normalized by β_{11} , i.e., β_{ij}/β_{11} . The latter provide all of the polarization properties of the scattering.

unity. This is thought to be the result of the non-sphericity of the particles, as exact scattering computations of $N_{ij}(\Theta)$ for randomly oriented isotropic spheroids and cylinders show $N_{22}(\Theta) < 1$ for at least some range of scattering angles.

As we have seen above, certain matrix elements in Eq. (5.17) for the oceanic average can be well-represented by the results of the Rayleigh-Gans approximation for particles with anisotropic (but constant) polarizability (Figures 5.19-5.21). For example, an excellent approximation to N_{12} is the Rayleigh-Gans result

$$N_{12}(\Theta) = N_{21}(\Theta) = -\frac{(1 - \delta) \sin^2 \Theta}{(1 + \delta) + (1 - \delta) \cos^2 \Theta},$$

and, while not as good as for N_{12} and N_{21} ,

$$N_{22}(\Theta) = \frac{1 + \cos^2 \Theta}{1 + \cos^2 \Theta + 2\delta/(1 - \delta)}, \quad N_{33}(\Theta) \approx \frac{2 \cos \Theta}{1 + \cos^2 \Theta},$$

and

$$N_{44}(\Theta) = \frac{1 - 2\delta}{1 + \delta/2} N_{33}(\Theta).$$

In these equations δ , the depolarization ratio, is given by

$$\delta = \frac{1 - DOP(90^\circ)}{1 + DOP(90^\circ)},$$

where $DOP(90^\circ) = -N_{12}(90^\circ)$. The fact that the normalized scattering matrix of particulates in the open ocean resembles that scattering in the Rayleigh-Gans approximation is a manifestation of the low refractive index of such particulates — a requirement for the validity of the Rayleigh-Gans approximation.

Samples of cultured phytoplankton have normalized scattering matrices that are remarkably similar to the oceanic averages. For example, Figure 5.22 provides the result of measurement of the degree of (linear) polarization, $-N_{12} = P_{12}(\Theta)/P_{11}(\Theta)$, for several species of phytoplankton. On the figure the species labeled “S” are approximately spherical in shape, while those labeled with “F” have a filamentous structure, i.e., cylindrical in shape. The species Sb, Sc, Fc, and Fe contain gas vacuoles, while the others do not. The oceanic average $-N_{12}$ from Figure 5.19 is presented as well (the thick black curve). Note that there is no characteristic of the curves that distinguishes the shape or the presence (or absence) of gas vacuoles, i.e., there appear to be no features in the DOP data for various species that are characteristic of their shape (or size) or internal structure.

5.3.3.5 Empirical Model for the Particle Phase Function in Case 1 Waters

The fact that measurements show that B_p varies with C demands that the particle phase function $P_p(\Theta)$ must also vary with C . However, the variation with Θ for $\Theta > 90^\circ$ must still retain the WetLab Average shape. A simple (empirical) two component model that can accomplish the desired behavior of $P_p(\Theta)$, and satisfy the variation of B_p with C , is easy to develop.

We start with the Petzmas phase function, which has the desired angular behavior for $\Theta > 90^\circ$. Then, we can produce a phase function that has any desired backscattering probability (less than Phase-T) by increasing the Phase-T small angle scattering by an appropriate amount. Table 5.3 provides the factor f that the Petzmas phase function must be increased by in order to provide the given backscattering probability B_p . For completeness, we have also included the factors for Phase-T as well.

Table 5.3: The factor f by which the indicated phase function must be augmented at scattering angles between 0 and 2° in order to yield the given backscattering probability B_p . The result after the multiplication must be renormalized so that the integral of the modified $P(\Theta)$ over all solid angles is still 4π .

B_p	Petzmas	Petzold
0.0019	35.862	35.522
0.0020	33.917	33.594
0.0050	11.741	11.612
0.0070	7.517	7.426
0.0100	4.348	4.285
0.0140	2.237	2.192
0.0181	1.000	1.000

Ideally, for a two component model one would like the detrital component (smaller particles) to have a B -value a little higher than 0.012, and the phytoplankton component (larger particles) a little smaller than 0.002 (Section 5.3.3.3). The choice $B_{pn} = 0.0140$, where “ pn ” stands for the nonliving (detrital) particle component and $B_{pl} = 0.0019$, where “ pl ” stands for the living (phytoplankton) particle component meet these desires. Letting α_n represent the fraction of B_p due to the nonliving component, then

$$B_p = \alpha_n B_{pn} + (1 - \alpha_n) B_{pl}.$$

For consistency with Eq. (5.16),

$$\alpha_n = \frac{0.002 + 0.01(1/2 - 1/4 \log_{10} C) - B_{pl}}{B_{pn} - B_{pl}}$$

or for our choice of the backscattering probabilities,

$$\alpha_n = 0.00826 + 0.826(1/2 - 1/4 \log_{10} C).$$

The particle phase function for a given C is

$$P_p(\Theta) = \alpha_n P_{pn}(\Theta) + (1 - \alpha_n) P_{pl}(\Theta),$$

where P_{pn} and P_{pl} are the nonliving and living particle phase functions developed using the results in Table 5.3. These end member phase functions, P_{pn} and P_{pl} , are provided in Figure 5.23. To reiterate, this proposed phase function model (1) expresses what we expect for the phase functions of detritus and phytoplankton, and (2) directly provides the modeled B_p , i.e., Eq. (5.16), as a function of C .

5.4 Influence of Single Particle Properties – Size, Composition, Shape – on IOPs

Now that we have described the experimental studies regarding the IOPs of the constituents of Case 1 waters, we will try to develop an understanding of some of the experimental observations, based on the properties of the individual particles, through the use of light scattering theory. In addition we will describe how to employ the measurements to develop a more fundamental understanding of some of the particle properties.

The physical properties determining particle IOPs are their size, composition and shape. The goal of this section is to gain an understanding of how these factors influence absorption, scattering, and backscattering. We have already examined some of these questions in Chapters 3 and 4 in the case of spherical particles. Clearly, the shape of most phytoplankton departs considerably from spherical; however, examining the IOPs for spheres represents a first step in understanding the optics of phytoplankton. Here we will apply that theory to particles in Case 1 waters: phytoplankton and detrital particles. We will use Mie theory (or its anomalous diffraction approximation) for spherical particles, and the DDA for cylindrically-shaped particles, when we try to understand some of the effects of departures from sphericity.

5.4.1 The Absorption Coefficient

Our main particle of interest, phytoplankton, contains absorbing pigments (Chlorophyll a , etc.). For electromagnetic waves, absorption is manifest in the imaginary part of the

refractive index. The complex index of refraction is $m = m_r + im_i$, with m_r governing the reflection and refraction from surfaces and m_i the loss in irradiance along the path. As developed in Chapter 1, the imaginary part, m_i is related to the absorption coefficient of the pigment material a_i by

$$m_i = \frac{a_i \lambda_{\text{Vacuum}}}{4\pi}. \quad (5.18)$$

Note that, here the subscript i on m_i stands for “imaginary,” i.e., the imaginary part of the complex refractive index, while, on a_i it stands for “internal,” i.e., it is the absorption coefficient of the material inside the particle. The particles in our case are suspended in water, so we must use the indices relative to water, i.e., $\tilde{m} = m/m_{\text{Water}}$, then

$$\tilde{m}_i = \frac{a_i \lambda_{\text{Vacuum}}}{4\pi m_{\text{Water}}} = \frac{a_i \lambda_{\text{Water}}}{4\pi}. \quad (5.19)$$

Thus, given the absorption coefficient of the absorbing material (at a given wavelength), the value of \tilde{m}_i required to apply scattering theory is provided. Consider a single absorbing component, and let c_i be its volume concentration within the particle (mass of pigment \div particle volume). The absorption coefficient can be written $a_i = a_i^* c_i$, where a_i^* is called the *specific* absorption coefficient with respect to the absorbing pigment.²¹ For a volume V of solution containing N such particles, the total concentration of the pigment is

$$C = \frac{N}{V} v c_i, \quad (5.20)$$

where v is the volume of one particle. Combining these ideas we have

$$a_i = \frac{V a_i^*}{N v} C \quad \text{and} \quad \tilde{m}_i = \left(\frac{V a_i^*}{N v} \right) \left(\frac{C \lambda_{\text{Water}}}{4\pi} \right).$$

Now, in general we need to know \tilde{m}_r as well as \tilde{m}_i to apply Mie scattering theory (spherical particles) to determine the absorption properties of the solution; however, knowing \tilde{m}_r is not necessary if we use the van de Hulst approximation to Mie theory (Chapter 3). In this approximation we can compute the absorption efficiency of the particle, $Q_a = \sigma_a/A_p$, where σ_a is the absorption cross section²² of the particle and A_p is the particle's projected area on a plane normal to the direction of the incident beam (πR^2 , with R the radius), from a knowledge of \tilde{m}_i alone. In this approximation,

$$Q_a = 1 + 2 \frac{\exp(-\rho')}{\rho'} - 2 \frac{1 - \exp(-\rho')}{\rho'^2}, \quad (5.21)$$

²¹The specific absorption coefficient is also called the absorption cross section, with units m^2/kg or an equivalent.

²²Recall that the absorption cross section σ_a is related to the absorption coefficient a through the number density of particles n : $a = n\sigma_a$.

where

$$\rho' = \frac{4\pi D}{\lambda_{\text{Water}}} \tilde{m}_i = \frac{4\pi D}{\lambda_{\text{Water}}} \frac{a_i \lambda_{\text{Water}}}{4\pi} = D a_i, \quad (5.22)$$

and D is the diameter of the particle. Thus, ρ' is the absorption *optical thickness* associated with a light ray that would pass through the center of the particle. Clearly, Q_a approaches unity as ρ' becomes large, meaning that the absorption cross section is just the projected area of the sphere. Figure 5.24 provides Q_a as a function of ρ' .

The particle absorption coefficient a_{Part} of a sample containing N particles in a volume V is

$$a_{\text{Part}} = \frac{N}{V} \sigma_a = \frac{N}{V} \frac{\pi D^2}{4} Q_a.$$

If the concentration of the pigment in V is C , and the specific (to C) absorption coefficient of the collection of particles is defined as $a_{\text{Part}}^* \triangleq a_{\text{Part}}/C$, then

$$a_{\text{Part}}^* = \frac{N}{V} \frac{\pi D^2}{4C} Q_a = \frac{N}{V} \frac{\pi D^2}{4c_i(N/V)v_i} Q_a = \frac{3}{2} \frac{a_i}{c_i} \frac{Q_a}{\rho'} = \frac{3}{2} a_i^* \frac{Q_a}{\rho'}$$

or

$$\frac{a_{\text{Part}}^*}{a_i^*} = \frac{3}{2} \frac{Q_a}{\rho'}. \quad (5.23)$$

The quantity Q_a/ρ' is plotted in Figure 5.25. The figure shows the relationship between a_{Part}^*/a_i^* and the internal concentration c_i , or equivalently the internal absorption coefficient. For a given particle diameter D , the greater the internal pigment concentration, the smaller the ratio of a_{Part}^* to a_i^* . This makes sense physically as can be seen if we consider two extreme cases with the same C in a large macroscopic volume V : (1) the absorbing pigment is within a large number of very small particles ($D \rightarrow 0$); and (2) the absorbing pigment is all within a single large particle. Let the incident beam be cylindrical with cross sectional A . Then if L is the thickness of the sample, $V = AL$. The first case is equivalent to having the pigment uniformly dispersed within the volume, and the absorption coefficient of V , $a_{\text{Sample 1}}$ is just $a_i^* C$. Thus, in the first case $a_{\text{Sample 1}}^* = a_i^*$, and the optical thickness of the sample is $a_{\text{Sample 1}} L = a_i L$, which need not be small compared to unity.

The second case is more complicated. For simplicity, let us assume that c_i within the particle is so high that $c_i a_i^* D \gg 1$, then all of the light incident on the particle is absorbed.²³ Then the fraction of light removed from the beam is the scattering cross section of the particle divided by A . This fraction (assumed small because $\pi D^2/4 \ll A$) is the absorption coefficient of the sample $a_{\text{Sample 2}}$ times L , i.e.,

$$a_{\text{Sample 2}} L = \frac{\pi D^2}{4A} \quad \text{or} \quad a_{\text{Sample 2}}^* = \frac{\pi D^2}{4ALC}.$$

²³This assumption is not necessary, but simplifies matters. If we do not make it, then the final result must be multiplied by $1 - \exp(-2a_i D/3)$, where $2D/3$ is the mean path of photons through the sphere. This factor is the fraction of photons, incident on the *sphere*, that are absorbed.

Therefore,

$$\frac{a_{\text{Sample } 2}^*}{a_i^*} = \frac{\pi D^2}{4ALCa_i^*} = \frac{\pi D^2}{4A} \frac{1}{a_{\text{Sample } 1} L}$$

which, because $\pi D^2/4 \ll A$, is $\ll 1$. Thus, in the first case $a_{\text{Sample } 1}^*/a_i^* \approx 1$, and in the second $a_{\text{Sample } 2}^*/a_i^* \ll 1$. These predictions follow Eq. (5.23).

5.4.1.1 Applications

These simple formulae can be applied to the analysis of actual absorption data to obtain a deeper understanding of their properties. Here we examine several applications: (1) estimate of the spectrum of a_i and of \tilde{m}_i ; (2) demonstration of the so-called “package effect;” and (3) possibly understanding some aspects of the detrital absorption spectrum.

5.4.1.1.1 Estimation of a_i and \tilde{m}_i from a_{Part}

Suppose that we measure the absorption spectrum $a_{\text{Part}}(\lambda)$ in a given sample of, say, a cultured species of phytoplankton, which we assume to be of uniform size. (Here, and henceforth in this chapter by the symbol λ we mean λ_{Vacuum} .) In addition, assume we also measure their diameter (assumed spherical, but see below) and the number per unit volume ($n = N/V$, where N is the number in V), both of which can be determined with a Coulter counter.²⁴ Then we can determine $Q_a(\lambda) = 4a_{\text{Part}}(\lambda)V/N\pi D^2$, and using Figure 5.24 or Eqs. (5.21) and (5.22), we can derive ρ' , and thus $a_i(\lambda)$ the true absorption spectrum of the absorbing pigments within the cells. From this we can determine $\tilde{m}_i(\lambda)$ from

$$\tilde{m}_i(\lambda) = \frac{\lambda a_i(\lambda)}{4\pi m_{\text{Water}}}.$$

This is the method used to estimate the $\tilde{m}_i(\lambda)$ from particle absorption measurements.

To get an idea of the range of values for \tilde{m}_i that we can expect, we need to look at absorption spectra for phytoplankton pigments. Figure 5.26 provides the specific absorption coefficient (specific to the concentration of each pigment) of several absorbing pigments that are found in phytoplankton, and Figure 5.27 provides the specific absorption coefficient of Chlorophyll *a*. Chlorophyll *a* is the principal photosynthetic pigment and the dominate absorber in spectral regions near 440 and 675 nm. Measurement of the Chlorophyll *a* concentration in addition to the measurements of N and D , which provides v , determines c_i for Chlorophyll *a*.²⁵ It is found that the intercellular concentration c_i is in the range 0.01 to

²⁴The Coulter counter is a commercial resistive pulse particle counter that measures the volume of individual particles suspended in a conducting fluid, e.g., sea water. It is described in Chapter 4.

²⁵Since the Coulter counter measures the particle volume, if the cells are all the same size and shape, such measurements (along with C) yield c_i *without* even assuming that the particles are spherical.

15 kg/m³, or in units more appropriate for Figure 5.27, 0.01×10^6 to 15×10^6 mg/m³. Near 440 nm, $a_i^* \approx 0.025$ m²/mg, so a_i there is in the range 250 to 3.75×10^5 m⁻¹, or 2.5×10^{-4} to 0.375 μm^{-1} . Inserting these into Eq. (5.19) yields $6.6 \times 10^{-6} \lesssim \tilde{m}_i \lesssim 0.01$ at 440 nm. Thus, considering the pigment Chlorophyll *a*, for modeling phytoplankton scattering and absorption, it is likely unreasonable to employ an \tilde{m}_i that exceeds 0.01 by very much.

In general, as can be seen in Figure 5.26, there are several pigments that contribute to the particle absorption. How are these handled? Let $(c_i)_j$ be the concentration of the j^{th} pigment and $(a_i^*)_j$ the specific absorption coefficient of the j^{th} pigment. Then the internal absorption coefficient is

$$a_i = \sum_j (c_i)_j (a_i^*)_j.$$

Given $(a_i^*)_j$ (Figure 5.26) and measurement of a_{part} at several wavelengths, the $(c_i)_j$ could be estimated from linear regression. Or conversely, given $(c_i)_j$, the a_i spectrum could be computed and $\tilde{m}_i(\lambda)$ determined:²⁶

$$\tilde{m}_i(\lambda) = \frac{\lambda}{4\pi m_{\text{Water}}} \sum_j (c_i)_j (a_i^*(\lambda))_j. \quad (5.24)$$

Interestingly, Eq. (5.24) can also be used to learn something about $\tilde{m}_r(\lambda)$. Recall from Chapter 1 that if the spectrum can be decomposed into N absorption features, Eqs. (1.32) relate the real and imaginary parts of the refractive index (here relative to water):

$$\tilde{m}_i = \sum_{i=1}^N \frac{\tilde{C}_i}{1 + \eta_i^2} \quad \text{and} \quad \tilde{m}_r - 1 = \tilde{C}_0 + \sum_{i=1}^N \frac{\tilde{C}_i \eta_i}{1 + \eta_i^2}. \quad (5.25)$$

In these equations, the \tilde{C} 's are constants, $\eta_i = 2(\omega_i - \omega)/\gamma_i$, $\omega_i = 2\pi c_\ell/\lambda_i$, $\omega = 2\pi c_\ell/\lambda$ and γ_i is the “width” of the i^{th} absorption feature. The absorption features are centered on the λ_i 's. We achieve the same result by letting $\eta_i = \alpha_i(\lambda - \lambda_i)/\lambda$ and fitting the spectrum to the absorption features by choosing N sets of λ_i , α_i , and \tilde{C}_i . When this is accomplished, \tilde{m}_r is known to within a single constant \tilde{C}_0 , i.e., the complete *spectral shape* of \tilde{m}_r is known.²⁷

²⁶Although not particularly useful, one could compute the contribution to \tilde{m}_i from each pigment using,

$$(\tilde{m}_i)_j = \lambda \frac{(c_i)_j (a_i^*)_j}{4\pi m_{\text{Water}}},$$

and then $\tilde{m}_i = \sum_j (\tilde{m}_i)_j$. The contribution to \tilde{m}_i at 440 nm from Chlorophyll *a* is $6.6 \times 10^{-6} \lesssim (\tilde{m}_i)_{\text{Chl}} \lesssim 0.01$.

²⁷It is not really necessary to decompose the spectrum of \tilde{m}_i into individual oscillators. It can be shown that

$$\tilde{m}_r(\lambda_0) = 1 + \frac{2\lambda_0^2}{\pi} \oint_0^\infty \frac{\tilde{m}_i(\lambda)}{\lambda(\lambda_0^2 - \lambda^2)} d\lambda,$$

All that remains to find \tilde{m} is the determination of \tilde{C}_0 . We shall see that in some cases this can be accomplished by measuring the spectrum of the extinction or scattering coefficient.

5.4.1.1.2 The “Package Effect”

Consider a cell with a single absorbing pigment. Figure 5.25 shows that the particle absorption coefficient specific to that pigment a_{Part}^* is a non-linear and decreasing function of $\rho' = c_i D a_i^*$. Thus, for a given c_i , a_{Part}^* decreases (increases) as the particle size increases (decreases), or for a given particle size, a_{Part}^* decreases (increases) as the intracellular concentration c_i increases (decreases). Thus, a_{Part}^* depends how the absorbing pigment is “packaged.” In addition, consider intracellular absorption at two different wavelengths with two different a_i^* ’s. For example, let $a_i^*(\lambda_2) = 2a_i^*(\lambda_1)$. Then $\rho'(\lambda_2) = 2\rho'(\lambda_1)$. If the cells are tiny, i.e., $\rho' \ll 1$, then

$$\frac{a_{\text{Part}}^*(\lambda_2)}{a_{\text{Part}}^*(\lambda_1)} = \frac{a_i^*(\lambda_2)}{a_i^*(\lambda_1)} = 2.$$

However, if $\rho'(\lambda_1) = 1$, then $\rho'(\lambda_2) = 2$, and calculating $Q_a(\rho')/\rho'$, we find

$$\frac{a_{\text{Part}}^*(\lambda_2)}{a_{\text{Part}}^*(\lambda_1)} \approx 0.73 \frac{a_i^*(\lambda_2)}{a_i^*(\lambda_1)} \approx 1.5,$$

while if $\rho'(\lambda_1) = 20$, the corresponding ratio is approximately 1. Thus, as the particles increase in size from tiny to very large, the ratio $a_{\text{Part}}^*(\lambda_2)/a_{\text{Part}}^*(\lambda_1)$ varies from 2 (the actual pigment ratio) to 1, and the absorption difference in the spectrum of a_{Part}^* that is seen in the tiny particles is missing in the large particles. This “flattening” of the spectrum of a_{Part}^* with increasing particle size (or increasing intracellular concentration) is also referred to as the “package effect.”

5.4.1.1.3 Detrital Absorption

The package effect can also, partially at least, explain the detrital particle (non-algal particle) absorption spectrum: $a_{\text{NAP}}(\lambda)$. Let’s assume that $a_i(\lambda)$ for detrital particles has the same spectrum as CDOM, i.e., the same spectral shape as $a_y(\lambda)$ (Eq. (5.11)). If we are

where the λ ’s are the wavelengths *in* the water, and \wp indicates that the *principle value* of the integral be used at the singularity ($\lambda = \lambda_0$). This is called the Kronig-Kramers dispersion relationship. Note that it unfortunately requires knowing \tilde{m}_i at all wavelengths 0 to ∞ . Usually \tilde{m}_i is known only over a small portion of the spectrum, e.g., in studying plankton cultures, the visible spectrum. In such a case, the difference formula

$$\tilde{m}_r(\lambda_0) = \tilde{m}_r(\lambda_1) + \frac{2(\lambda_1^2 - \lambda_0^2)}{\pi} \wp \int_0^\infty \frac{\lambda \tilde{m}_i(\lambda)}{(\lambda_0^2 - \lambda^2)(\lambda_1^2 - \lambda^2)} d\lambda,$$

obtained from the above equation by simple subtraction, has been shown to be more insensitive to the assumptions that must be made about the spectrum outside the region of measurements in order to carry out the integration; however, it leaves undetermined the constant $\tilde{m}_r(\lambda_1)$, the analog to C_0 .

given S_y , then, noting that $Da_i = \rho'$ and inserting this spectral shape for ρ' into Eq. (5.21), the result is the spectral shape of a_{NAP} as a function of $Da_i(440)$. Taking $Da_i(440) = 1$ and $S_y = 0.018 \text{ nm}^{-1}$ yields an approximately exponential spectral shape for a_{NAP} , with $S_{NAP} \approx 0.015 \text{ nm}^{-1}$ ($S_y = 0.014 \text{ nm}^{-1}$ yields $S_{NAP} \approx 0.011 \text{ nm}^{-1}$). For smaller $Da_i(440)$, S_{NAP} will be closer to S_y , but, if $a_i(\lambda) \propto a_y(\lambda)$ for non-algal particles, the package effect will yield an a_{NAP} spectrum that is approximately exponential, with $S_{NAP} < S_y$.

5.4.2 The Extinction (Attenuation) and Scattering Coefficient

Consider a homogeneous spherical particle of diameter D with index $\tilde{m} = \tilde{m}_r + \tilde{m}_i i$. Let $\rho = 2x(\tilde{m}_r - 1)$, where $x = \pi D/\lambda_{\text{Water}}$, and $\tan \gamma = \tilde{m}_i/(\tilde{m}_r - 1)$.²⁸ The extinction efficiency ($Q_c = 4\sigma_c/\pi D^2$, where σ_c is the extinction cross section) in the van de Hulst approximation is

$$Q_c = 2 \left[1 - 2 \frac{\cos \gamma}{\rho} \sin(\rho - \gamma) \exp(-\rho \tan \gamma) - 2 \frac{\cos^2 \gamma}{\rho^2} \cos(\rho - 2\gamma) \exp(-\rho \tan \gamma) + 2 \frac{\cos^2 \gamma}{\rho^2} \cos 2\gamma \right]. \quad (5.26)$$

Figure 5.28 compares the result of Eq. (5.26) with exact computations provided by Mie theory (Chapter 3) for particles with no absorption. The approximation is clearly more accurate as $\tilde{m}_r \rightarrow 1$; however, even for indices as high as 1.2, the error rarely exceeds 10%. Clearly, the approximation describes the general features of the extinction and can be used to investigate the general behavior of extinction with the size and refractive index. Figure 5.29 provides the exact extinction efficiency as a function of ρ for absorbing spheres with a low value of \tilde{m}_r (similar to that of phytoplankton). Note that the oscillations in Q_c are increasingly damped as \tilde{m}_i increases. Also plotted on Figure 5.29 are the approximate values of Q_c and Q_a derived from Eqs. (5.26) and (5.21), respectively. These give an indication of the error to be expected using the approximate formulas (usually less than 10% in the efficiencies). Note that when the particle is absorbing, at large values of ρ , $Q_c \rightarrow 2$ and $Q_a \rightarrow 1$, so the scattering efficiency $Q_b = Q_c - Q_a \rightarrow 1$, and the extinction is equally produced by absorption and scattering. The scattering efficiency Q_b determined from $Q_c - Q_a$ in Figure 5.28 is provided in Figure 5.30. The most efficient scatterers have $\rho \approx 4$ or $D \approx 2\lambda_{\text{Water}}/\pi(\tilde{m}_r - 1)$. At 500 nm (in vacuum) this yields $D \approx 4.8 \text{ } \mu\text{m}$ for $\tilde{m}_r = 1.05$.

It is important to note that Figure 5.29 shows that the variation in Q_c from non-absorbing to strongly-absorbing particles is very weak: basically just a damping of the oscillations. In contrast, Figure 5.30 shows that Q_b is considerably reduced by absorption.

²⁸Note that ρ and ρ' are related through $\rho'/\rho = \tilde{m}_i/(\tilde{m}_r - 1) = \tan \gamma$.

This suggests that for particles with a strong absorption feature (e.g., phytoplankton near 670 nm – Figure 5.10) the spectral scattering coefficient would be depressed over the feature, and so the feature may be nearly missing entirely in the beam attenuation coefficient. This is precisely what is observed. Furthermore, Figure 5.29 suggest that \tilde{m}_r and the particle size determine the beam attenuation coefficient, with \tilde{m}_i having only a secondary affect.

If the extinction coefficient for a spherical particle at a particular wavelength and particle diameter are measured, then Q_c can be determined. Assuming that the absorption coefficient at the same wavelength has also been measured, then \tilde{m}_i is determined as described earlier. In this case Eq. (5.26) can be used to determine \tilde{m}_r , and so, the complex refractive index of the particle could be determined from measurements of c and a . Unfortunately, examination of Figures 5.28 and 5.29 show that for a given $Q_c > 1.75$ there are multiple values of ρ , so in this case the determination of \tilde{m}_r is not unique. For $Q_c > 1.75$, spectral information concerning σ_c is required to resolve this ambiguity. If the spectral absorption cross section $\sigma_a(\lambda)$ is available, then the spectrum of $\tilde{m}_i(\lambda)$ can be determined, and the spectrum of $\tilde{m}_r(\lambda)$ found from Eq. (5.25) to within an additive constant \tilde{C}_0 . The constant \tilde{C}_0 can then be estimated by fitting the $Q_c(\lambda)$ spectrum to that computed from

$$\rho(\lambda) = \frac{\pi D}{\lambda_{\text{Water}}} \left[\tilde{C}_0 + \sum_{i=1}^N \frac{\tilde{C}_i \eta_i}{1 + \eta_i^2} \right],$$

using Eq. (5.26). This is how it is known that for phytoplankton $\tilde{m}_r \sim 1.05$ and $\tilde{m}_i \lesssim 0.01$.²⁹ Note that if the particle is too large, or if the absorption is too large, then Q_c and Q_a are very insensitive to ρ' and ρ , respectively, and these methods for estimating \tilde{m} fail.

Given an absorbing pigment within the cell, e.g., Chlorophyll a , and following arguments similar to those leading to Eq. (5.23), we can derive the specific (to the given pigment) scattering coefficient of particles through³⁰

$$b_{\text{Part}}^* = \frac{b_{\text{Part}}}{C} = \frac{3}{2} \frac{Q_b}{c_i D}.$$

Replacing D by its dependence on ρ , we can write this as

$$b_{\text{Part}}^* = \left(\frac{3\pi(\tilde{m}_r - 1)}{c_i \lambda_{\text{Water}}} \right) \frac{Q_b}{\rho}, \quad (5.27)$$

²⁹If higher accuracy is required (which is usually not the case for phytoplankton) then Eqs. (5.26) and (5.21) can be used to approximate \tilde{m} and Mie theory can then be used to provide better determinations.

³⁰The relationships between a_{Part}^* , b_{Part}^* , Q_a and Q_b can also be put in the useful form

$$Q_a = \frac{2}{3} c_i D a_{\text{Part}}^* \quad \text{and} \quad Q_b = \frac{2}{3} c_i D b_{\text{Part}}^*.$$

and see that the variation of the specific absorption coefficient with particle size is governed by Q_b/ρ . Figure 5.31 provides Q_b/ρ as a function of ρ for three values of \tilde{m}_i with $\tilde{m}_r = 1.05$. It shows that, for $\rho \gtrsim 4$, the specific scattering coefficient decreases as the particle size increases for a given intracellular concentration c_i , while for $\rho \lesssim 4$, it increases as the particle size increases. Measurements of b_{part}^* for cultured suspensions of phytoplankton confirm these observations.

5.4.3 The Backscattering Coefficient

Because of the direct significance of backscattering to remote sensing, we now discuss general aspects of the backscattering of spherical particles in relation to particle size and refractive index. Later, we briefly consider the effect of particle shape by examining cylindrically-shaped particles.

The backscattering properties of non-absorbing homogeneous spherical particles are provided in Figure 5.32. They were computed using the Mie scattering theory described in Chapter 3. The abscissa is the size parameter $x = \pi D/\lambda_{\text{Water}}$. The ordinates describe the scattering in several different ways: the backscattering probability B ; the scattering efficiency Q_b ; the backscattering efficiency $Q_{bb} = BQ_b$; and the non-dimensional backscattering scattering cross section $\sigma_{bb} = x^2 Q_{bb}$.³¹ To put the size scale in perspective, note that for a particle that is one micrometer in diameter, at $\lambda_{\text{vacuum}} = 500$ nm, $x \approx 8.4$, and thus phytoplankton would fall in the range $x \approx 8$ to 80 or somewhat larger (or smaller) depending on λ .

One sees that the backscattering probability B is 0.5 for very small particles (Rayleigh scattering), and then decreases as the particle size increases.³² Clearly low-index particles backscatter much less than high-index particles. Note that the minimum observed in B ($x \approx 10$ for $m = 1.15$ and $x \approx 100$ for $m = 1.025$) corresponds approximately to the first maximum in the scattering efficiency Q_b . The backscattering efficiency Q_{bb} reaches a maximum and then rapidly falls to about half that maximum as the particle size increases.³³

³¹The actual backscattering cross section is $Q_{bb} \times \pi R^2$ where R is the particle radius. Since $x = 2\pi R/\lambda_{\text{Water}}$, the actual σ_{bb} is $x^2 Q_{bb} \times (\lambda_{\text{Water}}^2/2\pi)$. So the actual backscattering cross section is that given in the lower right panel of Figure 5.32 (and Figures 5.33 and 5.34) multiplied by $\lambda_{\text{Water}}^2/2\pi$. If λ_{Water} is in μm , σ_{bb} is in μm^2 .

³²For the very low index particles ($m = 1.025$) the Rayleigh-Gans theory (Chapter 3) agrees well with these computations of B up to $x \lesssim 10$, while for the high index ($m = 1.15$) the Rayleigh-Gans computation agrees with Mie only up to $x \lesssim 2$.

³³The strong oscillations seen in the Q_{bb} and B are real but not well represented in these graphs. Correctly characterizing the oscillations requires high resolution in x . Here the resolution in x , i.e., Δx , is 0.1 for $0 \leq x \leq 500$ and 1 for $x \geq 500$. The strange behavior of Q_{bb} for $200 \leq x \leq 500$ is due to “beats” between the frequency of the oscillation and the frequency of the calculation.

The position of the maximum depends strongly on the refractive index with the higher indices moving the maximum to smaller particle sizes. For very large particles, B , Q_{bb} and σ_{bb} differ by approximately an order of magnitude from the low to the high refractive index. It is important to note that regardless of their refractive indices, larger particles always backscatter more strongly than smaller particles with the same index, i.e., they have a larger σ_{bb} .

The effect of particle absorption on scattering is provided in Figure 5.33. The most obvious difference between absorbing and non-absorbing particles is the absence of the high frequency oscillations in the former. Also, B and Q_{bb} become essentially constant for large x , and absorption strongly attenuates the backscattering for larger particles. In addition increasing absorption moves the maximum of Q_{bb} to smaller particle sizes. Note that for this case where the real part of \tilde{m} is 1.075 (perhaps near the upper limit for phytoplankton) the maximum (with absorption) is near $x = 100$, close to that of a particle with $D = 12 \mu\text{m}$, e.g., a large phytoplankter. For lower values of \tilde{m}_r this maximum will shift to still larger particle sizes. Figure 5.33 suggests that the range of B for phytoplankton is likely to be about 0.001 to 0.003, as was assumed for B_{pl} in the empirical models in Section 5.3.3.4.

Finally, it is interesting to understand the effect of internal structure on scattering by particles. Mie-like calculations are possible for coated spherical particles. One might imagine phytoplankton having an interior that is mostly water-like (with absorbing pigments) surrounded by a membrane that keeps the cell intact. We provide here computations for a coated sphere in which the core of the sphere is absorbing ($m = 1.075 + 0.0050i$) and the coating is a higher-index non-absorbing material ($m = 1.15$).³⁴ Figure 5.34 provides results similar to those in Figures 5.32 and 5.33. In Figure 5.34 $x = 2\pi R/\lambda_{\text{Water}}$, where here R is the overall radius of the sphere – core *plus* coating). The parameter t is the thickness of the coating in micrometers ($t = 0, 0.04, 0.08, 0.16, 0.24, 0.32, 0.48, 0.64 \mu\text{m}$). The overall radius of the particle is $R = R_c + t$, where R_c is the radius of the core, so $x = x_c + 2\pi t/\lambda$. For $\lambda_{\text{vacuum}} = 500 \text{ nm}$ and $t = 0.08 \mu\text{m}$, $x = x_c + 1.33$. Thus, for a given value of x at a different wavelength, the physical thickness of the coating will be different, so it is best to think of these results as being only for a single wavelength, i.e., $\lambda_{\text{vacuum}} = 500 \text{ nm}$.

The most obvious effect of the coating is to significantly increase the backscattering and move the maximum of Q_{bb} to smaller particle sizes. It is interesting to note that the increase in Q_{bb} is so large in some cases that σ_{bb} is no longer a monotonically increasing function of x . Also, for large x the particles with various coatings tend to constant B and Q_{bb} such that $t = 0.08, 0.16, 0.32$, and $0.64 \mu\text{m}$ approach asymptotes of similar value (“similar” meaning within approximately 20%), $t = 0.04$ and $0.48 \mu\text{m}$ similar values, and

³⁴The refractive index of the coating used in the computations is much higher than one might expect; however, the results are presented here for illustrating the possible effects of a cell membrane, not as a realistic model of phytoplankton.

$t = 0$ and $0.24 \mu\text{m}$ similar values. We have no quantitative explanation of this behavior; however, qualitatively the effect must be due to interference among multiple reflections within the coating. Figure 5.35 compares the backscattering efficiency Q_{bb} of the coated sphere with successively thicker coatings to that of a homogeneous sphere with the same index as the coating. Interestingly, for $10 \lesssim x \lesssim 60$, the backscattering for a sphere with a $0.64 \mu\text{m}$ coating is very close to the backscattering of the homogeneous sphere with the coating index.³⁵ As the sphere becomes larger ($x > 60$), Q_{bb} falls by more than one order of magnitude.

In sum, the message here is clear: a high-index coating on a low-index particle can significantly increase the backscattering, even if the coating thickness is significantly less than λ .

5.4.4 The Scattering Phase Function

We will not discuss in detail the dependence of the scattering phase function on size and refractive index as these are discussed in some detail in Chapters 3 and 4. In addition, the phase functions are considerably changed when a distribution of particle sizes is considered (later in this chapter) rather than a single size. However, we do provide one example just to show the general nature of the variation of the phase function with scattering angle and with the strength of absorption (\tilde{m}_i). Figure 5.36 shows the phase function computed for a particle with $x = 20$ and $\tilde{m}_r = 1.05$. The three curves shown are for $\tilde{m}_i = 0, 0.0075$, and 0.0150 , i.e., no absorption to strong absorption. There are two effects that are evident from the figure: (1) the large variation in the magnitude of the phase function over the range of scattering angles (8 orders of magnitude); and (2) the strong oscillations with an angular frequency of approximately 1 oscillation per 15° . Both of these are strongly influenced by the particle size. Larger particles show a larger range of variation and more oscillations in a given angular range (higher angular frequency of oscillations). This figure also shows that, at least for $x = 20$, the influence of absorption is small. Absorption damps out the oscillations, lowering the peaks and increasing the troughs. As we have seen earlier, absorption causes the backscattering to decrease. When a wide distribution of particle sizes is considered, the oscillations are completely smoothed out.

³⁵At the maximum in Figure 5.35 $x \approx 40$ so, at a vacuum wavelength of 500 nm , the overall radius of the particle is about $2.75 \mu\text{m}$. The thickest coating ($0.64 \mu\text{m}$) is then 23% of the particles radius at the maximum of Q_{bb} .

5.4.5 Influence of Particle Shape on IOPs

Phytoplankton (and most marine particles) have shapes that deviate significantly from spheres. To try to demonstrate the influence of shape on the IOPs, we take the example of a collection of homogeneous cylinders in random orientation. (Some species of phytoplankton collect in long chains that are actually reasonable to model as homogeneous cylinders.) The aspect ratio of the cylinder is defined by $AR = L/D$, where L is the length and D is the diameter. The IOPs for finite cylinders can be computed using the DDA described in some detail in Chapter 3. In addition, a computation method similar to Mie theory for spheres is available for infinitely long cylinders. We will compute the IOPs as a function of D and AR for cylinders with $1/3 \leq AR \leq 30$ and compare these to those of spheres and infinite cylinders.

5.4.5.1 Absorption

We begin by looking at the absorption coefficient of cylinders. For an infinite cylinder, the only dimensions in the electromagnetic scattering problem are the diameter of the cylinder D and the vacuum wavelength λ . Taking a clue from spheres, we take the size parameter for cylinders to be $x = m_{\text{Water}}(\pi D/\lambda)$, i.e., the IOPs of infinite cylinders can depend on size only through x . For finite cylinders, the IOPs will depend on both x and AR .

Figure 5.37 (left panel) provides the absorption efficiency Q_a for finite absorbing-cylinders as a function of $\rho' = 4x\tilde{m}_i$. The value of Q_a is determined by computing the absorption cross section σ_a for the particle and dividing it by the orientationally-averaged projected area³⁶ of the cylinder: $\pi D(L + D/2)/4$. As with spheres, ρ' is the absorption optical thickness of the cylinder for a ray passing through the axis, normal to the surface, i.e., Da_i . The computations (symbols) are for refractive indices $\tilde{m} = 1.05 + 0.002i$ and $1.05 + 0.010i$, i.e., \tilde{m}_r in the midrange for phytoplankton and \tilde{m}_i values representing (relatively) weak and strong absorption for phytoplankton. The solid curves are the computed values of Q_a for an infinite cylinder, while the dotted curve is Q_a for a homogeneous sphere with index $\tilde{m} = 1.05 + 0.002i$ and the given value of x . There are several points for each ρ' , and these correspond to different values of AR , with the largest AR values closest to the solid lines (infinite cylinders). Points with $AR < 1$ are all below the dotted line (homogeneous spheres). These are disk-shaped particles. The points very close to the dotted line all have $AR = 1$. These are the most compact cylinders (minimum surface area for a given volume)

³⁶The projected area of an object in a given orientation is just the area of the geometric shadow of the object when illuminated by a parallel beam (neglecting diffraction). The orientationally-averaged projected area is the average area of the shadows over orientation. For an object that is convex (no indentations) the orientationally-averaged projected area is the object's actual surface area divided by 4.

and, in this respect, the closest to spheres. All points above the dotted line have $AR > 1$, with AR increasing as Q_a becomes larger for a given ρ' .

The right panel of Figure 5.37 is the same as the left panel except that all points with $AR < 3$ have been removed (and Q_a for spheres as well). The removed values include $AR = 1/3, 1/2, 1$, and 2 (note that not all values of AR are computed for each D). It shows that Q_a can be computed for a cylinder with $AR > 3$ with an error less than about 10%, by assuming that the cylinder is infinite. Conversely, given Q_a (i.e., through measurement of σ_a and the particle's L and D), when $AR \geq 3$, \tilde{m}_i can be deduced from measurements with an error of roughly the same order (nominally 10%, but dependent on the actual value of Q_a) by assuming that the cylinder is infinite.

Can we deduce \tilde{m}_i if D and L are unknown, but the (cylindrically shaped) particle volume is known, e.g., from Coulter counter measurements? In this case, all one can do is assume a particle shape – traditionally spherical. From the volume measurement, one can compute the diameter of the equal-volume sphere D_{Equiv} . Then from the measurement of the absorption cross section σ_a (the absorption coefficient divided by the number density in the beam through the sample volume) one can form the volume equivalent absorption efficiency $Q_a = 4\sigma_a/\pi D_{\text{Equiv}}^2$. The volume-equivalent size parameter is $x = \pi D_{\text{Equiv}}/\lambda_{\text{Water}}$ and $\rho' = 4x\tilde{m}_i$. Thus, all computed quantities from σ_a and the particle volume assume a spherical shape. Figure 5.38 shows the result of this exercise. The solid lines on the figure are Q_a for $\tilde{m} = 1.05 + 0.002i$ (black) and $\tilde{m} = 1.05 + 0.010i$ (red) computed from Mie theory (spheres), and the dotted line is Eq. (5.21) (spheres). Again we see that \tilde{m}_i could be estimated with an error $\lesssim 10\%$ by analyzing the experimental data assuming the particles are homogeneous, volume-equivalent spheres.³⁷

5.4.5.2 Extinction and Scattering

The extinction efficiency Q_c of non-absorbing finite cylinders ($\tilde{m} = 1.2$) is presented in Figure 5.39. Again, $\rho = 2\pi D(\tilde{m}_r - 1)/\lambda_{\text{Water}}$, where D is the diameter of the cylinder. The left panel shows the results for all aspect ratios, while the right panel is restricted to $AR \geq 3$. The solid curve is Q_c for an infinite cylinder. As in the case of absorption, when $AR \geq 3$, treating the particle as being infinitely long results in only a small error in Q_c . Figure 5.40 compares Q_c for a non-absorbing cylinder with that for a sphere with the same \tilde{m} and shows that the extinction efficiency for each varies in a similar manner, but with a phase shift in the oscillations between the two. The case of absorbing cylinders

³⁷Noting that there is error associated with Eq. (5.21), one could employ Mie theory instead for higher accuracy. In the cases examined here, the dependence of Q_a on \tilde{m}_r (unknown) is quite weak, so higher accuracy is possible.

is presented in Figure 5.41 with the left panel including all AR and the right panel only $AR \geq 3$. All of the cases in Figure 5.37 are included on Figure 5.41. Clearly, when $AR \geq 3$ the particle's extinction efficiency can be computed with reasonable accuracy by using the infinite- L assumption.³⁸

Finally, what if we interpret extinction data for cylinders by assuming they are spheres with the same volume (the *volume-equivalent sphere* approximation)? Figure 5.42 provides the result of such an exercise. Here, the diameter D_{Equiv} of the sphere having the same volume as the cylinder is computed, σ_c is divided by $\pi D_{\text{Equiv}}^2/4$ to form Q_c , and $\rho = 2x_{\text{Equiv}}(\tilde{m}_r - 1)$, where $x_{\text{Equiv}} = \pi D_{\text{Equiv}}/\lambda_{\text{Water}}$. Clearly, treating absorption-scattering process using the *volume-equivalent sphere* approximation does not work as well for extinction as it does for absorption (Figure 5.38). One method for estimating \tilde{m}_r from σ_c and particle volume data is to use D_{Equiv} to form Q_c . Then from Q_c , try to find a value of \tilde{m}_r that is consistent between the two. For example, consider the large open diamond point on Figure 5.42 ($\rho \approx 2.36$, $Q_c \approx 1.40$). Using this value of Q_c one would assume from the extinction efficiency curves for spheres that ρ should be approximately 1.8. This means that the ρ estimated from Q_c is too small by a factor of about 3/4, hence the estimated value of $\tilde{m}_r - 1$ will be too small by the same factor, and the derived \tilde{m}_r will be approximately 1.037 instead of the correct 1.050. Thus, all the points to the right of the curves in Figure 5.42 will yield \tilde{m}_r values that are too small. Note that not knowing the precise value of \tilde{m}_i leads to a much smaller error in \tilde{m}_r than we are discussing here. In fact we could assume $\tilde{m}_i = 0$ with little additional error in \tilde{m}_r . For large values of \tilde{m}_r or for larger cylinders, this method of finding the index for cylinders fails completely. It is easy to understand why. Since $Q_c \rightarrow 2$ for large particles, we have $\sigma_c \rightarrow 2\pi D(L + D/2)$ for large L . Now, let Q_c^{Sph} be the extinction efficiency computed by assuming that the cylinder is a sphere of equal volume, i.e., as sphere with diameter D_{Equiv} , where $\pi D_{\text{Equiv}}^3/6 = \pi D^2 L/4$. Then, for large L

$$Q_c^{\text{Sph}} \triangleq \frac{4\sigma_c}{\pi D_{\text{Equiv}}^2} \rightarrow 2\pi D(L + D/2) \left(\pi \left(\frac{3}{2} \right)^{2/3} D^{4/3} L^{2/3} \right)^{-1} \approx 1.5(AR)^{1/3},$$

which shows that Q_c^{Sph} has no upper bound, while Q_c for spheres does not exceed 4 (Figure 5.28). Thus, care must be taken when trying to use the equivalent sphere approximation to derive refractive indices for cylinders from extinction measurements. Note that similar considerations apply to deriving \tilde{m}_i from Q_a^{Sph} . These methods can only be used when

³⁸If we assume that the cylinder contains an absorbing pigment of concentration c_i within the cylinder, the equations such as Eqs. (5.23) and (5.27) can be derived for cylinders:

$$\frac{a_{\text{Part}}^*}{a_i^*} = \left(1 + \frac{1}{2AR} \right) \frac{Q_a}{\rho'} \quad \text{and} \quad b_{\text{Part}}^* = \frac{2\pi(\tilde{m}_r - 1)}{c_i \lambda_{\text{Water}}} \left(1 + \frac{1}{2AR} \right) \frac{Q_b}{\rho},$$

and Q_a and Q_b can be computed assuming the cylinder is infinite if $AR \geq 3$.

$Q_c^{\text{Sph}} < 4$ and $Q_a^{\text{Sph}} < 1$, and then, only when ρ is sufficiently small to avoid the multivalued nature of ρ given Q_c .

5.4.5.3 Backscattering

How much does shape influence particle backscattering? To look at this question we again consider homogeneous cylinders. As in the case of Q_c and Q_a , the backscattering probability becomes independent of the length of the cylinder (and equal to that of an infinite cylinder) as the aspect ratio becomes large. Figure 5.43 shows B as a function of AR for low index absorbing and non-absorbing cylinders to demonstrate the influence of (or lack of influence of) absorption on backscattering probability and Figure 5.44 shows the influence of \tilde{m}_r for non-absorbing cylinders. Clearly AR need only be larger than 3 to 5 in order that B is effectively independent of AR . Figure 5.45 provides examples of B and Q_{bb} as a function of the size parameter. Included for comparison are the same quantities for homogeneous spheres.³⁹ In these figures, the points for cylinders at a given x cluster together when the AR is large (see Figure 5.43). Thus, in the case of B , the points falling farthest below the curves for spheres correspond to cylinders with the largest aspect ratios, and for these B for cylinders is about half that of spheres. For the low index cylinders computations were also carried out for aspect ratios smaller than one (disk-shaped) and equal to one (maximum volume cylinder for a given surface area, i.e., most compact). The open symbols are all for $AR < 1$, while all the $AR = 1$ cases are very close to the curve for spheres, as one might expect. For low index cylinders, Q_{bb} values are roughly the same for spheres and cylinders; however, that does not mean the backscattering cross sections σ_{bb} are similar, because they are related to Q_{bb} through $\sigma_{bb} = Q_{bb} \times A_p$, where A_p is the orientationally-averaged projected area of the cylinder, $\pi D(L + D/2)/4$, which is directly proportional to L for large AR . However, there is a simplification that is particularly useful in the low-index case, where $Q_{bb}^{\text{Cylinder}} \approx Q_{bb}^{\text{Sphere}}$. The backscattering cross section can be written

$$\sigma_{bb}^{\text{Cylinder}} = \frac{\pi D}{4}(L + D/2) Q_{bb}^{\text{Cylinder}} = \frac{\pi D^2}{4}(AR + 1/2) Q_{bb}^{\text{Cylinder}},$$

while

$$\sigma_{bb}^{\text{Sphere}} = \frac{\pi D^2}{4} Q_{bb}^{\text{Sphere}}.$$

Therefore, for the low-index case

$$\sigma_{bb}^{\text{Cylinder}} \approx (AR + 1/2) \sigma_{bb}^{\text{Sphere}},$$

³⁹It is important to emphasize again that the size parameter here for the cylinder has nothing to do with it's length, it depends only the diameter, i.e., it is *not* the size parameter for the volume-equivalent or surface area-equivalent sphere, as is often used in comparison of scattering by particles of different shapes.

and the backscattering cross section of a cylinder of aspect ratio AR is approximately equal to the backscattering cross section of a sphere with the same diameter times AR . For the high index case, Q_{bb} increases more slowly with x than for spheres; however, a similar approximation for the high-index cylinders suggests that this result still holds for them to well within a factor of 2. One should also note that for the largest x here, the backscattering efficiency of the high-index cylinder is roughly a factor of 50 times that of the low-index cylinder, indicating the very strong dependence of backscattering on the refractive index.

5.4.5.4 The Scattering Phase Function

In this subsection we compare the scattering phase function for spheres and cylinders and the dependence of the phase function for cylinders on their aspect ratio. Figure 5.46 provides the phase function for a non-absorbing cylinder with $x = 2.6$ and $\tilde{m}_r = 1.20$ as a function of the aspect ratio.⁴⁰ Note that for $AR \geq 5$ there is very little difference between the various phase functions except at very small angles, where the longer cylinders have increasingly larger values.⁴¹ It is remarkable how quickly with increasing AR the phase function (except at small angle) becomes essentially static. The phase function for the sphere is somewhat similar to that for the cylinder with $AR = 1$. We have already seen this similarity in the variation of B with AR (Figures 5.43 and 5.44). Figures 5.47 and 5.48 compare the phase function and degree of linear polarization, respectively, for an absorbing cylinder with $x = 7$, $\tilde{m}_r = 1.05$ and $\tilde{m}_i = 0.002$, with that of a similar diameter sphere. Note that the similarity between the two aspect ratios and the deeper minima for the larger AR .⁴² The phase function for the sphere is similar in magnitude to the cylinder, but differs in detail.

The motivation behind this section was to show some of the differences and the similarities between IOPs computed for spheres (the usual assumption) and cylinders (used as an example of non-spherical particles). The cylindrical shape was used only because of its simplicity. Ellipsoidal shapes (spheroids) could have been used as well. The main conclusion is that, although there are significant quantitative differences between particles of different shapes, there are qualitative similarities. Also, scattering and absorption for

⁴⁰In Figure 5.46, $Q_c = 0.487, 0.410, 0.999, 1.007, 1.032, 1.033$, and 1.033 , and $B = 0.0326, 0.0349, 0.0212, 0.0212, 0.0209, 0.0218$, and 0.0220 , for the sphere, $AR = 1, 5, 10, 100, 200$, and 250 , respectively, i.e., for the sphere $Q_c = 0.487$ and $B = 0.0326$, etc.

⁴¹For $AR > 10$ the apparent noise in the computations is due to using an insufficient number of orientations in the orientational averaging. The computation time is directly proportional to the number of orientations and excessive time would be required to produce smooth curves at high AR .

⁴²In Figure 5.47, $Q_c = 0.271, 0.464$, and 0.537 , and $B = 0.00300, 0.00176$, and 0.00153 , for the sphere, $AR = 3$ and 20 , respectively.

cylinders has been shown to be similar in magnitude to those of some associated sphere, i.e., an equal volume sphere in the case of absorption and $AR + 1/2$ times the backscattering cross section of a sphere with the same diameter in the case of backscattering. In this respect, the optics of spherical particles can serve as a “guide” toward understanding the IOPs of particles of other shapes.

5.5 Influence of The Distribution of Particle Size on IOPs

In the previous subsection we discussed the theory of the IOPs for a particle with a specific size and composition; however, a suspension of single-size particles in nature is very difficult to come by. Even the particles in a suspension of cultured phytoplankton have a distribution of particles sizes (sometimes rather wide compared to the mean size). In natural waters the range encountered for particle sizes can cover many orders of magnitude. Figure 5.49 provides the size distribution of particles in the surface waters of the Tongue of the Ocean, Bahamas, measured with a Coulter counter. In the figure, the abscissa is the number of particles with diameter (D) greater than a given value (ordinate). The measurements range from $0.5 \leq D \leq 8 \mu\text{m}$, but the distribution undoubtedly continues to smaller and larger particles. Clearly, it is critical to understand the influence of the distribution of particle sizes on the IOPs.

5.5.1 Analytical Representations of the Size Distribution

The solid lines in Figure 5.49 piecewise represent functions of the form $N_{>D} = kD^{-\nu}$, where k and ν are constant for each segment. If D is in μm , then k is the number of particles per ml with $D > 1 \mu\text{m}$. For a case like Figure 5.49 with multiple straight line segments, then each must be extrapolated to $D = 1 \mu\text{m}$ to determine the associated k . For example, in Figure 5.49 for 3 segments, the smallest- and the middle-size segments have $k \approx 17,000$ particles per ml and $\nu \approx 6.5$ and 2.5 , respectively, while the largest segment has $k \sim 60,000$ particles per ml and $\nu \approx 3$. If one were to try to fit the data to a single function, then the ν value would be ~ 3 with $k \sim 40,000 \text{ ml}^{-1}$.

The “size distribution” is defined by $n(D) = dN/dD$, where dN is the number of particles per unit volume with diameters in the range $D - dD/2 \leq D \leq D + dD/2$. This is related to $N_{>D}$ through

$$n(D) = \frac{dN}{dD} = -\frac{dN_{>D}}{dD},$$

so, if we assume the distribution can be represented by a single segment, which is roughly

the case for $D > 1 \mu\text{m}$, then⁴³

$$n(D) = \frac{\nu k}{D^{\nu+1}} \triangleq \frac{K}{D^{\nu+1}}.$$

If the $N_{>D}$ graph contains several segments, then this formula applies to each segment with their individual values of k and ν , and the size distribution $n(D)$ will not be continuous, but will have jump discontinuities. The actual size distribution of course has no discontinuities, only its segmented representation. Size distributions of this type are called power-law distributions; also called Junge distributions. (See Chapter 4 for a review of power-law distributions and log-normal distributions, discussed below.)

The total number of particles per ml ($N_0 = \int_0^\infty n(D) dD$) is infinite if the distribution of sizes continues to follow the power-law all the way to $D = 0$, so if it continues to have this form, there must be a smallest particle size D_S . If there is a largest size D_L , then for this truncated distribution the constant K is related to the total number of particles:

$$N_0 = \frac{k}{D_s^\nu} - \frac{k}{D_L^\nu} = \int_0^\infty dN = K \int_{D_s}^{D_L} D^{-(\nu+1)} dD,$$

so

$$K = \nu k = \frac{N_0 \nu}{(D_s^{-\nu} - D_L^{-\nu})}.$$

Other distributions are also used to describe marine particle sizes. For example, the distribution of sizes in phytoplankton cultures of a single species is often represented as a *log-normal* distribution:⁴⁴

$$\frac{dN}{d(\ln D)} = \frac{N_0}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{\ln D - \ln D_N}{\sigma} \right)^2 \right], \quad (5.28)$$

where N_0 (the total number of particles per unit volume), D_N (the *median* or *modal* diameter) and σ are constant. An equivalent form of this distribution is

$$\frac{dN(D)}{dD} = \frac{N_0}{\log_e(10) D \sqrt{2\pi\sigma_{10}^2}} \exp \left[-\frac{1}{2} \left(\frac{\log_{10}(D/D_N)}{\sigma_{10}} \right)^2 \right], \quad (5.29)$$

⁴³In Chapter 4, the size distribution is defined in terms of particle radius R . Here, we use diameter D because that is how Coulter counter results are usually reported. Clearly, $dN/dR = 2dN/dD$.

⁴⁴Note the similarity between Eq. (5.28) and the *Gaussian* or *Normal* distribution. For particles following a Gaussian distribution in diameter,

$$\frac{dN}{dD} = \frac{N_0}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{D - D_m}{\sigma} \right)^2 \right].$$

For this distribution the mean of D , $\langle D \rangle = D_m$ and the variance $\langle (D - \langle D \rangle)^2 \rangle = \sigma^2$. Thus, for the log-normal distribution (Eq. (5.28)) $\langle \ln D \rangle = \ln D_N$ and $\langle (\ln D - \langle \ln D \rangle)^2 \rangle = \sigma^2$ — the log normal distribution is a Gaussian distribution in which $D \rightarrow \ln D$ everywhere.

where $\sigma_{10} = \sigma \log_{10}(e)$. Size distributions of the type shown in Figure 5.49 are sometimes modeled as different populations of particles, with the individual populations having log-normal distributions. For example, a distribution with ν values similar to that shown in Figure 5.49 can be well modeled by a combination of three log-normal distributions:

$$\frac{dN}{dD} = \sum_{i=1}^3 \frac{dN_i}{dD}, \quad \text{where} \quad \frac{dN_i}{dD} = \frac{N_{0i}}{\log_e(10)D\sqrt{2\pi\sigma_{10i}^2}} \exp \left[-\frac{1}{2} \left(\frac{\log_{10}(D/D_{Ni})}{\sigma_{10i}} \right)^2 \right],$$

with $N_{0i} = (6.5 \times 10^{11}, 7 \times 10^4, 3 \times 10^3) \text{ ml}^{-1}$, $D_{Ni} = (0.011, 0.5, 1.0) \mu\text{m}$, and $\sigma_{10i} = (0.35, 0.30, 0.35)$, where $i = (1, 2, 3)$. This is shown in Figure 5.50. The right panel in the figure compares favorably to the cumulative distribution in Figure 5.49. This kind of representation has the advantage that each component of the distribution might be physically linked to a particular kind of particle, e.g., here possibly viruses ($i = 1$), detritus ($i = 2$), and phytoplankton ($i = 3$).

5.5.2 Influence of the Distribution of Sizes on IOPs

Recall that the IOPs (a_{Part} , b_{Part} , c_{Part} , and β_{Part}) for single-sized particles are related to the respective cross sections σ_a , σ_b , σ_c , and $d\sigma_b/d\Omega$, by the number of particles per unit volume, e.g.,

$$b_{\text{Part}} = \frac{N}{V} \sigma_b,$$

etc., where N is the number of particles in V . When we have a distribution of sizes governed by $n(D) = dN(D)/dD$, with dN equal to the number of particles per unit volume with diameters in the range $D - dD/2 \leq D \leq D + dD/2$, then for particles in this size range,

$$db_{\text{Part}}(D) = \sigma_b(D) dN(D) = \sigma_b(D) \frac{dN(D)}{dD} dD = \sigma_b n(D) dD,$$

etc, for a_{Part} , c_{Part} , and β_{Part} , so

$$b_{\text{Part}} = \int \sigma_b n(D) dD,$$

etc. If the efficiencies are given, e.g., by $Q_b(D) = 4\sigma_b/\pi D^2$, then

$$b_{\text{Part}} = \frac{\pi}{4} \int Q_b D^2 n(D) dD,$$

with similar expressions for a and c . In a similar manner, VSF is

$$\beta_{\text{Part}}(\Theta) = \int \frac{d\sigma_b}{d\Omega} n(D) dD.$$

In these expressions, the integrals are taken over all particle sizes.

For some applications it is useful to define mean efficiencies with respect to a given size distribution in the following manner. The mean scattering efficiency is defined according to

$$\langle Q_b \rangle = \frac{\int D^2 Q_b(D) n(D) dD}{\int D^2 n(D) dD}$$

so

$$b_{\text{Part}} = \frac{\pi}{4} \langle Q_b \rangle \int D^2 n(D) dD = \langle Q_b \rangle G,$$

where

$$G = \frac{\pi}{4} \int D^2 n(D) dD$$

is the total projected area of the particles in the size distribution. Similar definitions are made for $\langle Q_a \rangle$ and $\langle Q_c \rangle$.

5.5.2.1 Estimation of a_i and \tilde{m} for Size-Distributed Plankton

In Section 5.4.1.1.1 we examined a method for estimation of particle refractive index, e.g., for single-sized phytoplankton, from measurements of the absorption coefficient. How does the distribution of particle sizes influence this estimation? We note that Q_a is a monotonically increasing function of $\rho' = a_i D$, where a_i is the absorption coefficient of the intracellular material. Therefore,

$$\langle Q_a(a_i) \rangle = \frac{\pi}{4G} \int D^2 Q_a(a_i D) n(D) dD$$

is also monotonic in a_i , so $a_{\text{Part}} = \langle Q_a(a_i) \rangle G$ admits a unique solution for a_i and therefore \tilde{m}_i . If the complete spectrum of $a_{\text{Part}}(\lambda)$ is measured yielding $\tilde{m}_i(\lambda)$, then it can be fit to a collection of oscillators from which \tilde{C}_i in Eq. (5.25) can be determined. This means that $\tilde{m}_r - 1$ can be determined to within an additive constant \tilde{C}_0 . This constant can sometimes (but not always) be determined from c_{Part} :

$$c_{\text{Part}}(\lambda) = \langle Q_c(\tilde{m}, \lambda) \rangle G.$$

If the particle sizes (and \tilde{m}) are actually in the range where $Q_c(x, \tilde{m})$ varies significantly with x over the spectrum, then \tilde{C}_0 can be determined, otherwise it cannot. This requires that the mean size is not so large (or \tilde{C}_0 not so large) that $\langle Q_c \rangle \approx 2$. So, with this latter proviso, the basic method of estimating \tilde{m} , described for single-size particles, will also work for particles distributed in size.

For specific absorption coefficients, we can define the mean intracellular pigment concentration $\langle c_i \rangle$ to be the bulk pigment concentration C divided by the total particle volume V_{Total} , i.e.,

$$\langle c_i \rangle \triangleq \frac{C}{V_{Total}} = \frac{C}{(\pi/6) \int D^3 n(D) dD}.$$

Then it is easy to show that

$$a_{Part}^* = \frac{a_{Part}}{C} = \frac{\langle Q_a \rangle (\pi/4) \int D^2 n(D) dD}{\langle c_i \rangle (\pi/6) \int D^3 n(D) dD} = \frac{3}{2} \frac{\langle Q_a \rangle \langle D^2 \rangle}{\langle c_i \rangle \langle D^3 \rangle},$$

where $\langle D^p \rangle = \int D^p n(D) dD$. This is the analog of $a_{Part}^* = 3Q_a/2c_i D$ for distributed particles. Also, noting that $a_i^* = a_i/\langle c_i \rangle$, we have

$$\frac{a_{Part}^*}{a_i^*} = \frac{3}{2} \frac{\langle Q_a \rangle \langle D^2 \rangle}{a_i \langle D^3 \rangle},$$

which is the analog of Eq. (5.23) for distributed particles. In a similar manner,⁴⁵

$$b_{Part}^* = \frac{3}{2} \frac{\langle Q_b \rangle \langle D^2 \rangle}{\langle c_i \rangle \langle D^3 \rangle}.$$

5.5.2.2 IOPs of Power-Law Distributed Particles

As mentioned earlier, if we modeled the distribution of ocean particles in Figure 5.49 by a single power-law, one would find $\nu \approx 3$. Thus, it is important to examine the IOPs associated with a size distribution

$$n(D) = \frac{K}{D^{\nu+1}},$$

with D extending over the entire size range D_{Min} to D_{Max} . Consider the total scattering coefficient,

$$b_{Part} = \frac{\pi}{4} \int_{D_{Min}}^{D_{Max}} D^2 Q_b n(D) dD = \frac{\pi K}{4} \int_{D_{Min}}^{D_{Max}} \frac{D^2}{D^{\nu+1}} Q_b dD. \quad (5.30)$$

The integral can be computed by deriving Q_b from Mie theory or from the anomalous diffraction approximation, and will obviously depend on D_{Min} and D_{Max} . It is instructive

⁴⁵The relationships between a_{Part}^* , b_{Part}^* , $\langle Q_a \rangle$ and $\langle Q_b \rangle$ can also be put in the useful form

$$\langle Q_a \rangle = \frac{2}{3} \frac{\langle D^3 \rangle}{\langle D^2 \rangle} \langle c_i \rangle a_{Part}^* \quad \text{and} \quad \langle Q_b \rangle = \frac{2}{3} \frac{\langle D^3 \rangle}{\langle D^2 \rangle} \langle c_i \rangle b_{Part}^*.$$

to see under what circumstances one can extend the limits to $D_{\text{Min}} \rightarrow 0$ and $D_{\text{Max}} \rightarrow \infty$. For small D we are in the Rayleigh regime in which the scattering is proportional to the square of the particle volume, i.e. $D^2 Q_b \propto D^6$, while for large D , $Q_b \rightarrow 2$. So, for small D the integrand is $\propto D^{5-\nu}$, while for large D it is $\propto D^{1-\nu}$. Therefore, if the limits are extended, the integral will exist,⁴⁶ i.e., be $< \infty$, if $2 < \nu < 6$. Thus, even though the total number of particles and the total projected area of the particles (and for $\nu > 2$, the total volume of particles) are infinite, the scattering is finite for this range of the parameter ν . Then, for this range of ν we can compute

$$b_{\text{Part}} = \frac{\pi K}{4} \int_0^\infty \frac{D^2 Q_b}{D^{\nu+1}} dD.$$

To do this we note that Q_b depends only on x and \tilde{m} , so we change variables in the integral from D to $x = \pi D / \lambda_{\text{Water}}$. Making this substitution ($D = \lambda_{\text{Water}} x / \pi$) we find

$$b_{\text{Part}} = \frac{\pi K}{4} \left(\frac{\lambda_{\text{Water}}}{\pi} \right)^{2-\nu} \int_0^\infty x^{1-\nu} Q_b dx.$$

The integral in this expression is simply a number, so this tells us that the spectral dependence of the particle scattering coefficient is

$$b_{\text{Part}}(\lambda) \propto \lambda^{2-\nu},$$

a relationship we have already found for aerosols (Chapter 4). Similar arguments can be made for a_{Part} , c_{Part} , and β_{Part} , i.e., they all vary spectrally according to $\lambda^{2-\nu}$, which implies that the scattering phase function for such a distribution of particles is *independent of wavelength*. Note that these conclusions require that the size distribution be extended over the entire size range ($0 \rightarrow \infty$).⁴⁷ If the size range is finite, then

$$b_{\text{Part}} = \frac{\pi K}{4} \left(\frac{\lambda_{\text{Water}}}{\pi} \right)^{2-\nu} \int_{x_{\text{Min}}}^{x_{\text{Max}}} x^{1-\nu} Q_b dx, \quad (5.31)$$

⁴⁶Note that

$$\left| \int_c^\infty \frac{1}{y^a} dy \right| < \infty \quad \text{if } a > 1 \quad \text{and } c > 0 \quad \text{and} \quad \left| \int_0^c \frac{1}{y^b} dy \right| < \infty \quad \text{if } b < 1 \quad \text{and } c < \infty.$$

⁴⁷There is another interesting relationship can be derived when the power-law size distribution is valid for all sizes from 0 to ∞ . This concerns small-angle scattering. Small-angle scattering is dominated by diffraction. In the diffraction-geometrical optics approximation (Chapter 3) the small angle volume scattering function is given by

$$d\beta(\Theta) = \frac{\kappa^2 R^4 J_1^2(\kappa R \sin \theta)}{(\kappa R \sin \theta)^2},$$

where $\theta = 2 \sin(\Theta/2)$, Θ is the scattering angle, and R is the particle radius. Now, $\kappa R = x$ so $R = \lambda x / 2\pi$, and for small angle scattering $\theta \approx \Theta$. These yield

$$d\beta(\Theta) = \left(\frac{\lambda}{2\pi} \right)^2 x^2 \frac{J_1^2(x\Theta)}{\Theta^2} dN$$

where $x_{\text{Max}} = \pi D_{\text{Max}}/\lambda_{\text{Water}}$ and $x_{\text{Min}} = \pi D_{\text{Min}}/\lambda_{\text{Water}}$. In this case, the integral itself also depends on wavelength (through the dependence of the limits of integration on wavelength) so $b_{\text{Part}}(\lambda)$ is no longer $\propto \lambda^{2-\nu}$, nor are a_{Part} , c_{Part} , and β_{Part} . In this case, the phase function becomes dependent on wavelength.

It is of obvious interest to know how the various scattering quantities depend on the limits to the size distribution. We examine this by computing the scattering coefficient, the backscattering coefficient and the VSF at various angles as a function of the integration limits. Specifically, for a given quantity, e.g., the scattering coefficient, consider

$$f_b(x) = \int_0^x y^{1-\nu} Q_b(y, \tilde{m}) dy, / \int_0^\infty y^{1-\nu} Q_b(y, \tilde{m}) dy,$$

as $x = \pi D/\lambda_{\text{Water}}$ varies from 0 to ∞ . Equation (5.31) shows that

$$b_{\text{Part}} = [f_b(x_{\text{Max}}) - f_b(x_{\text{Min}})] \frac{\pi K}{4} \left(\frac{\lambda_{\text{Water}}}{\pi} \right)^{2-\nu} \int_0^\infty y^{1-\nu} Q_b(y, \tilde{m}) dy, \quad (5.32)$$

with similar relationships for β and b_b . Actually, because we carry out the integrations numerically, the upper limit is some large value (here the maximum value of x is typically 2500, but sometimes 6300). Similar computations were carried out for b_b and β . The results with $\nu = 3$ are shown in Figure 5.51 for b , b_b and $\beta(20^\circ)$ with $\tilde{m} = 1.05 + 0.000i$. The results shown in Figure 5.51 are quite typical for power-law distributions. We note that the scattering coefficient and the backscattering coefficient are mostly determined by particles in different regions of the size distribution: 80% of b is determined from $6 < x < 100$, while 80% of b_b is determined by $0.7 < x < 13$. Note that near the center of the visible spectrum (550 nm or $\lambda_{\text{Water}} \approx 413$), x is related to the particle diameter through $D \approx 0.13x$, so the size range determining most of b and b_b are $0.8 \lesssim D \lesssim 13 \mu\text{m}$ and $0.09 \lesssim D \lesssim 1.7 \mu\text{m}$, respectively. The size range of importance for $\beta(\Theta)$ depends considerably on Θ . For $\Theta > 90^\circ$, $\beta(\Theta)$ is determined by approximately the same size range as b_b , while for $\Theta \lesssim 10^\circ$, $\beta(\Theta)$ is determined by approximately the same size range as b . In between, there is a gradual decrease in the size range of maximum influence as Θ increases. The VSF for an intermediate angle 20° is shown in the figure.

Letting $y = x\Theta$ and noting that $dN = (K/D^{\nu+1})dD$, we find that

$$\beta(\Theta) = K \left(\frac{\lambda}{\pi} \right)^{(2-\nu)} \Theta^{(\nu-4)} \int_0^\infty y^{(1-\nu)} J_1^2(y) dy.$$

The integral is finite as long as $1 < \nu < 4$, so we see that $\beta(\Theta) \propto \Theta^{(\nu-4)}$. (Note that if $\nu \geq 4$ there are so few large particles that diffraction no longer dominates at small angles and the fact that the integral diverges is irrelevant.) It was mentioned earlier that at small angles the Petzold phase function varied according to $P(\Theta) \propto \Theta^{-m}$ with $m > 0$. So according to this analysis, $\nu - 4 = -m$ or $m = 4 - \nu$, and ν can be estimated from m and vice versa. For the Petzold Phase-T phase function, $m \approx 1.4$, so this suggests that $\nu \approx 2.6$.

How are the size ranges in Figure 5.51 changed by adding absorption? Figure 5.52 provides a plot similar to Figure 5.51 (with β omitted) for $\tilde{m}_r = 1.05$ with $\tilde{m}_i = 0, 0.002$, and 0.010 . The introduction of absorption moves the range of influence for b to slightly smaller particle sizes, but changes the upper limit for b_b significantly, from $x = 13$ to about $x = 5$ as \tilde{m}_i increases for 0 to 0.010 (larger value of D decreases from 1.7 to about $0.65 \mu\text{m}$). Thus, introducing absorption can significantly reduce the size range governing backscattering.

How are the size ranges in Figure 5.51 changed by changing \tilde{m}_r ? Figure 5.53 provides a plot similar to Figure 5.52 for $\tilde{m}_r = 1.025, 1.050$, and 1.075 with $\tilde{m}_i = 0$. Varying \tilde{m}_r strongly affects the range of influence for b . Changing \tilde{m}_r from 1.025 to 1.075 changes the range of influence for b from $4 < x < 63$ to $10 < x < 200$, while the upper limit for b_b increases from 8 to 25 . Thus, increasing \tilde{m}_r shifts the range of influence to larger sizes for b , and shifts the upper limit of the size range for b_b .

Finally, how are the size ranges in Figure 5.51 changed by changing ν ? Figure 5.54 provides a plot similar to Figure 5.52 for $\tilde{m} = 1.05 + 0i$. When ν increases from 3 to 3.5 , the range of influence for both b and b_b moves to smaller particle sizes, but decreasing ν to 2.5 greatly increases the upper limit of the range of influence to the point where it is approximately $x = 800$ for b and $x \approx 1000$ for b_b . For b_b this means the upper limit of the range of influence is $D \approx 130 \mu\text{m}$! The reason for this is easy to understand. For $\nu = 2$ the integral diverges, i.e., its value is ∞ . This means that the closer ν gets to 2 , the larger the range contributing significantly to the scattering quantities.⁴⁸ The peculiar shape of the b_b curve for $\nu = 2.5$ can be traced to the shape of Q_{bb} as a function of x (see Figure 5.32). There is a maximum in Q_{bb} ranging from about $x = 200$ to $x = 1000$. For $\nu = 3.5$ this size range doesn't contribute, but for $\nu = 3$ it starts to have some effect. When ν is reduced to 2.5 , there are sufficient numbers of particles in this size range that it becomes very important and the peculiar shape emerges. For completeness we include the values of the integrals for all of the cases in Figures 5.51 to 5.54 in Table 5.4.

Considering that x over the visible spectrum varies by a factor of ~ 2 , these figures clearly suggest that for $\nu \gtrsim 3$, and the low indices of refraction generally found for marine particles, the size distribution only needs to range over $0.5 \lesssim D \lesssim 15 \mu\text{m}$ for the spectral variation of b_{Part} to be nearly $\propto \lambda_{\text{Water}}^{(2-\nu)}$, as it is when the sizes vary over the entire range of diameters ($0 \rightarrow \infty$). Only if the distribution is narrow, or if either x_{Max} or x_{Min} falls near the steep parts of the curves in these figures will the spectral dependence depart

⁴⁸The astute reader will note that the curves in Figure 5.54 do not reach a value of 1.0 . That is because the integrals have been computed numerically only up to $x \approx 6300$. However, it is easy to estimate the contribution from $6300 < x < \infty$ because Q_b , which is equal to Q_c if $\tilde{m}_i = 0$, and is given approximately by Eq. (5.26), becomes oscillatory around a more or less constant value (2). Likewise, Q_{bb} becomes oscillatory around a value of ~ 0.0055 (Figure 5.32).

Table 5.4: Values of the integrals encountered in Figures 5.51 to 5.54.

\tilde{m}	ν	$\int_0^\infty y^{1-\nu} Q_b(y, \tilde{m}) dy$	$\int_0^\infty y^{1-\nu} Q_{bb}(y, \tilde{m}) dy$
$1.025 + 0.000i$	3.0	1.035×10^{-1}	3.518×10^{-4}
$1.050 + 0.000i$	3.0	2.068×10^{-1}	1.476×10^{-3}
$1.050 + 0.002i$	3.0	1.870×10^{-1}	1.402×10^{-3}
$1.050 + 0.010i$	3.0	1.522×10^{-1}	1.370×10^{-3}
$1.050 + 0.000i$	2.5	1.299×10^{-0}	3.580×10^{-3}
$1.050 + 0.000i$	3.5	4.964×10^{-2}	1.160×10^{-3}
$1.075 + 0.000i$	3.0	3.091×10^{-1}	3.535×10^{-3}

significantly from the power-law spectrum.

Using Eq. (5.32) to numerically compute b_{Part} requires some care. Given the cumulative size distribution $N_{>D} = kD^{-\nu}$ we found earlier that

$$\frac{dN}{dD} = \frac{K}{D^{\nu+1}} = \frac{\nu k}{D^{\nu+1}}.$$

If D is in μm , then k is the total number of particles per ml larger than 1 μm . However, the units of k are *not* ml^{-1} , they are $\text{ml}^{-1}(\mu\text{m})^\nu$. The corresponding K also has units $\text{ml}^{-1}(\mu\text{m})^\nu$. To use Eq. (5.32) and get the answer in m^{-1} , we need to convert all of the units of length to meters. Thus, the units for k must be changed from $\text{ml}^{-1}(\mu\text{m})^\nu$ to

$$\frac{(10^{-6}\text{m})^\nu}{10^{-6}\text{m}^3} \quad \text{or} \quad \frac{dN}{dD} = \frac{\nu k (10^{-6})^{(\nu-1)}}{D^{\nu+1}},$$

where now k is numerically equal to the number of particles per ml with diameters larger than 1 μm . Then Eq. (5.32) becomes

$$b_{\text{Part}} = 10^{-6} \times [f_b(x_{\text{Max}}) - f_b(x_{\text{Min}})] \frac{\nu k}{4} \pi^{(\nu-1)} \lambda_{\text{Water}}^{(2-\nu)} \int_0^\infty y^{1-\nu} Q_b(y, \tilde{m}) dy,$$

where λ_{Water} is in now μm , and direct substitution gives b_{Part} in m^{-1} . Thus, for example, in the case of $\nu = 3$, and taking $x_{\text{Min}} = 0$ and $x_{\text{Max}} = \infty$, we have for $\lambda_{\text{Water}} = 0.375 \mu\text{m}$ (500 nm in air)

$$b_{\text{Part}} = 10^{-6} \times \frac{3k\pi^2}{4\lambda_{\text{Water}}} \int_0^\infty y^{-2} Q_b(y, \tilde{m}) dy = 1.97 \times 10^{-5} k \int_0^\infty y^{-2} Q_b(y, \tilde{m}) dy.$$

If $\tilde{m} = 1.05 + 0.000i$, the integral is ~ 0.2 , so $b_{\text{Part}} \approx 4 \times 10^{-6} k$. Fitting the distribution in Figure 5.49 to a single segment, $N_{>D} = kD^{-\nu}$, yields $\nu \sim 3$ and $k \sim 40,000$ particles per ml (although the fit would be rather poor) so this would predict that $b_{\text{Part}} \sim 0.16$

m^{-1} . This is close to the observed scattering coefficient in the Tongue of the Ocean surface waters at the time the size distribution samples were obtained ($0.10 \rightarrow 0.14 \pm 0.03 \text{ m}^{-1}$).

We complete this section by looking at the scattering phase functions for the power-law size distribution with $\nu = 3$. For this case, with $\tilde{m} = 1.050 + 0.000i$, Figure 5.51 shows that the scattering and backscattering coefficients are determined mostly by particles in the $x = 1$ to 1000 range. In Figure 5.55 we provide the phase functions for distributions ranging over $0 \rightarrow x_{\text{Max}}$ for $x_{\text{Max}} = 1, 10, 100$, and 1000. Note how the phase function changes shape from Rayleigh-like to more and more sharply peaked in the forward direction, characteristic of scattering by larger and larger particles. Notice also that the phase function changes very little from $x_{\text{Max}} = 100$ to 1000, except in the extreme forward and backward directions. (The increase in the extreme backward direction is called the “glory” and is characteristic of spherical particles, and disappears for other particle shapes.) For large x_{Max} the phase function varies as $P(\Theta) \propto \Theta^{-m}$, where $m \approx 0.87$. This is consistent with diffraction theory that suggests $m = 4 - \nu$, which in this case is unity.

Figure 5.56 shows how the phase function for $x_{\text{Max}} = 1000$ varies with \tilde{m} . The phase function for $\Theta \gtrsim 10^\circ$ generally increases with increasing \tilde{m}_r . Figure 5.56 also shows that \tilde{m}_i has little influence on the phase function. In addition, Figure 5.56 compares the Petzmas phase function (described in Section 5.3.3.2) to the others. This comparison shows that the $\nu = 3$ phase functions have the same general shape as Petzmas; however, the values of B depend significantly on \tilde{m} , with B for $\tilde{m} = 1.05 + 0i$ being about half of that for Petzmas. It is gratifying that such a simple model comes so close to describing what is actually observed in nature.

Finally, the degree of polarization (DOP) of the scattering by this size distribution is provided in Figure 5.57. Recall that the DOP is defined as $-P_{12}/P_{11}$, where the subscripts indicate the elements of the phase matrix (Chapter 3). Included for comparison is the DOP of Rayleigh scattering for an isotropic scatterer (e.g., a very small homogeneous sphere). Several conclusions are clear: (1) when \tilde{m}_r is small, e.g., near 1, the DOP is very close to Rayleigh scattering; (2) in this case absorption has little influence on the DOP; and (3) $\tilde{m}_r \gtrsim 1.075$ are required for significant departures from the DOP of isotropic Rayleigh scattering. Thus, we would expect that the (P_{11} -normalized) phase matrix for scattering by Case 1 waters would be similar to Rayleigh scattering. That is indeed the case (Section 5.3.3.4).

In summary, we have seen that including a distribution of particle sizes into the theory of the IOPs is straightforward. The procedures for estimation of \tilde{m} for monodisperse phytoplankton cultures in Section 5.4 are easily extended to polydisperse systems. For size distributions similar to those observed in Case 1 waters, the effective size range influencing the various scattering properties is easily carried out (Figures 5.51 – 5.54). Interestingly,

the same regions of the size distribution rarely contribute to scattering and backscattering: b is mostly determined by larger particles ($> 1 \mu\text{m}$), while b_b is mostly determined by smaller particles ($< 1 \mu\text{m}$). Thus, measurements of b_b made either in-situ or remotely, will typically provide information about particles in smaller size ranges. The scattering phase functions are mostly influenced by \tilde{m}_r , with \tilde{m}_i having little impact; however, the value of ν can have a large effect, especially when it falls below $\nu = 3$ bringing relatively more larger particles into the scattering process, and these scatter mostly at smaller angles. The DOP of the scattering by the low- \tilde{m}_r particles that are found in the oceans is very close to that of Rayleigh scattering by isotropic spheres. It is tempting to try to model the IOPs by considering a size distribution consisting of a number of log-normal functions each representing a particular kind of particle (e.g., as in Figure 5.50) to develop a systematic model for the particle phase function as described in Section 5.3.3.5; however, this is a topic for future research and will not be attempted here.

Here, we have only looked at spherical particles; however, the calculations can be extended to non-spherical particles with simple shapes using publicly-available computer codes. The details of the IOPs will differ somewhat from those for spheres, but the general nature of the behavior with size, etc., will be similar to that seen here for spheres.

5.6 Summary

Here, we have discussed those IOPs of natural waters that play a direct role in ocean color remote sensing. For simplicity we have limited the discussion to Case 1 waters. Principally, we described in detail the influence of particles containing Chlorophyll *a* on the IOPs. We applied scattering theory to the known physical and chemical properties of particles in the water to understand the empirical measurements in a more fundamental manner. The reader should understand that this is a very dynamic subject and that we have presented only the most rudimentary ingredients, i.e., those essential for understanding the foundations of ocean color. However, the material presented should provide a solid departure point for the examination of more up-to-date research.

5.7 Appendix 1: The Fournier-Forand Phase Function

In Chapter 3 we developed an analytic the expression for the extinction efficiency, that we referred to as the “van de Hulst approximation.” In this approximation the interference, in the exact forward direction, of light passing through a particle with the light passing by (diffracted by) the particle was computed, and the extinction cross section determined with

the help of the optical theorem. Of course, interference occurs at other scattering angles as well as at $\Theta = 0$. When interference is carefully examined at $\Theta \neq 0$ the result is what is called the anomalous diffraction approximation. Application of this approximation to the scattering by a collection of non-absorbing particles with a power-law size distribution provides the Fournier-Forand (FF) phase function:⁴⁹

$$P_{FF}(\Theta) = \frac{1}{4\pi(1-\delta)^2\delta^\alpha} \left[\alpha(1-\delta) - [(1-\delta^\alpha) + \delta(1-\delta^\alpha) - \alpha(1-\delta)] \sin^{-2} \left(\frac{\Theta}{2} \right) \right] + \frac{1-\delta_{180}^\alpha}{16\pi(\delta_{180}^\alpha-1)\delta_{180}^\alpha} (3\cos^2\Theta - 1), \quad (5.33)$$

where

$$\alpha = \frac{4-\nu}{2} \quad \text{and} \quad \delta = \frac{4}{3(\tilde{m}-1)^2} \sin^2 \left(\frac{\Theta}{2} \right).$$

The quantity \tilde{m} is the (real) refractive index of the particle, ν is the slope parameter in the power-law size distribution,⁵⁰ and δ_{180} is the value of δ when $\Theta = 180^\circ$. The backscattering probability for this phase function is

$$B_{FF} = 1 - \frac{1 - \delta_{90}^{\alpha+1} - 0.5(1 - \delta_{90}^\alpha)}{(1 - \delta_{90})\delta_{90}^\alpha}, \quad (5.34)$$

where δ_{90} is the value of δ when $\Theta = 90^\circ$. For small angles,

$$P_{FF}(\Theta) \sim \frac{1}{\Theta^{(4-\nu)}},$$

so for $\nu < 4$ the phase function is infinite at $\Theta = 0$. In spite of this, the total scattering coefficient will still be finite as long as $\nu > 2$. This behavior of the small-angle phase function was seen earlier in the diffraction approximation (Footnote 45).

Other than the fact that it has an analytical representation (with only two parameters, \tilde{m} and ν), the main reason for the popularity of the FF phase function in marine optics is that for certain sets of parameters it yields an excellent analytic fit to the Petzold (and

⁴⁹This result is not exact even in the anomalous diffraction approximation, as further approximations were necessary to make the integral over particle size tractable. In addition, the result is applicable only to non-absorbing particles.

⁵⁰Recall that the power-law size distribution is given by

$$\frac{dN}{dD} \propto \frac{1}{D^{\nu+1}}$$

where dN is the number of particles per unit volume with (volume-equivalent spherical) diameters between D and $D + dD$.

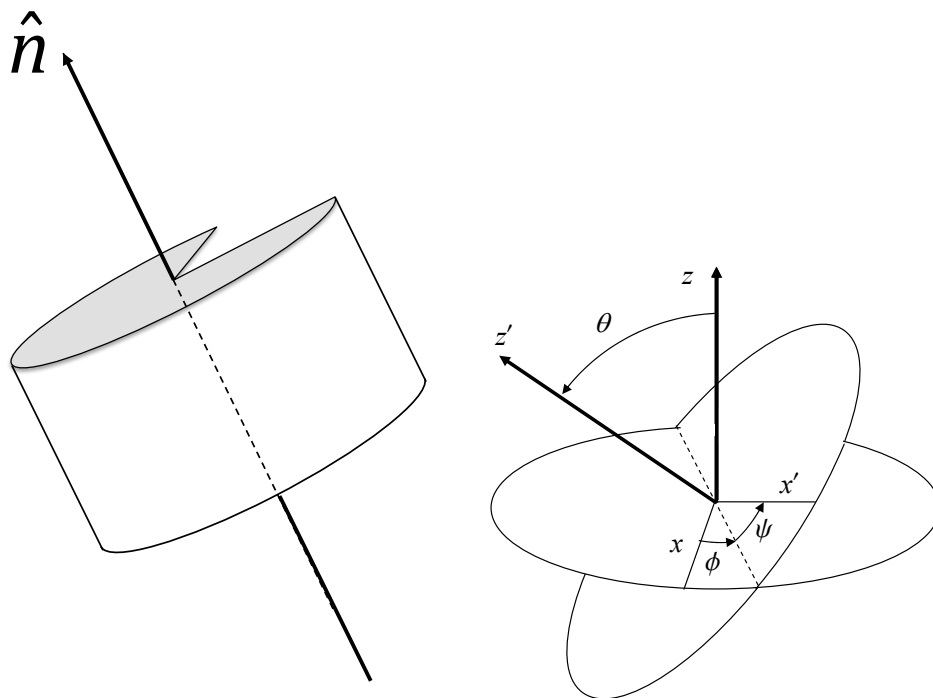
Petzmas) phase function(s). This is shown in Figure 5.58, where the FF parameters have been chosen so that all three phases functions have the same backscattering probability. (Note that the choice of \tilde{m} and ν for a given B_{FF} is *not* unique.) Clearly an excellent fit to both the Petzold and Petzmas phase functions over the entire range of Θ is obtained. In particular, the FF phase function differs from Petzmas for $\Theta > 90^\circ$ by less than 10%. The FF phase function is the default for particles in the HydroLight radiative transfer code.

In spite of the success of reproducing the Petzold and Petzmas phase functions, one must be careful in interpretation of the parameters (especially \tilde{m}) if the FF is used to retrieve them from experimental light-scattering data. This is particularly true when backscattering data are considered, i.e., Eq. (5.34). An example of the erroneous behavior of FF in the backward direction, compared to Mie theory, is provided in Figure 5.59, which shows the phase functions for a collection of power-law distributed spherical particles with refractive index $\tilde{m} = 1.05$ and $\nu = 3$. Since Mie theory is exact, we see that the FF formulation strongly overestimates the backscattering in this case if the correct values of $\tilde{m} = 1.05$ and $\nu = 3$ are used in Eq. (5.33). To closely approximate the backscattering observed in the Mie calculation, using the correct ν requires $\tilde{m} \approx 1.01180$ in the FF formula. A cursory examination of other cases shows similar behavior. Although this degrades the value of the FF formulation, it is not surprising that there should be significant error at the larger scattering angles, as the FF phase function is based on anomalous diffraction theory, which is valid only for small-angle scattering.

5.8 Appendix 2: Particle Orientational Effects and Averaging

In the text we often referred to situations in which non-spherical particles were in *random* orientation; however, we never defined with any precision what random orientation means, nor did we examine any effects of non-random particle orientation. This is remedied in this appendix, where we describe how to specify the orientation of a particle and then compare random and non-random orientation using scattering by a thin disk as an example.

In order to ask how an object is oriented, we need to ask, oriented to what? So, we pick a certain orientation (the “basic” orientation) and specify a new orientation relative to the basic one. Consider the object on the left in the figure above. It is a cylinder with a sector cut out (so the object has no rotational symmetries). Let’s specify the basic orientation as that in which \hat{n} points toward the north pole of a spherical coordinate system, i.e., is aligned with the z axis, and where one edge of the sector is along the y axis. Then any other orientation can be defined by specifying the direction the new \hat{n} is pointing, along



with a rotation by some angle *around* the new vector \hat{n} . One way to accomplish this in a systematic manner is shown on the right panel. Here the base orientation is the disk in the x - y plane of the unprimed cartesian coordinate system. The object is first rotated around the z axis through an angle ϕ . The z axis is then rotated through an angle θ in a plane normal to the dashed line in the figure (the position to which the x axis was rotated in the first rotation and called *line of nodes*) to a new z axis labeled z' . Finally, the object is rotated around the z' axis through an angle ψ .⁵¹ The angle ϕ , θ , and ψ are called *Euler angles* and describe the orientation of a rigid object relative to some fixed orientation or fixed coordinate axes. Note, if the sector were not cut out of the object in the figure, i.e., the object is a solid cylinder, then the orientation would independent of ψ , and the direction of \hat{n} is all that would be required to specify the orientation.

Using Euler angles we can specify the probability that the orientation is given by (ϕ, θ, ψ) , where $\phi - d\phi/2 \leq \phi \leq \phi + d\phi/2$, etc., through a probability density $p(\phi, \theta, \psi)$, with

$$\int_0^{2\pi} d\phi \int_0^{2\pi} d\psi \int_0^\pi p(\phi, \theta, \psi) \sin \theta d\theta = 1.$$

⁵¹In a spherical coordinate system defined in the usual manner, the polar angle of the spherical system corresponds to the angle θ , and the azimuth angle of the spherical system corresponds to $\phi - \pi/2$.

If all orientations are equally probable (random orientation), clearly

$$p(\phi, \theta, \psi) = \frac{1}{8\pi^2}.$$

The orientational average of a quantity $Q(\phi, \theta, \psi)$ that depends on orientation is

$$\langle Q \rangle = \int_0^{2\pi} d\phi \int_0^{2\pi} d\psi \int_0^\pi Q(\phi, \theta, \psi) p(\phi, \theta, \psi) \sin \theta d\theta,$$

and for random orientation

$$\langle Q \rangle = \frac{1}{8\pi^2} \int_0^{2\pi} d\phi \int_0^{2\pi} d\psi \int_0^\pi Q(\phi, \theta, \psi) \sin \theta d\theta.$$

In what follows the object we will be considering is a disk, for which the probability density is independent of ψ . For disks,

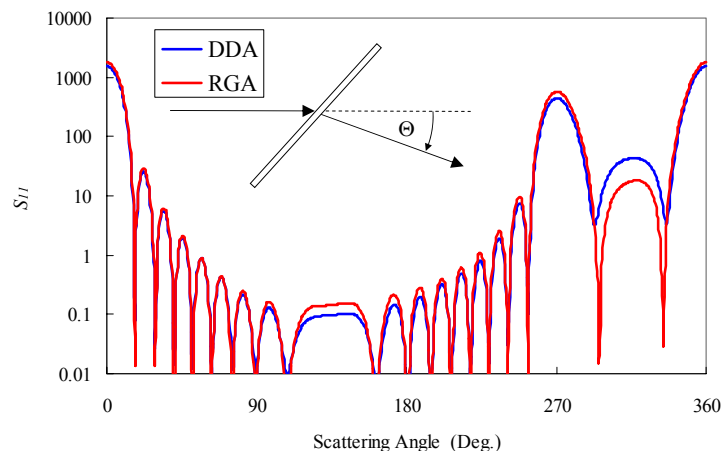
$$\langle Q \rangle = \int_0^{2\pi} d\phi \int_0^\pi Q(\phi, \theta) p(\phi, \theta) \sin \theta d\theta,$$

with

$$\int_0^{2\pi} d\phi \int_0^\pi p(\phi, \theta) \sin \theta d\theta = 1,$$

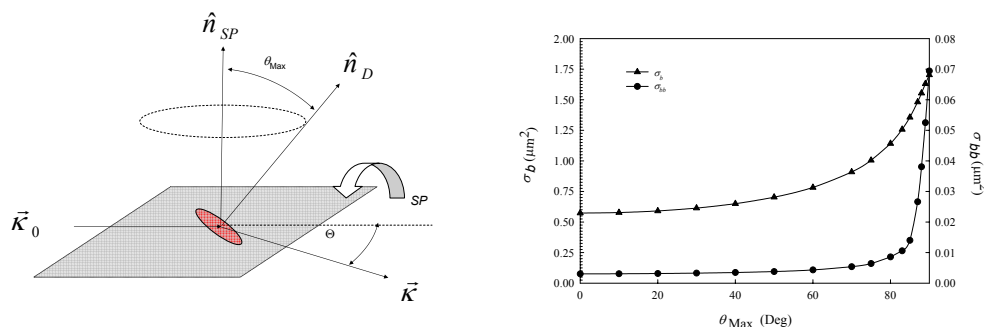
and for random orientation $p(\phi, \theta) = 1/4\pi$.

In general, orientational averaging in computations is carried out by simply solving the scattering problem for a large number of orientations, and computing the averaging integral(s) numerically. To illustrate some of the effects of averaging for a particle for which the scattering is strongly affected by orientation, we will consider scattering by a disk. To make the problem tractable, we show in the figure below that the Rayleigh-Gans approximation (RGA) provides an excellent approximation to the exact scattering solution provided by the discrete-dipole approximation (DDA). Since the RGA solution is analytical, orientational averaging is relatively simple to carry out. The figure shows the quantity $S_{11}(\Theta)$, where the volume scattering function $\beta(\Theta) = nS_{11}(\Theta)/\kappa^2$, for scattering by a disk oriented as shown with its normal parallel to the plane of scattering (the plane of the inserted figure) making an angle of 45° with the incident beam. The diameter of the disk is $2.7 \mu\text{m}$, the thickness is $0.1352 \mu\text{m}$, the refractive index is 1.20, and the wavelength of the radiation is 700 nm. The scattering angle Θ ranges from 0 to 2π . Clearly, the RGA solution is very close to the DDA over many orders of magnitude and can be used to demonstrate some effects of averaging. It is particularly interesting to note the presence of the large scattering peak near $\Theta = 270^\circ$. This peak is in the position for specular reflection off the surface of the disk and propagation vertically in the figure. It is nearly as strong



as the forward peak ($\Theta = 0$), and as such will have a significant affect in orientational averaging.

As an example of the possible influence of orientation, consider scattering by a collection of disks whose orientation is restricted as shown in the figure below. The left panel shows the geometry, where the normal to the disk is uniformly distributed over the surface of a spherical cap of half-angle θ_{Max} . When $\theta_{\text{Max}} = 90^\circ$ the disks are in random orientation. The shaded area is the scattering plane (SP), which is horizontal in this drawing. Typically, when making a measurement of scattering with a bench top scattering photometer it is assumed that the particles are in random orientation, so the scattering plane is representative of any other plane in which one might care to make measurements. The right panel provides σ_b and σ_{bb} for a disk of the dimensions in the figure above (at 700 nm) as a function of θ_{Max} . Clearly, if any process restricts the disk normal from being in the scattering plane ($\theta_{\text{Max}} = 90^\circ$), even by a small amount, measurements of both the total scattering and the backscattering will be drastically reduced. When the scattering plane is horizontal (as for most bench scattering photometers), gravitational settling of the particles could provide such an orientation restriction, causing an underestimation of both both σ_b and σ_{bb} . (It is just such an orientation restriction, resulting from gravitational settling of ice-crystals, that gives rise to atmospheric halo phenomena such as Sun dogs, Sun halos and Moon halos.) If θ_{Max} is close to 0 (or even as large as $\sim 45^\circ$) the error in measuring σ_b would be close to a factor of three and σ_{bb} would be too low by a factor of 20 or more. The primary cause of this particular orientation effect is obviously the strong scattering peak at $\Theta = 270^\circ$ that comes into the averaging when the disk normal is parallel to the scattering plane. In cases such as this, one would have to make measurements over a number of scattering planes from the one shown in the figure to one perpendicular to it and intersecting it along \vec{k}_0 , in order to obtain accurate measurements of σ_b and σ_{bb} .



5.9 Bibliographic Notes

5.2.1 The Absorption Coefficient

The spectra in Figure 5.2 are from a combination of Pope and Fry [1997] and Sogandares and Fry [1997], and the description of the measurement principle described in Footnote (2) follows that of Fry et al. [1992]. Although these measurements were considered the “gold standard” for the absorption coefficient, they have been called into doubt near the minimum and in the UV by two very sensitive methods. The first employs irradiance measurements taken in very clear ocean water in the South Pacific near Easter Island to infer the absorption coefficient through inversion (Chapter 6) [Morel et al., 2007], while the second employs a technique similar to Sogandares and Fry [1997], but reportedly much more sensitive [Cruz et al., 2009]. Both believe that the Pope and Fry [1997] and Sogandares and Fry [1997] coefficients are too large in the UV, with Cruz et al. [2009] suggesting they are too large from the absorption minimum (~ 420 nm) into the UV. There are no questions regarding the absorption coefficient at wavelengths greater than about 430–450 nm in Figure 5.2. Figure 5.3 is from Hale and Querry [1973]. The temperature (and salinity) dependence of the absorption coefficient were measured by Pegau et al. [1997].

5.2.2 The Scattering of Pure Water and Sea Water

The theory and measurements of the scattering of water and sea water through 1974 is thoroughly discussed in Morel [1974]. The measurements in Tables 5.1 and 5.2 were taken from Morel [1968] based on measurements in Morel [1966]. Their stated uncertainty (for $\beta(90^\circ)$) is $\pm 2\%$; however, Morel’s measurements of δ are considered to be too high. The recommended value now is $\delta = 0.039$, the lowest value determined by Farinato and Rowell [1976]. Recent calculations of $\beta(90^\circ)$ using Eq. (5.5) by Zhang and Hu [2009] and the thermodynamic properties of water agree with Morel’s measurements (within their

experimental uncertainty) using $\delta = 0.039$. In addition, [Zhang et al. \[2009\]](#) used the fluctuation theory of light scattering along with the thermodynamics of ionic solutions and sea water to explain the roughly 30% increase in scattering as the salinity varies from 0 to 38.5 ‰. The comments regarding the contribution to b_b in the waters near Hawaii were based on the analysis of [Gordon et al. \[2009\]](#).

5.2.3 The Raman Scattering of Pure Water

The Raman scattering coefficient for illumination at 488 nm was deduced by [Marshall and Smith \[1990\]](#). They actually measured the differential scattering cross section per molecule integrated over the entire Raman band using illumination polarized normal to the scattering plane and scattering at 90° analyzed with a polarizer with pass direction also normal to the scattering plane (proportional to $\beta_{r \rightarrow r}$ using the notation in Appendix 3 to Chapter 3 and integrated over the whole Raman band). They state a relative uncertainty of approximately 12% in their measurement. They also review previous measurements. The shape of the Raman spectrum as a function of ν_s in Figure 5.5 was taken from [Chang and Young \[1974\]](#) who measured $\beta_{r \rightarrow r}$ and $\beta_{r \rightarrow \ell}$ at a scattering angle of 180° with high spectral resolution, but without absolute calibration and with an illumination wavelength of 460 nm. Noting that $\beta_{r \rightarrow r}$ and $\beta_{r \rightarrow \ell}$ are independent of scattering angle, and that only $b_R(\nu_0 \rightarrow \nu_0 - \nu_s)$ depends on the value of ν_0 , the [Chang and Young \[1974\]](#) data can be calibrated using the value of b_R deduced by [Marshall and Smith \[1990\]](#) at 488 nm. This combination produced Figure 5.5. Figure 5.4 was prepared directly from the [Chang and Young \[1974\]](#) measurements. The spectral variation of b_R with wavelength was measured by [Bartlett \[1996\]](#), who also measured b_R for the Raman emission band at an excitation wavelength of 488 nm and found $(2.7 \pm 0.2) \times 10^{-4} \text{ m}^{-1}$, not significantly different from [Marshall and Smith \[1990\]](#). In addition, these authors concluded that, within the uncertainty of their measurements, salinity has no effect on the Raman scattering coefficient.

5.3 The IOPs of the Constituents of Natural waters

The classification of natural waters into Case 1 and Case 2 follows that presented in [Gordon and Morel \[1983\]](#).

5.3.1 Measurement of the Absorption of the Constituents of Case 1 waters

The “shiny tube” method for measuring the absorption by marine particles, based on collecting much of the light scattered by the particles, was inspired by the work of [Shibata \[1958\]](#) and [Yentsch \[1962\]](#), both of whom developed methods to collect the scattered light. The actual shiny tube device was developed by Zaneveld and co-workers (e.g., see [Zaneveld et al. \[1992\]](#)). The filter pad technique was first used by [Yentsch \[1962\]](#) and further

developed by Mitchell and Kiefer [Mitchell and Kiefer, 1984, 1988].

5.3.2 Absorption of the Constituents of Case 1 waters

The description of particle absorption is taken mostly from Bricaud et al. [1998]. The data for Figure 5.10 were provided by A. Bricaud.

5.3.3.1 Total Scattering Coefficient of Particles in Case 1 waters

The schematic of the small angle scattering device in Figure 5.12 is similar to that in Bauer and Ivanoff [1965] and Bauer and Morel [1967].

The early work on b_p , e.g., Eqs. (5.12) and (5.13) along with Figures 5.13 and 5.14 was taken directly from Gordon and Morel [1983]. The later work, i.e., the analysis leading to Eq. (5.15), was taken from Loisel and Morel [1998]. The relationship between $c_p(660)$ and POC was linearized here, based on the observation that a linear relation fit the data in Figure 4 of Loisel and Morel [1998] almost as well as the power-law in their analysis. In addition, other workers, e.g., Gardner et al. [1993], proposed a linear relationship between POC and $c_p(660)$. Examples of the fact that spectra of c_ϕ are nearly featureless, while b_ϕ is depressed in pigment absorption bands are provided in Bricaud et al. [1983] for several species of plankton. The question concerning the effect of the c -meter's acceptance angle on the measurement of c in Footnote 11 is discussed in Boss et al. [2009].

5.3.3.2 Volume Scattering Function of Particles in Case 1 Waters

The “Phase-T” phase function was measured by Petzold [1972] using a combination of instruments to measure the full volume scattering function near 514 nm [Mobley, 1994]. The measurements were made in San Diego (CA) Harbor containing very turbid water, but the results are similar to those measured in clearer off shore southern California waters (HAOCE). Other measurements reported by Petzold in clear water have been rejected as characteristic of ocean water because they show a more pronounced increase in scattering between 150° and 180° than that shown by Phase-T in Figure 5.17 which is not borne out by later measurements. The WetLab Avg./Petzmas phase function (Figures 5.16 and 5.17) was taken from Sullivan and Twardowski [2009]. This paper also describes the scattering instrument (MASCOT) used to make measurements of the volume scattering function. The name “Petzmas,” used here to describe the combination of Phase-T and the WetLab Avg., is an abbreviation for Petzold-MASCOT.

5.3.3.3 The Backscattering Coefficient of Particles in Case 1 Waters

The three-angle measurement of $\beta(\Theta)$ fitted to the β_{BZ} was first described by Gordon [1976]. The subscript “BZ” stands for Beardsley-Zaneveld who first used this functional

form in oceanic optics to represent the volume scattering function [Beardsley and Zaneveld, 1969]. The result that $B_p \sim 0.01$ in oligotrophic waters (and the concomitant covariation of b_{bp} and b_p) was described in Dall’Omo et al. [2009], although $B_p = 0.01$ was actually assumed by Gordon and Morel [1983] in explaining the physics underlying the relationship between the blue-green reflectance ratio and the chlorophyll concentration used in the original CZCS pigment algorithm. The $(^{1/2} - ^{1/4} \log_{10} C)$ factor in B_p first appeared in Morel [1988]. The discussion regarding B_p and the spectral parameter v follows that in Morel et al. [2002].

5.3.3.4 Polarization of Scattering in Case 1 Waters

The measured Mueller matrices for scattering by particles suspended in seawater were taken from Voss and Fry [1984] (Figures 5.19, 5.20, and 5.21). Additional scattering Mueller matrix measurements can be found in Fry and Voss [1985]. The degree of polarization measurements for phytoplankton in Figure 5.22 are from Volten et al. [1998].

5.3.3.5 Empirical Model for the Particle Phase Function in Case 1 Waters

The use of two components to interpret the VSF seems to have originated with Brown and Gordon [Brown and Gordon, 1973, 1974], who assigned the components to be either mineral (high refractive index) and planktonic (low refractive index), rather than detritus (small size) and phytoplankton (large size) as assumed here. A similar approach was made by Kopelevich [Kopelevich, 1983; Kopelevich and Mezhericher, 1983] based on a statistical analysis of experimental scattering data. The scheme of modifying the Phase-T/Petzmas phase functions to represent the two components was first used in a detailed study of R by Gordon [1992]. The end member values of B_p for the empirical two-component here (0.0019 and 0.0140) were developed in Morel et al. [2002] based on assumed size distributions and refractive indices for each component, i.e., a physical model of the particles. They used power-law distributions with $\nu = 2.1$ and 3.2 to represent “large” and “small” particles, respectively. In their model both populations have $\tilde{m} = 1.06$, and the scattering computations were carried out for a mixture of oblate and prolate spheroids, although spheres would produce similar results.

5.4.1 The Absorption Coefficient

The material here and Section 5.4.2 is based mostly the work of Morel and Bricaud [Bricaud and Morel, 1986; Morel and Bricaud, 1981]. Anomalous diffraction results formulas for coated particles can be found in Jonasz and Fournier [2007].

5.4.3 The Backscattering Coefficient

The calculations were all done by the author. A detailed review of backscattering by

marine particles can be found in [Stramski et al. \[2004\]](#).

5.4.5 Influence of Particle Shape on IOPs

This Section was taken from [Gordon \[2011\]](#). In addition, the original paper contains examination of the scattering by a two-layer cylinder, leading to another “package effect,” namely, the effect of how the absorbing pigment is distributed *within* the cylinder.

A through investigation of similar effects of non-sphericity has been carried out (with spheroidal particles) by [Clavano et al. \[2009\]](#), including a consideration of size-distributed particles. A study of the influence of internal structures (periodic and aperiodic) on the scattering of disk-like particles (to represent detach coccoliths from *E. huxleyi*) is presented in [Gordon \[2006, 2007b\]](#).

5.5 Influence of The Distribution of Particle Size on IOPs

The general nature of the size distribution in natural waters is from [Brown and Gordon \[1974\]](#). Figure 5.49 is from [Gordon et al. \[1972\]](#).

5.5.1 Analytical Representations of the Size Distribution

More material on the properties of size distribution of the type discussed in this section can be found in Chapter 4. An exhaustive treatment of present knowledge regarding the size distribution and refractive indices of marine particles is in [Jonasz and Fournier \[2007\]](#).

5.5.2.1 Estimation of a_i and \tilde{m} for Size-Distributed Plankton

This is based mostly on [Bricaud and Morel \[1986\]](#).

5.5.2.2 IOPs of Power-Law Distributed Particles

The computations present in this section were carried out by the author using a Mie scattering code. Further computations can be found in [Gordon and Brown \[1972\]](#), [Brown and Gordon \[1973\]](#) and [Brown and Gordon \[1974\]](#).

5.7 Appendix 1: The Fournier-Forand Phase Function

The Fournier-Forand phase function is developed of in [Fournier and Forand \[1994\]](#). (See also [Jonasz and Fournier \[2007\]](#).)

5.10 Figures

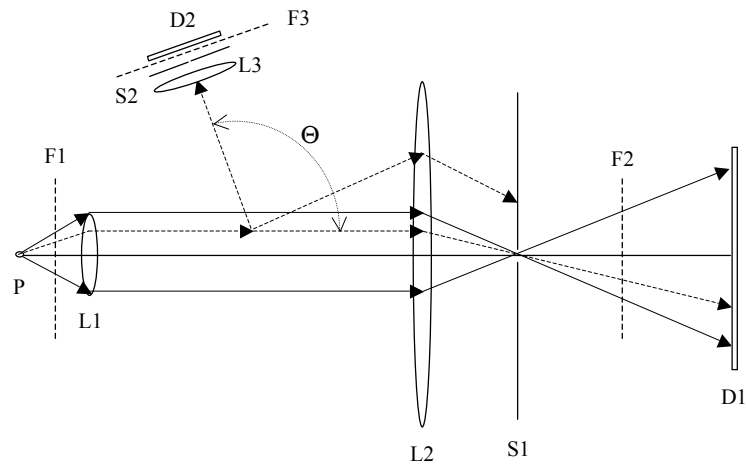


Figure 5.1: Schematic for measuring c and $\beta(\Theta)$, and therefore, a .

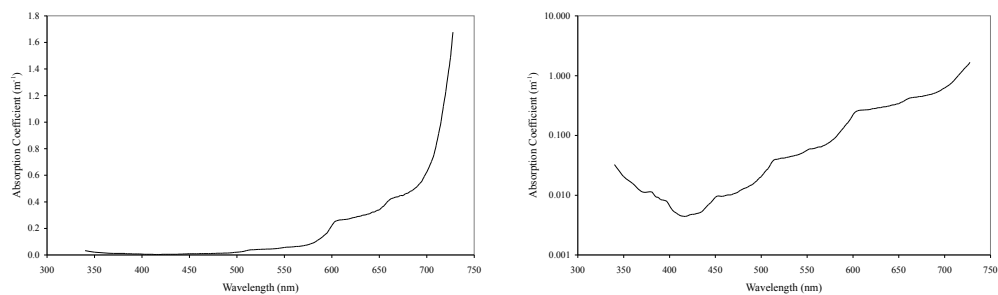


Figure 5.2: Spectra of the absorption coefficient of pure water in the visible.

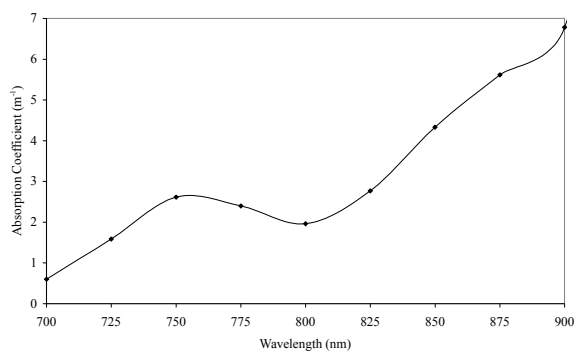


Figure 5.3: Spectra of the absorption coefficient of pure water in the near infrared.

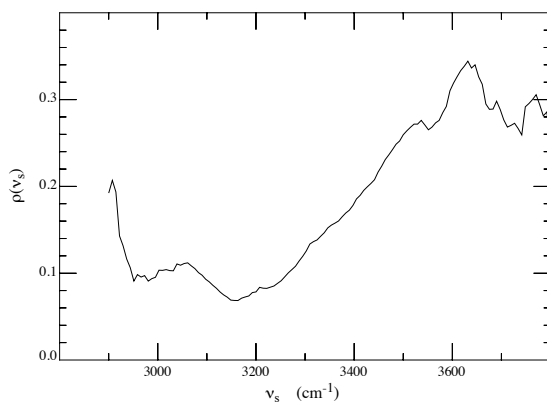


Figure 5.4: Spectrum of the depolarization ratio of Raman scattering. From Gordon [1999].

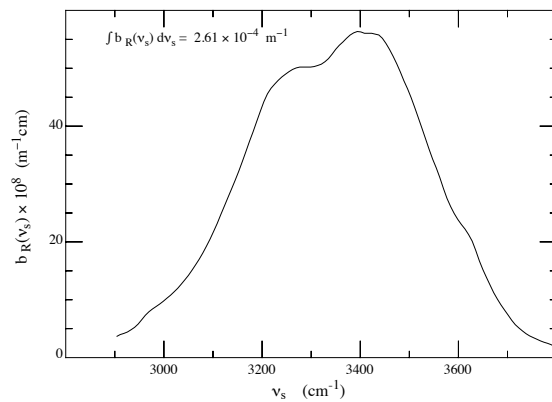


Figure 5.5: Spectrum of the Raman scattering coefficients of pure water. The quantity $b_R(\nu_s)$ refers to $b_R(\nu_0 \rightarrow \nu_0 - \nu_s)$ in the text. From [Gordon \[1999\]](#).

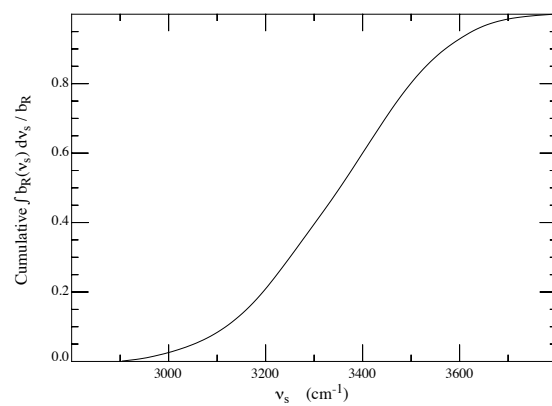


Figure 5.6: Cumulative integration of the data in Figure 5.5 as a function of the upper limit.

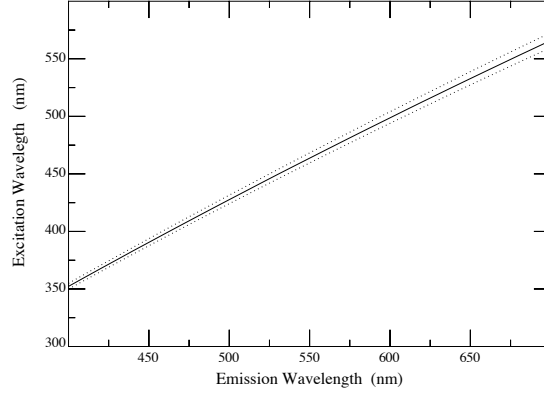


Figure 5.7: Raman excitation wavelength for a given Raman emission wavelength, λ_I . The dotted lines provide the limits for the excitation band resulting in 90% of the Raman emission. From [Gordon and Xu \[1996\]](#).

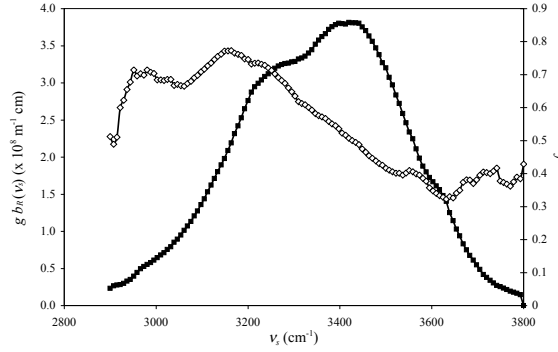


Figure 5.8: Spectrum of the Raman emission (filled symbols) and the polarization factor. The quantity g on the left axis is $(3/16\pi)(1 + 3\rho_R(\nu_s))/(1 + 2\rho_R(\nu_s))$ and the quantity f (open symbols) on the right axis is $((1 - \rho_s)/(1 + 3\rho(\nu_s)))$.

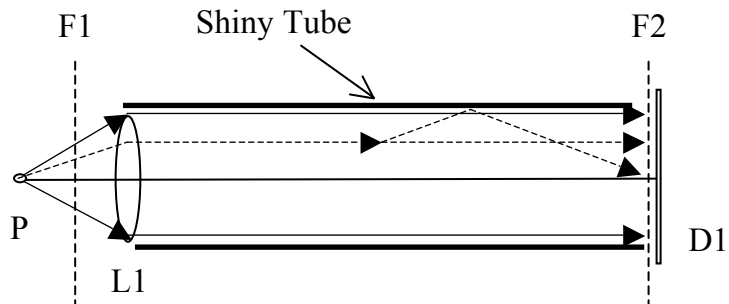


Figure 5.9: Schematic for measuring a for a suspension of particles that scatter mostly in the forward direction.

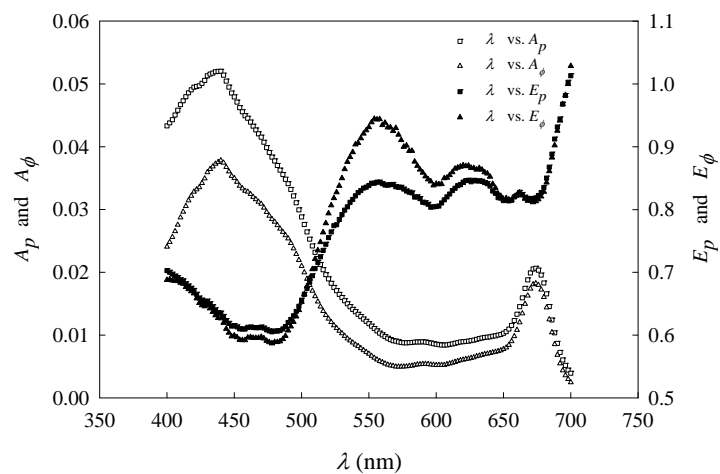


Figure 5.10: Average absorption spectrum of phytoplankton (A_ϕ) and all particles (A_p) with $C = 1 \text{ mg/m}^3$.

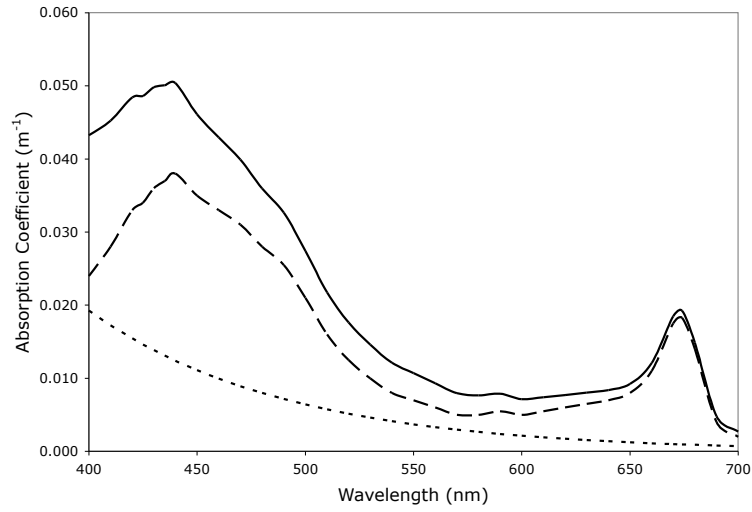


Figure 5.11: Average absorption spectrum of particles (solid curve) with $C = 1 \text{ mg/m}^3$. Dotted curve is for NAP, and dashed curve is for phytoplankton.

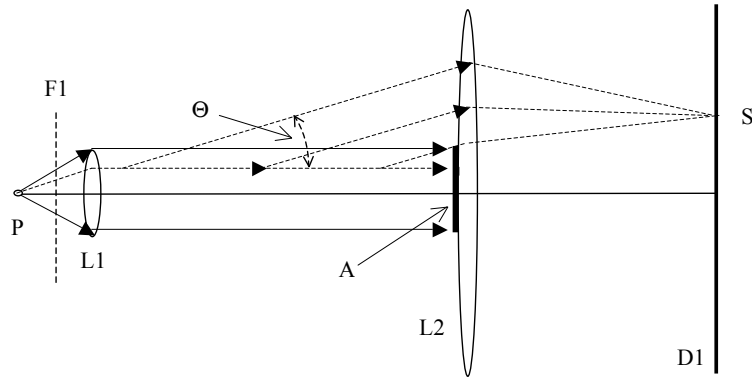


Figure 5.12: Schematic for measuring $\beta(\Theta)$ and/or b for a suspension of particles that scatter mostly in the forward direction. Here, “A” is an absorber to prevent the incident beam from P to reach the detector D1. Clearly, light scattered through an angle Θ from anywhere in scattering volume will be focussed by L2 at the point S (actually in a circle on which S is a part).

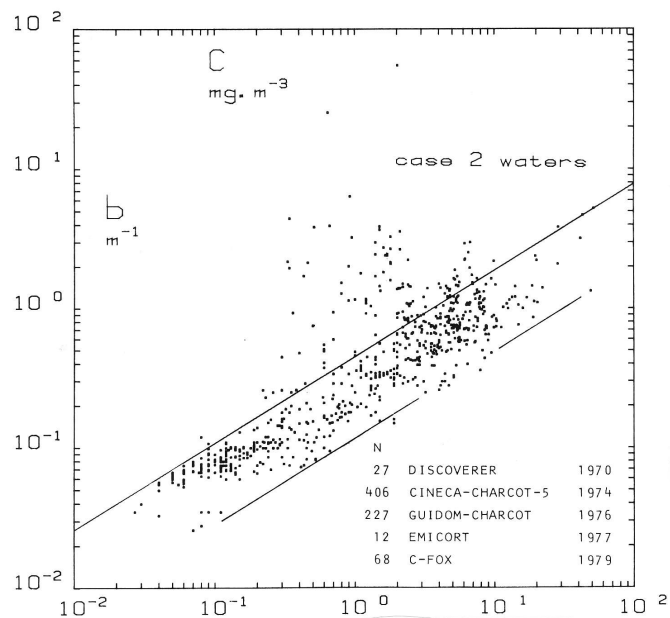


Figure 5.13: Relationship between the scattering coefficient and concentration of Chlorophyll *a* plus Phaeophytin *a* ($C_{C\&P}$) for a variety of oceanic and coastal waters, Case 1 and Case 2 (above the upper line). From [Gordon and Morel \[1983\]](#).

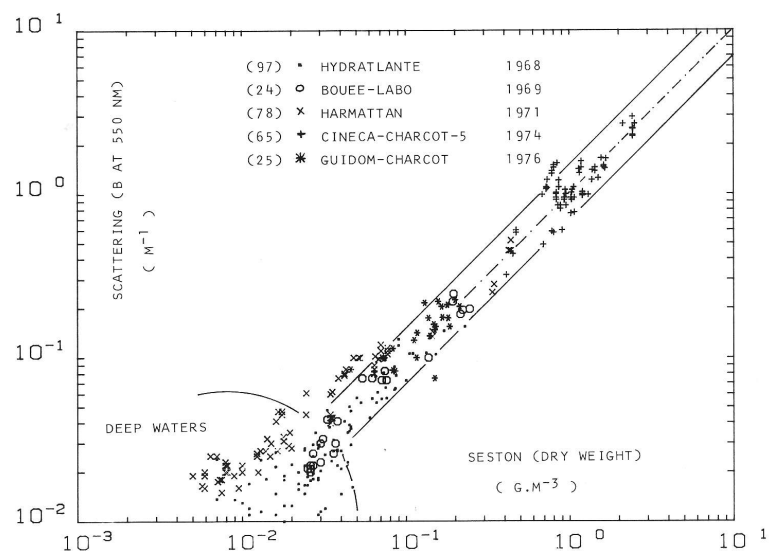


Figure 5.14: Relationship between the scattering coefficient and the seston concentration for a variety of oceanic and coastal waters (Case 1 and Case 2). From [Gordon and Morel \[1983\]](#).

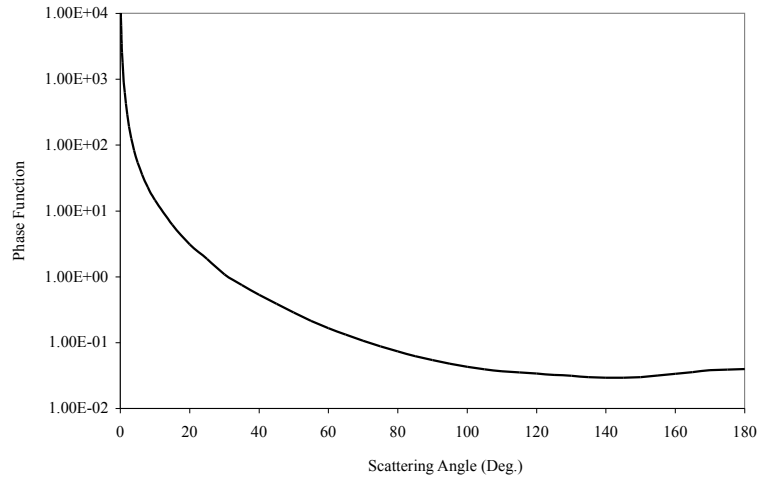


Figure 5.15: Marine particle phase function labeled “Phase-T.”

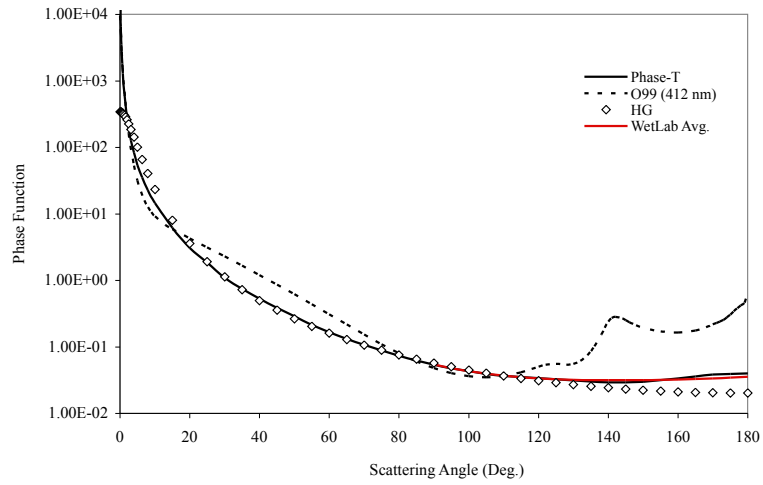


Figure 5.16: Marine particle phase function “Phase-T” compared with other phase functions in this work. “O99” is the phase function for the Oceanic aerosol model in Chapter 4, “HG” is a Henyey-Greenstein function with $g = 0.925$, and “WetLab Avg.” is the *average* phase function for marine particles (normalized to Phase-T at 90°).

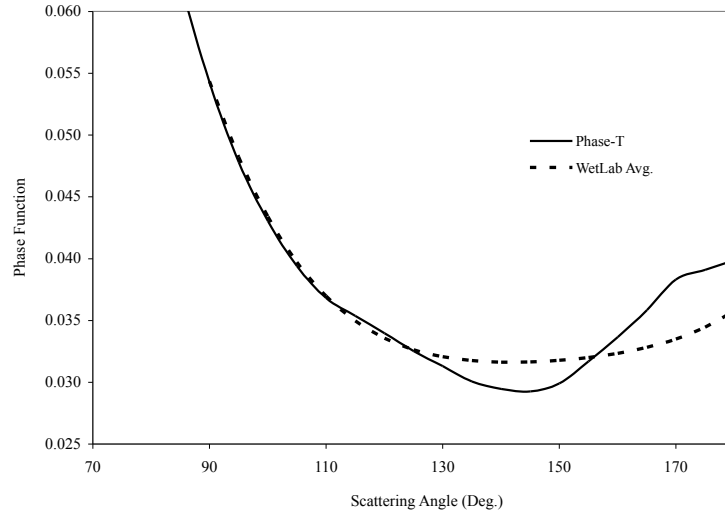


Figure 5.17: A more detailed comparison between “Phase-T” and “WetLab Avg.” Note that the phase function scale is now linear. The normalization at 90° is the same as in Figure 5.16.

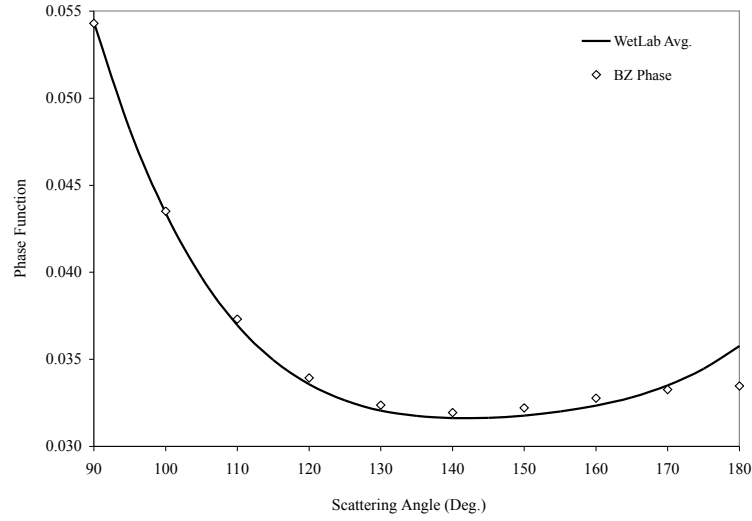


Figure 5.18: The fit of β_{BZ} to the “WetLab Avg.” phase function (extended with Phase-T to small angles to form the Petzmas phase function). The Petzmas phase function was sampled at $\Theta = 45^\circ$, 90° , and 135° to estimate $e_f = 0.71$ and $e_b = 0.34$. Note that the WetLab Avg. result at $\Theta = 180^\circ$ was not measured; measurements from $\Theta \leq 170^\circ$ were extrapolated to 180° .

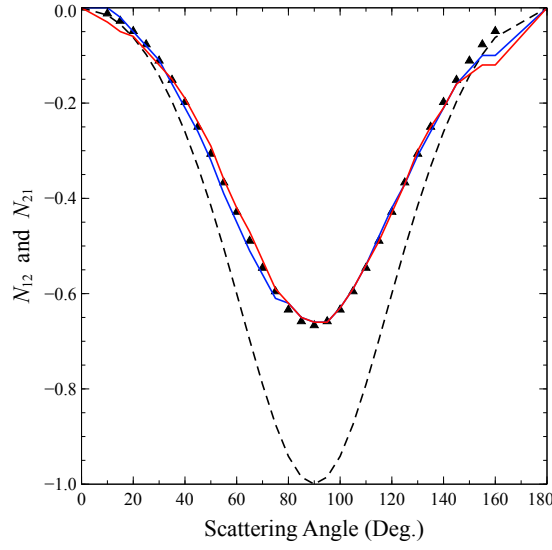


Figure 5.19: The elements N_{12} (red line) and N_{21} (blue line) of the average normalized scattering Mueller phase matrix for sixty stations in the Atlantic and Pacific Oceans. Dashed line is for isotropic Rayleigh (or Rayleigh-Gans) scattering, and triangles for anisotropic Rayleigh (or Rayleigh-Gans) scattering with $\delta = 0.2$.

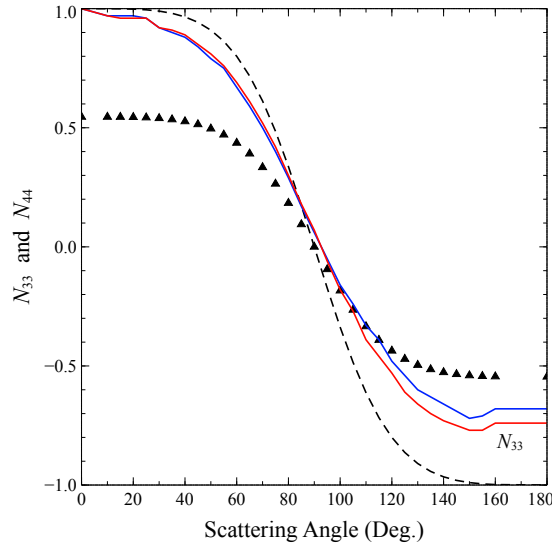


Figure 5.20: The elements N_{33} (red line) and N_{44} (blue line) of the average normalized scattering Mueller phase matrix for sixty stations in the Atlantic and Pacific Oceans. Dashed line is for isotropic Rayleigh (or Rayleigh-Gans) scattering, and triangles are N_{44} for anisotropic Rayleigh (or Rayleigh-Gans) scattering with $\delta = 0.2$.

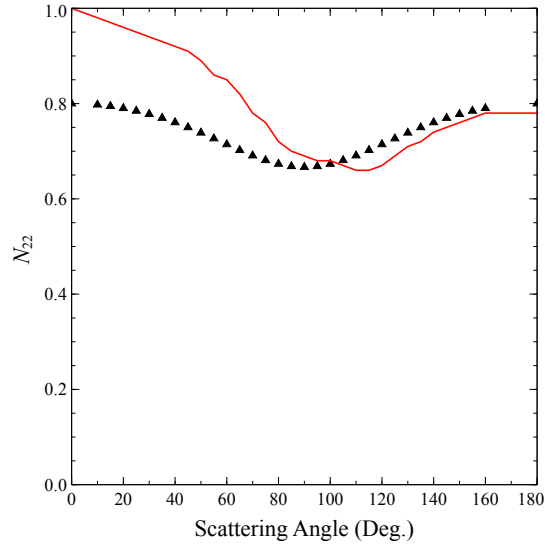


Figure 5.21: The element N_{22} (red line) of the average normalized scattering Mueller phase matrix for sixty stations in the Atlantic and Pacific Oceans. $N_{22} = 1$ for isotropic Rayleigh (or Rayleigh-Gans) scattering. The triangles are for anisotropic Rayleigh (or Rayleigh-Gans) scattering with $\delta = 0.2$.

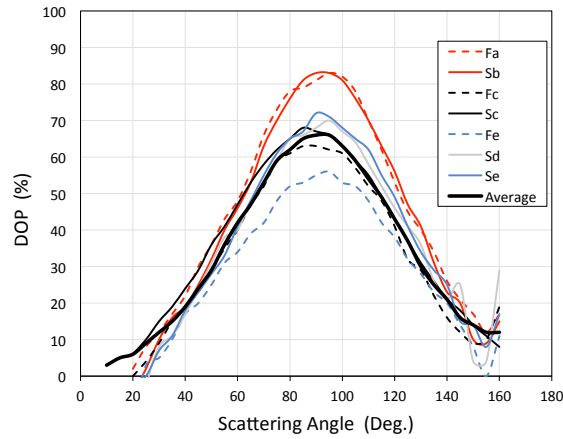


Figure 5.22: Degree of Polarization of scattering by marine phytoplankton compared to the oceanic average (“Average”). Those labeled with “S” are approximately spherical in shape. Those labeled with “F” have a filamentous structure, i.e., cylindrical in shape. The species are as follows: Fa, *Prochlorothrix hollandica*; Sb, *Microcystis aeruginosa*; Fc, *Oscillatoria agardhii*; Sc, *Microcystis* sp.; Fe, *Anabaena flos aquae*; Sd, *Phaeocystis*; Se, *Volvox aureus*. Note that the Degree of Polarization here is defined as $-N_{12}(\Theta) = -P_{12}(\Theta)/P_{11}(\Theta)$.

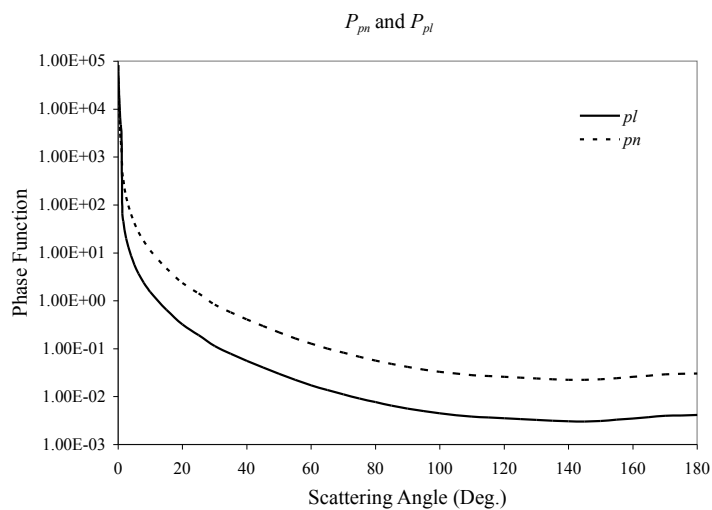


Figure 5.23: Proposed phase functions for the living and nonliving particulate matter in water based on modifying the Petzmas phase function according to Table 5.3. P_{pl} and P_{pn} have backscattering probabilities of 0.0019 and 0.0140, respectively.

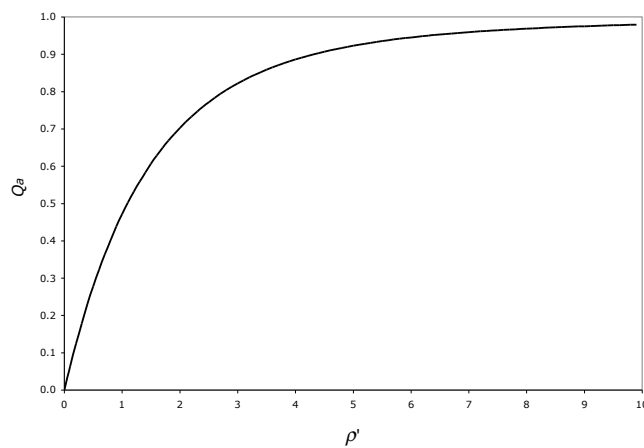


Figure 5.24: The absorption efficiency Q_a as a function of $\rho' = Da_i = Da_i^* c_i$.

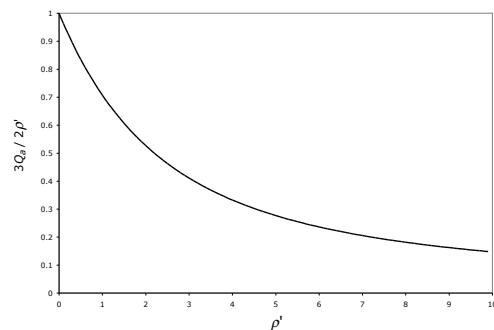


Figure 5.25: The function $3Q_a/2\rho' = a_{\text{Part}}^*/a_i^*$ as a function of $\rho' = Da_i = Da_i^* c_i$.

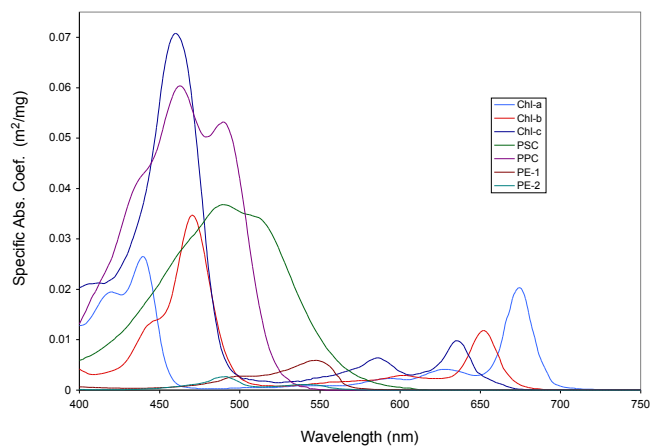


Figure 5.26: Specific (to the concentration of the given pigment) absorption spectra of some absorbing pigments found in phytoplankton. Chl-a: Chlorophyll *a*. Chl-b: Chlorophyll *b*. Chl-c: Chlorophyll *c*. PSC: Photosynthetic Carotinoids. PPC: Photoprotectant Carotinoids. PE-1: Phycoerythrin. PE-2: Phycoerythrin. These spectra are measured for pigments in solution (not in the cell). Within the cells, the pigment maxima shift a few (6-12 nm) to longer wavelengths.

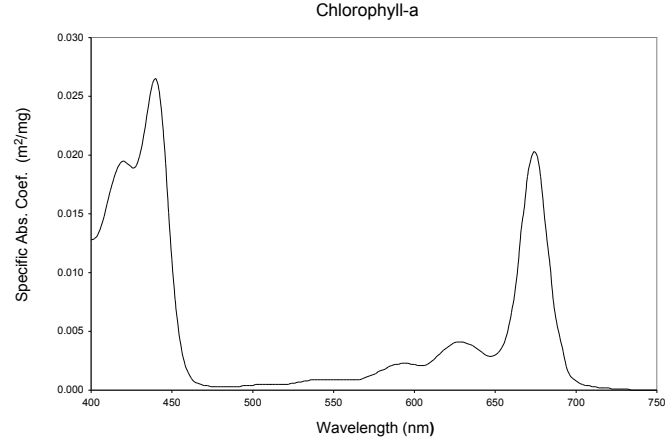


Figure 5.27: Specific absorption spectrum of Chlorophyll *a* from Figure 5.26.

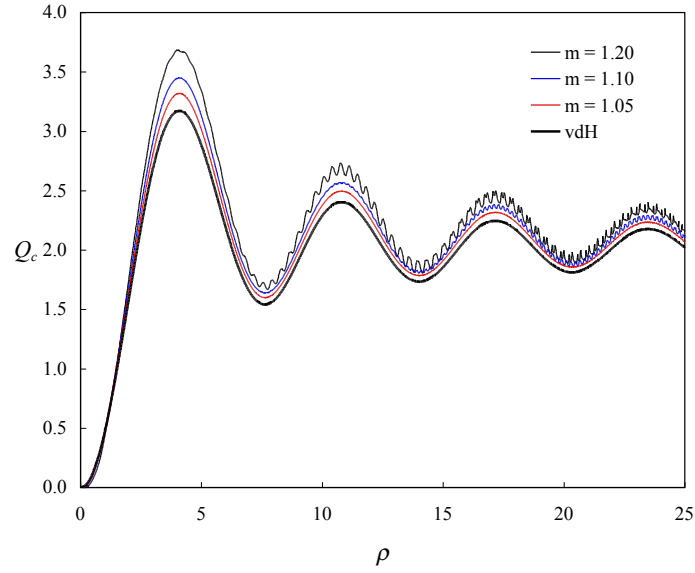


Figure 5.28: Extinction efficiency of non-absorbing spheres as a function of $\rho = 2x(\tilde{m}_r - 1)$. The top three curves are for $\tilde{m}_r = 1.05, 1.10$, and 1.20 . The lowest curve is the approximation Eq. (5.26). This is similar to Figure 3.3.

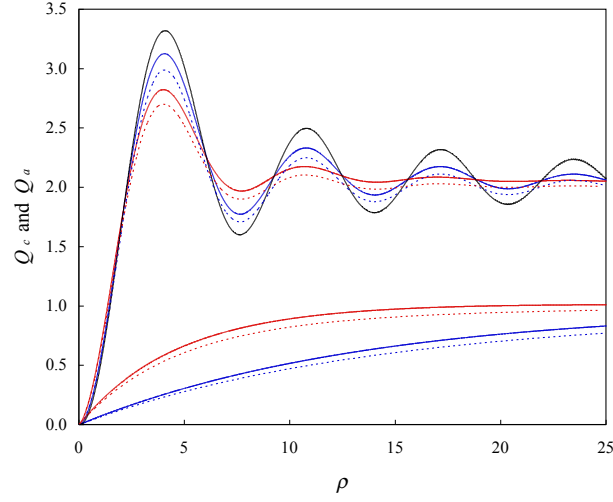


Figure 5.29: Extinction efficiency Q_c and absorption efficiency Q_a of absorbing spheres as a function of $\rho = 2x(\tilde{m}_r - 1)$. The solid curves are exact Mie computations for $\tilde{m}_r = 1.05$ (black), $\tilde{m}_r = 1.05 + 0.0025i$ (blue), and $\tilde{m}_r = 1.05 + 0.0075i$ (red). The associated dashed curves are the approximations to Q_c and Q_a , Eqs. (5.26) and (5.21), respectively.

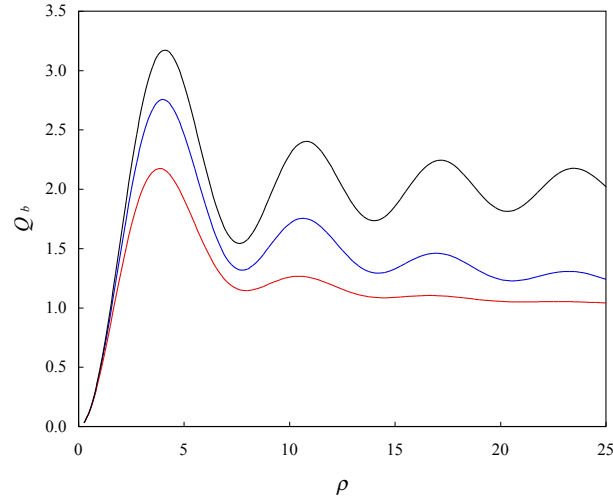


Figure 5.30: Scattering efficiency Q_b as a function of $\rho = 2x(\tilde{m}_r - 1)$. The curves are for $\tilde{m}_r = 1.05$ (black), $\tilde{m}_r = 1.05 + 0.0025i$ (blue), and $\tilde{m}_r = 1.05 + 0.0075i$ (red). They were calculated from $Q_c - Q_a$.

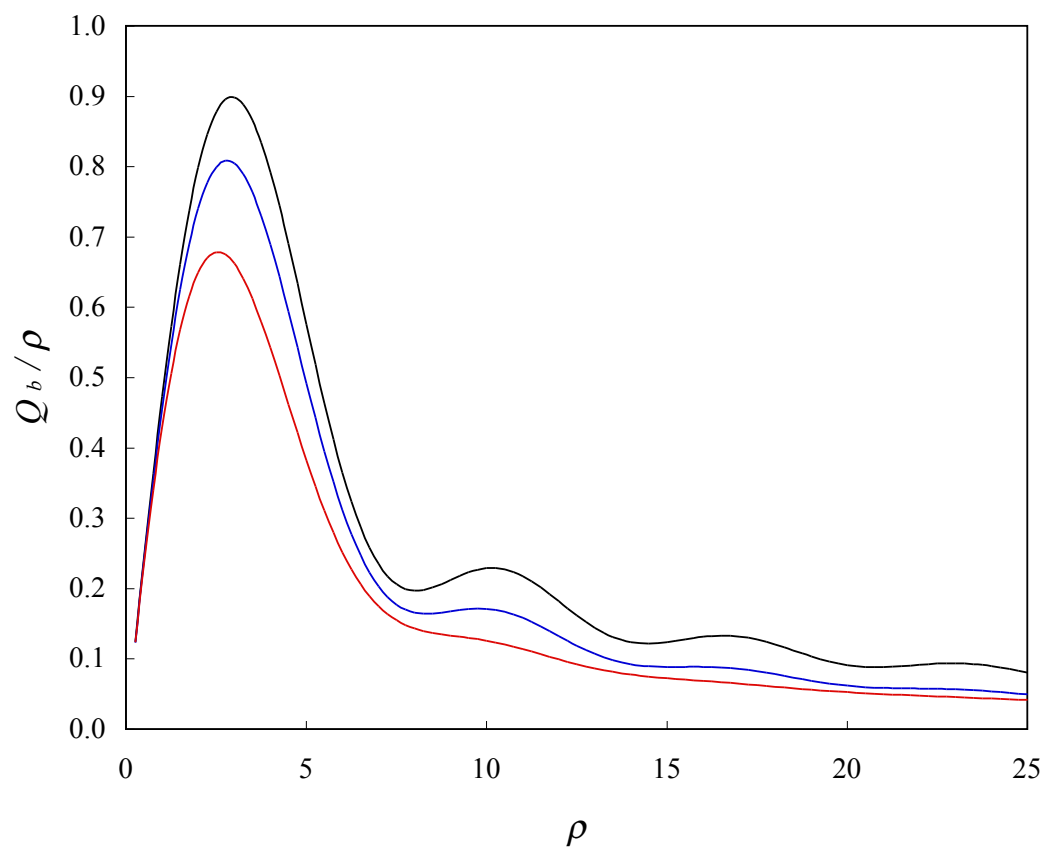


Figure 5.31: Scattering efficiency Q_b divided by ρ as a function of $\rho = 2x(\tilde{m}_r - 1)$. The curves are for $\tilde{m}_r = 1.05$ (black), $\tilde{m}_r = 1.05 + 0.0025i$ (blue), and $\tilde{m}_r = 1.05 + 0.0075i$ (red).

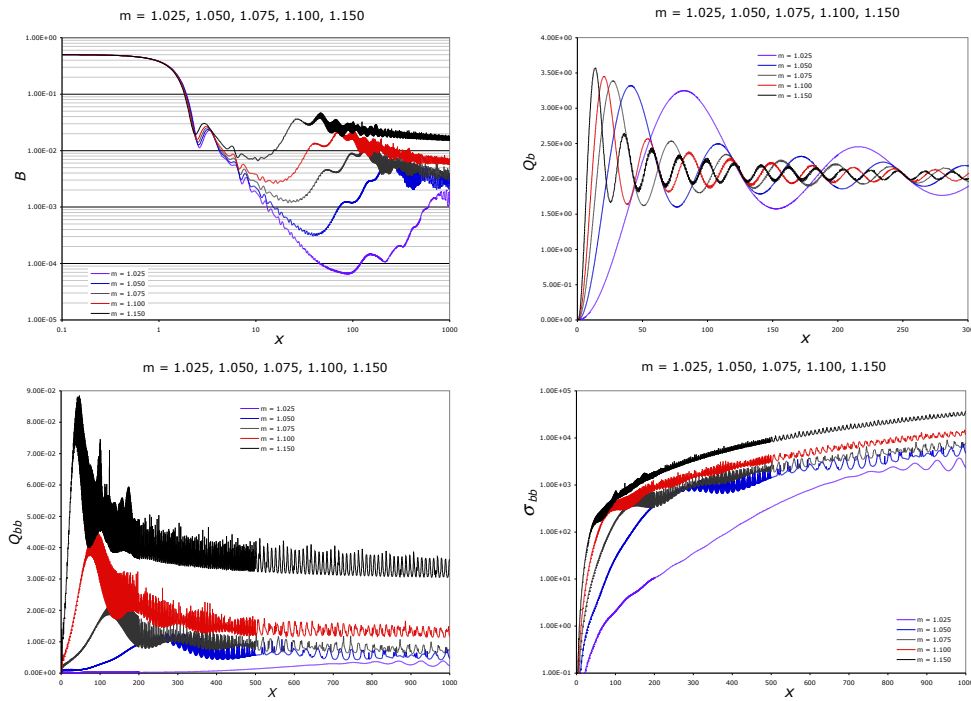


Figure 5.32: Optics of non-absorbing spherical particles as a function of $x = \pi D / \lambda_{\text{water}}$, where D is the diameter and λ_{water} is the wavelength. The parameter varied is the real part of the refractive index (m_r). Upper left: B . Upper right: Q_b . Lower left: Q_{bb} . Lower right: σ_{bb} .

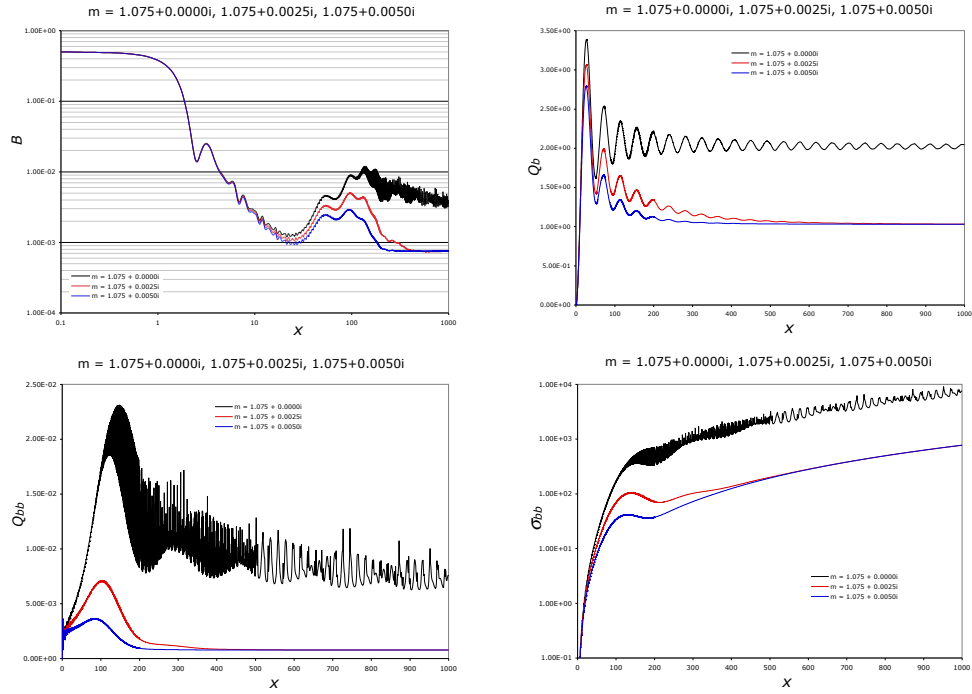


Figure 5.33: Optics of absorbing spherical particles as a function of $x = \pi D / \lambda_{\text{water}}$, where D is the diameter and λ_{water} is the wavelength. The parameter varied is the imaginary part of the refractive index (m_i , i.e., the absorption). Upper left: B . Upper right: Q_b . Lower left: Q_{bb} . Lower right: σ_{bb} .

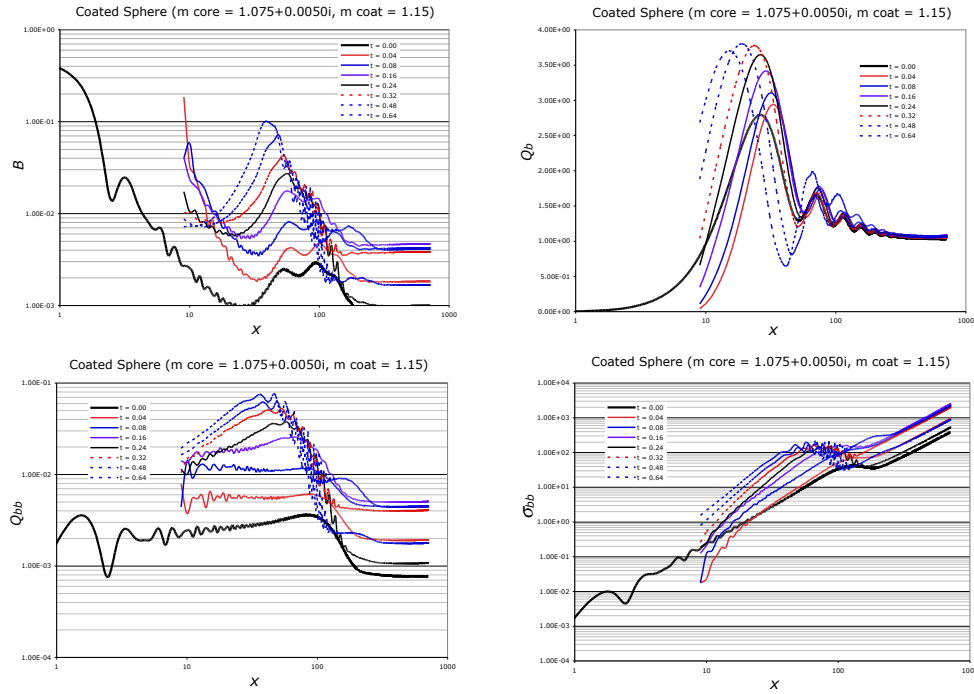


Figure 5.34: Optics of a coated sphere with refractive index $m = 1.075 + 0.0050i$. The refractive index of the coating is 1.15 and the thickness of the coating t is the parameter varied in the figures ($t = 0, 0.04, 0.08, 0.16, 0.24, 0.32, 0.48, 0.64 \mu\text{m}$). Upper left: B . Upper right: Q_b . Lower left: Q_{bb} . Lower right: σ_{bb} . The variable x is $\pi D/\lambda$, where D is the diameter of the core *plus* the coating.

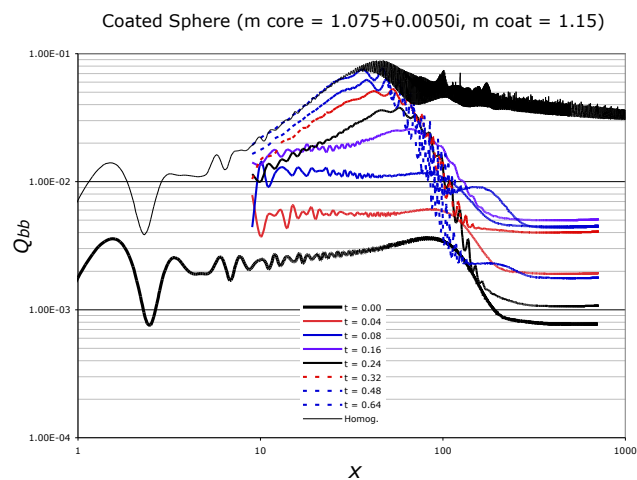


Figure 5.35: Comparison between backscattering scattering efficiency of a coated sphere with successively thicker coatings and a homogeneous sphere with the same refractive index as the coating (1.15).

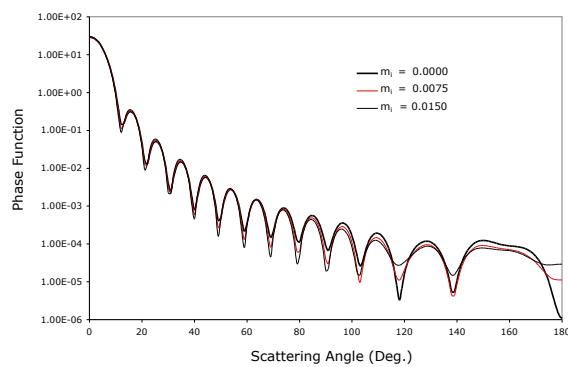


Figure 5.36: Phase function for a sphere with $x = 20$, $\tilde{m}_r = 1.05$ and $\tilde{m}_i = 0$, 0.0075, and 0.0150.

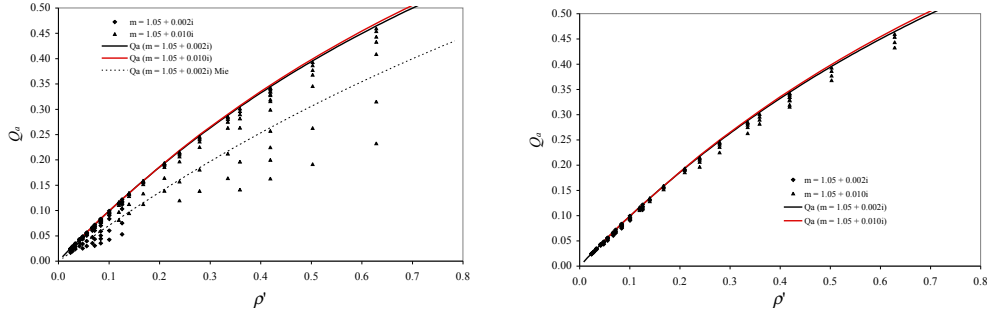


Figure 5.37: The absorption efficiency Q_a for randomly oriented absorbing cylinders ($m = 1.05 + 0.002i$ and $1.05 + 0.010i$) as a function of $\rho' = Da_i = Da_i^* c_i$, for aspect ratios ($AR = L/D$) ranging from 0.333 to 30 (left panel). The solid lines correspond to Q_a for an infinitely long cylinder. The cylinder diameters are 0.5, 1.0, and 1.5 μm , and the vacuum wavelength ranges from 400 to 700 nm. The dotted line is Mie theory for spheres with index $m = 1.05 + 0.002i$. In the right panel only computations for $AR \geq 3$ are shown.

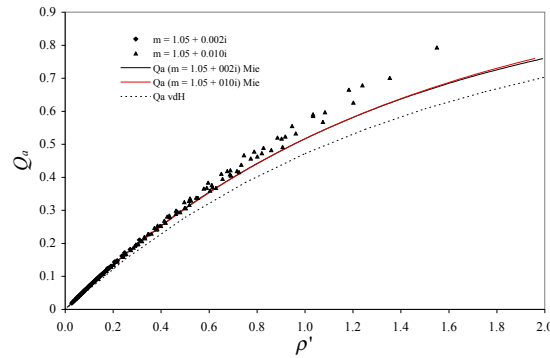


Figure 5.38: The absorption efficiencies of cylinders computed by dividing the cross sections (computed for cylinders) by the projected area of the *volume-equivalent sphere* as a function of $\rho' = 4x\tilde{m}_i$, where $x = \pi D_{\text{Equiv}} m_{\text{Water}} / \lambda$ and D_{Equiv} is the diameter of the *volume-equivalent sphere*. All the aspect ratios in Figure 5.37 are included.

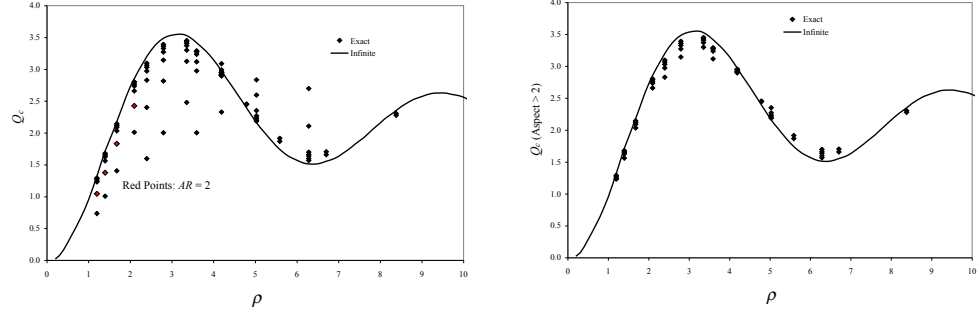


Figure 5.39: The extinction efficiency Q_c for randomly oriented non-absorbing cylinders ($\tilde{m} = 1.20$), as a function of $\rho = 2x(\tilde{m}_r - 1)$, where $x = \pi D m_{\text{Water}} / \lambda$ and D is the diameter of the cylinder, for aspect ratios ($AR = L/D$) ranging from 0.333 to 30 (left panel). The cylinder diameters are 0.5, 1.0, and 1.5 μm , and the vacuum wavelengths range from 400 to 700 nm. The solid line is Q_c for infinitely long cylinders. In the right panel only computations for $AR \geq 3$ are shown. From Gordon [2011].

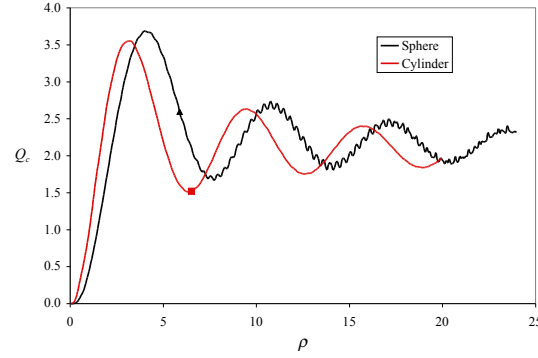


Figure 5.40: Comparison between Q_c for spheres and randomly oriented, infinitely long, cylinders. These curves are exact results for $\tilde{m} = 1.20$.

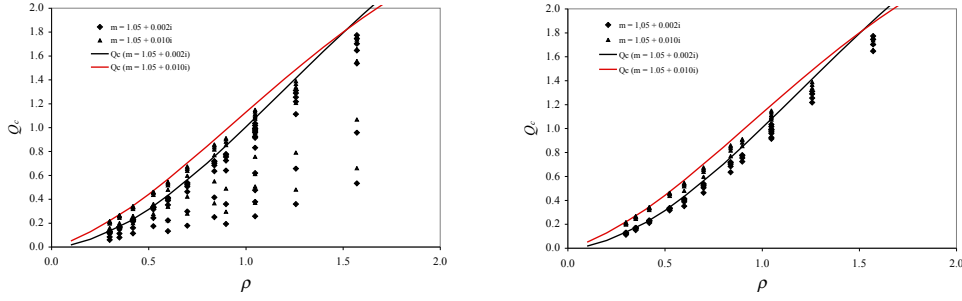


Figure 5.41: The extinction efficiency Q_c for randomly oriented absorbing cylinders ($m = 1.05 + 0.002i$ and $1.05 + 0.010i$), as a function of $\rho = 2x(\tilde{m}_r - 1)$, where $x = \pi D m_{\text{Water}} / \lambda$ and D is the diameter of the cylinder, for aspect ratios ($AR = L/D$) ranging from 0.333 to 30 (left panel). The cylinder diameters are 0.5, 1.0, and 1.5 μm , and the vacuum wavelengths range from 400 to 700 nm. The solid lines are Q_c for infinitely long cylinders. In the right panel only computations for $AR \geq 3$ are shown. From [Gordon \[2011\]](#).

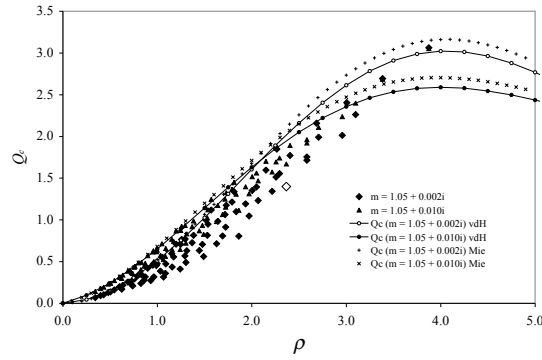


Figure 5.42: The extinction efficiencies of cylinders computed by dividing the cross sections (computed for cylinders) by the projected area of the *volume-equivalent sphere* as a function of $\rho = 2x(\tilde{m}_r - 1)$, where $x = \pi D_{\text{Equiv}} m_{\text{Water}} / \lambda$ and D_{Equiv} is the diameter of the *volume-equivalent sphere*. All the aspect ratios in [Figure 5.37](#) are included. From [Gordon \[2011\]](#).

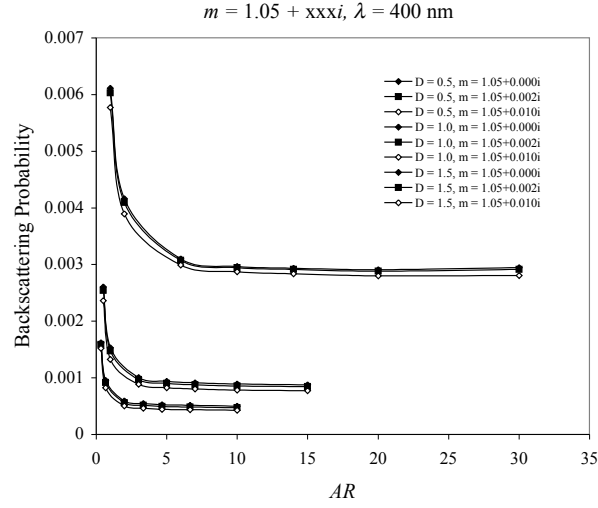


Figure 5.43: Examples of the variation of the backscattering probability with aspect ratio for cylinder diameters between 0.5 and 1.5 μm and refractive indices $\tilde{m} = 1.05$, $1.05 + 0.002i$ and $1.05 + 0.010i$ to demonstrate the effect of absorption on backscattering by finite cylinders. The vacuum wavelength is 400 nm. Note that the largest value of the size parameter here is $x \approx 16$.

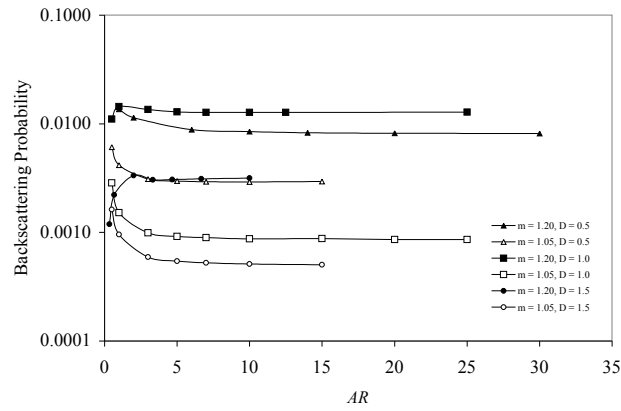


Figure 5.44: Examples of the variation of the backscattering probability with aspect ratio for cylinder diameters between 0.5 and 1.5 μm and refractive indices $\tilde{m} = 1.05$ and 1.20 to demonstrate the influence of refractive index on backscattering of finite cylinders. The vacuum wavelength is 400 nm.

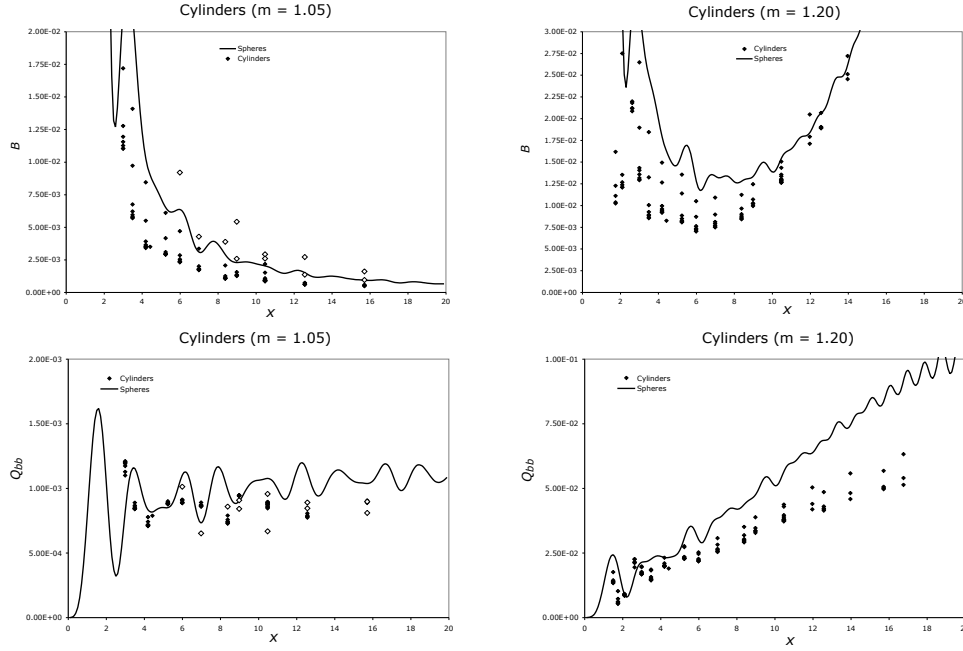


Figure 5.45: Comparison of backscattering probability and backscattering efficiency between cylinders (data points) and spheres (continuous line) as a function of the associated particle size parameter. Upper left: B for $m = 1.05$. Upper right: B for $m = 1.20$. Lower left: Q_{bb} for $m = 1.05$. Lower right: Q_{bb} for $m = 1.20$. For spheres the variable x is as usual $\pi D/\lambda_{\text{Water}}$, where D is the diameter of the sphere. For cylinders the variable x is $\pi D/\lambda_{\text{Water}}$, where D is the diameter of the cylinder. The open symbols for $m = 1.05$ correspond to cases in which the aspect ratio of the cylinder is less than one, i.e., a disk-shaped particle.

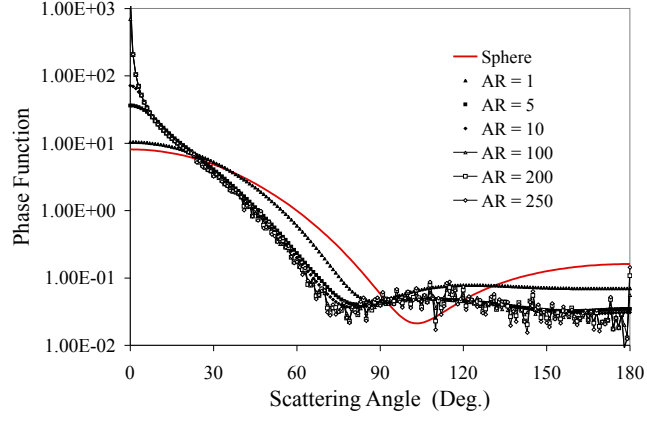


Figure 5.46: Phase function for a non-absorbing, randomly oriented, cylinder with $x = 2.62$, $\tilde{m}_r = 1.20$, $\tilde{m}_i = 0$ and various aspect ratios (AR) compared to a sphere with the same diameter.

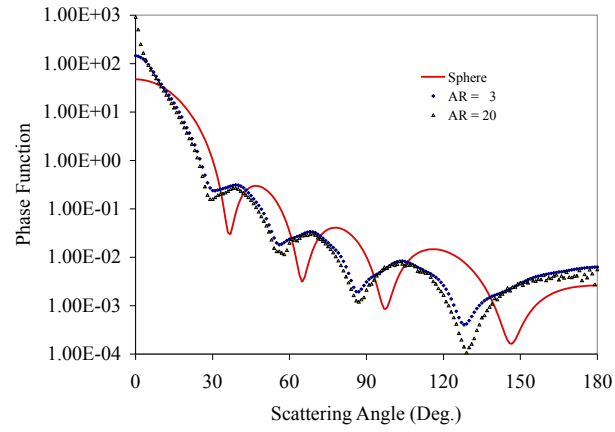


Figure 5.47: Phase function for an absorbing, randomly oriented, cylinder with $x = 7$, $\tilde{m}_r = 1.05$ and $\tilde{m}_i = 0.0020$.

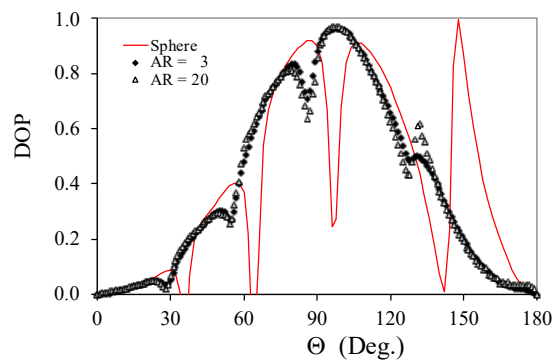


Figure 5.48: Degree of Linear Polarization ($-P_{12}(\Theta)/P_{11}(\theta)$) for an absorbing, randomly oriented, cylinder with $x = 7$, $\tilde{m}_r = 1.05$ and $\tilde{m}_i = 0.0020$.

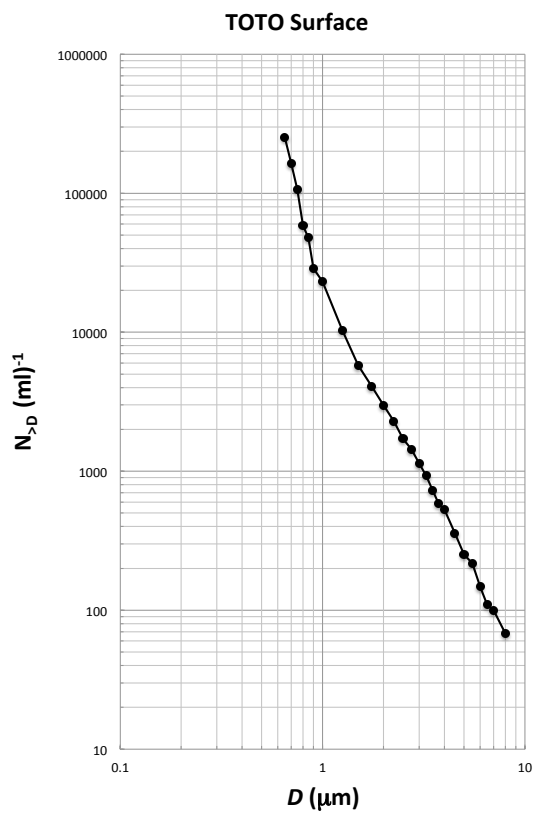


Figure 5.49: Size distribution of particles measured in the surface waters of the Tongue of the Ocean, Bahamas using a Coulter counter. Here, $N_{>D}$ refers to the number of particles per ml with sizes *greater* than the indicated diameter (D).

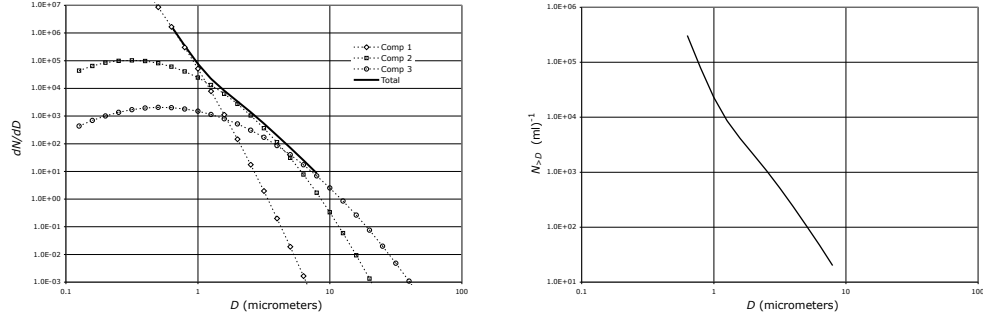


Figure 5.50: An exercise showing how three log normal distributions can be combined to produce a size distribution that could be fit reasonably well by segmented power law distributions. The right panel is the cumulative distribution resulting from the sum of the three log-normals in the left panel.

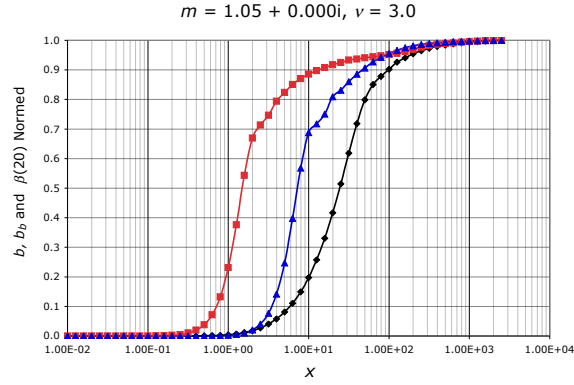


Figure 5.51: Scattering quantities as a function of the size parameter x used as the upper limit on the integrals, e.g., in Eq. (5.31) with $x_{\text{Min}} = 0$ and x_{Max} ranging from 0 to 2500 with $\tilde{m} = 1.050 + 0.0000i$. Diamonds (black): b . Squares (red): b_0 . Triangles (blue): $\beta(20^\circ)$.

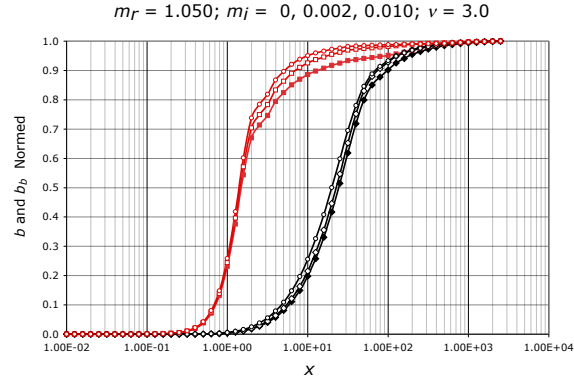


Figure 5.52: Same as Figure 5.51, but here $\tilde{m}_r = 1.050$ and $\tilde{m}_i = 0, 0.002$, and 0.010 . Black curves (on right) are for b , with $\tilde{m}_i = 0$ (Solid Diamonds), $\tilde{m}_i = 0.002$ (Open Diamonds), $\tilde{m}_i = 0.010$ (Open Circles). Red curves (on left) are for b_b , with $\tilde{m}_i = 0$ (Solid Squares), $\tilde{m}_i = 0.002$ (Open Squares), $\tilde{m}_i = 0.010$ (Open Circles).

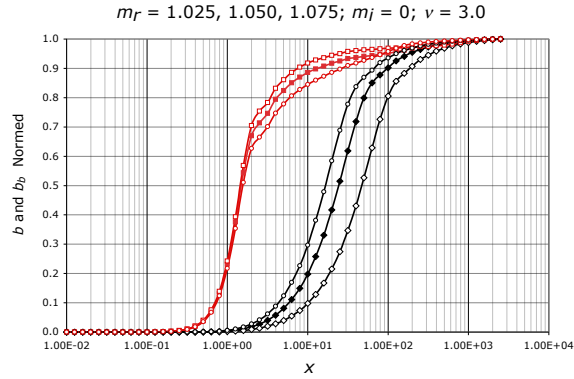


Figure 5.53: Same as Figure 5.51, but here $\tilde{m}_i = 0$ and $\tilde{m}_r = 1.025, 1.050$, and 1.075 . Black curves (on right) are for b , with $\tilde{m}_r = 1.025$ (Open Diamonds), $\tilde{m}_r = 1.050$ (Solid Diamonds), $\tilde{m}_r = 1.075$ (Open Circles). Red curves (on left) are for b_b , with $\tilde{m}_r = 1.025$ (Open Squares), $\tilde{m}_r = 1.050$ (Solid Squares), $\tilde{m}_r = 1.075$ (Open Circles).

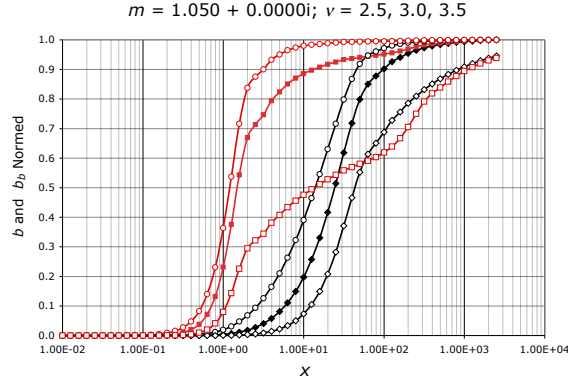


Figure 5.54: Same as Figure 5.51, but here $\tilde{m} = 1.050 + 0.0000i$ and $\nu = 2.5, 3.0$, and 3.5 . Black curves (on right) are for b , with $\nu = 2.5$ (Open Diamonds), $\nu = 3.0$ (Solid Diamonds), $\nu = 3.5$ (Open Circles). Red curves (on left) are for b_b , with $\nu = 2.5$ (Open Squares), $\nu = 3.0$ (Solid Squares), $\nu = 3.5$ (Open Circles).

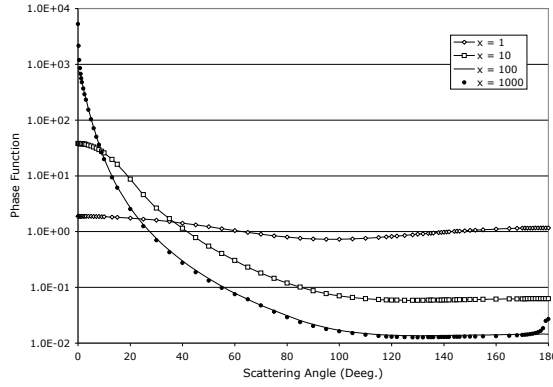


Figure 5.55: The phase function for particle scattering following a power-law distribution with $\nu = 3$. The size distribution extends from $0 \rightarrow x_{\text{Max}}$ (x_{Max} is indicated by x in the legend box). Note that there is an insignificant difference between $x_{\text{Max}} = 100$ and $x_{\text{Max}} = 1000$.

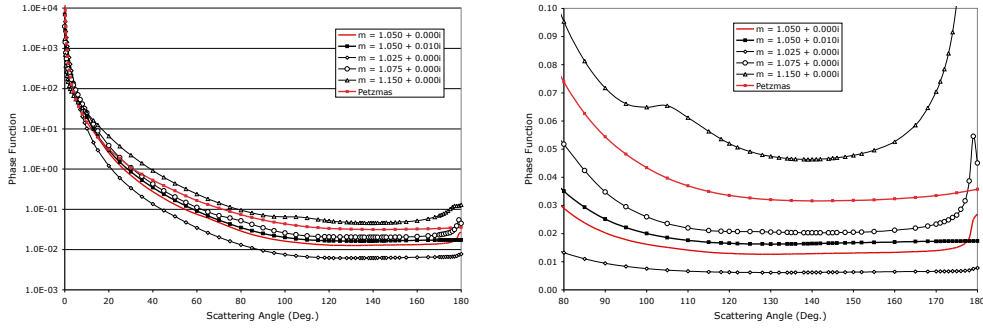


Figure 5.56: The phase function for particle scattering following a power-law distribution, with $\nu = 3$, as a function of \tilde{m} (\tilde{m} is indicated by m in the legend). The size distribution extends from $0 \rightarrow x_{\text{Max}} = 1000$. “Petzmas” is the modified Phase-T described in Section 5.3.3.2.

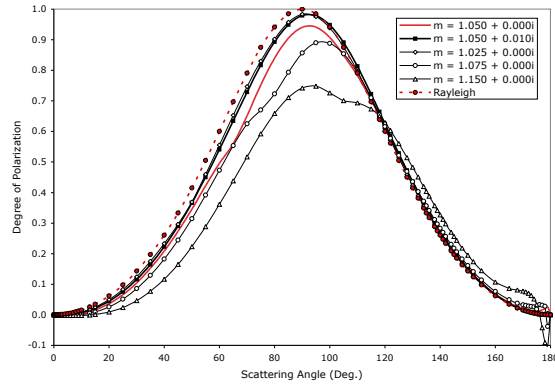


Figure 5.57: The degree of polarization (DOP) for particle scattering following a power-law distribution, with $\nu = 3$, as a function of \tilde{m} (\tilde{m} is indicated by m in the legend). The size distribution extends from $0 \rightarrow x_{\text{Max}} = 1000$. “Rayleigh” is the DOP of isotropic Rayleigh scattering.

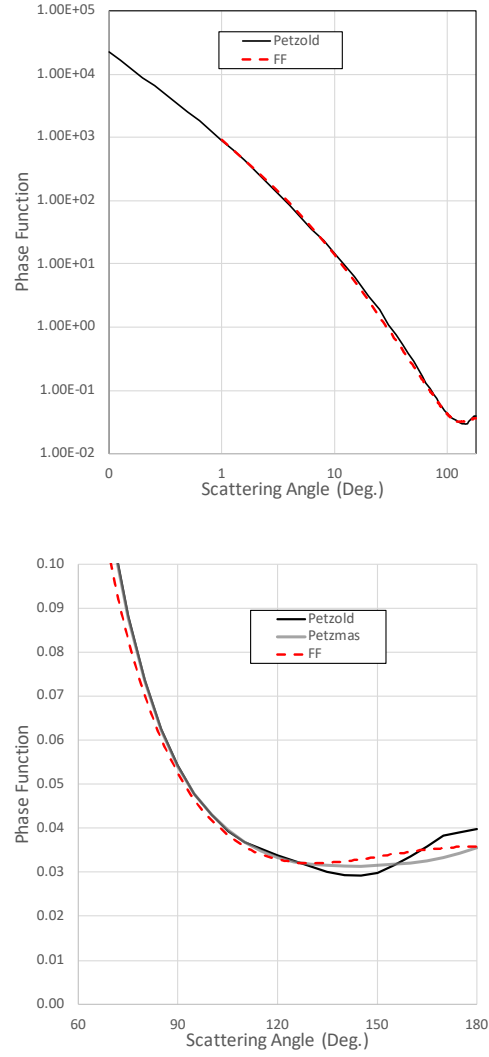


Figure 5.58: Comparison between the Petzold (Phase-T) and the FF phase functions (upper panel). The parameters used in generating the FF phase function were $m = 1.10$ and $\nu = 2.5835$. Lower panel is an expanded (linear) version of the left panel, with the Petzmas phase functions included as well. All of the phase functions have the same backscattering probability.

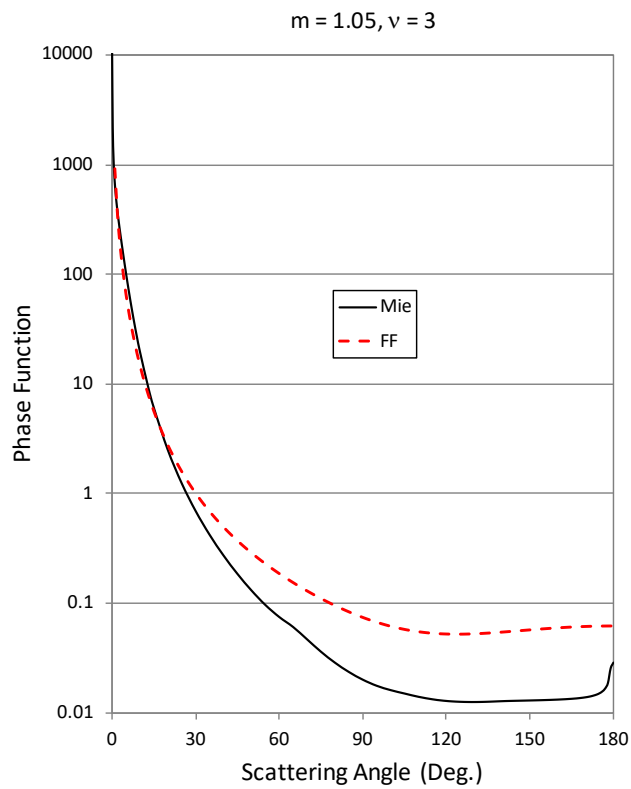


Figure 5.59: Comparison between the Mie-computed (for spheres) and the associated FF phase functions for $\tilde{m} = 1.05$ and $\nu = 3$. The associated backscattering probabilities are 0.00713 and 0.02885 for the Mie and FF calculations, respectively.

Chapter 6

Apparent Optical Properties (AOPs) of Natural Waters

6.1 Introduction

In this Chapter we discuss the *apparent optical properties* (AOPs) of natural waters. The AOPs are not really optical properties of the water and constituents, i.e., a characteristic of the medium like the IOPs (Chapter 5) but, under certain conditions, they may appear to be. In an ocean with spatially homogeneous IOPs, and a given illumination at the surface by the sun and sky, the AOPs are relatively weak functions of depth, and therefore appear to be properties of the water body. This the source of the word “apparent.” Unfortunately, when the illumination conditions change, the values of the AOPs change, so, as mentioned above, they are not characteristic of the medium, but depend on the illumination conditions as well. Their importance stems from the fact that historically, compared to the IOPs, the AOPs were much easier to measure. In addition, two of the AOPs are critically important in ocean color remote sensing: the remote sensing ratio and the diffuse attenuation coefficient. The former directly relates to the water-leaving radiance measured by a remote sensor above the surface, and the latter relates to the depth over which the remote sensor can “see” into the water.

Here we describe all of the AOPs that are important to oceanic optics with a particular emphasis on remote sensing. The goal is to familiarize the reader with the AOPs and their variation with the constituent concentrations, and to prepare the reader for understanding remote sensing applications through the derivation of approximate relationships that provide the explicit dependence of the AOPs on the IOPs, and in turn, on the constituent

concentrations through their influence on the IOPs.

We begin with definitions of most of the AOPs. Next we provide some examples of several AOPs derived from field measurements. In addition, we provide a set of examples of AOPs computed from solutions of the radiative transfer equation using IOPs that are representative of those actually observed in natural waters. The latter examples are more complete than the former in that a wider range of AOPs can be examined, and their variability with depth more clearly demonstrated. This examination is followed by a lengthy section developing a simple model for treating multiple scattering in radiative transfer (the quasi-single scattering approximation – QSSA). The QSSA provides approximate relationships for the AOPs in terms of the IOPs. These relationships are studied to determine their range of validity in a natural environmental setting, possible avenues for improvement, and their general application to ocean color remote sensing. Inelastic processes and the effects of vertical variations in the IOPs are then examined and incorporated into the QSSA. In an appendix the problem of estimating the IOPs given measurements of the AOPs is considered in some detail.

6.2 Apparent Optical Properties

6.2.1 Basic Radiometric Quantities

The radiometric quantity that is of fundamental importance in the water is the radiance, L (Chapter 2). Because of the strong attenuation of light in the water, the *attenuation length*, $\ell = 1/c$, is at most of the order of 100 m; however, horizontal spatial variability of the inherent optical properties in oceanic waters is usually at scales much greater than 100 m.¹ Picking a coordinate system in which the z -axis is increasing into the water from $z = 0$ at the surface and the x -axis and y -axis are parallel to the surface, this strong attenuation means that

$$\frac{\partial L}{\partial x} \ll \frac{\partial L}{\partial z} \quad \text{and} \quad \frac{\partial L}{\partial y} \ll \frac{\partial L}{\partial z} \quad \text{so} \quad \hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) \approx (\hat{\xi} \bullet \hat{e}_z) \frac{\partial L}{\partial z}.$$

If it is assumed that the IOPs are dependent only on depth and that the incident illumination from the Sun and the sky is independent of x and y , then the radiance $L(\vec{r}, \hat{\xi}, \lambda)$

¹In coastal and inland waters the scale of variability can be on the order of 100 m or even less, but in such waters c is typically much larger than for oceanic waters, so $\ell \ll 100$ m. Usually the assumption, that vertical variability in the radiance and IOPs is much greater than horizontal variability in such waters, is still valid.

becomes $L(z, \hat{\xi}, \lambda)$ and

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0 \quad \text{and} \quad \hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) = (\hat{\xi} \bullet \hat{e}_z) \frac{\partial L}{\partial z}.$$

Actually, it can almost always be assumed that the IOPs are dependent only on depth, so these are excellent approximations. Therefore, the radiative transfer equation becomes one-dimensional, and in this setting the water body is, from a radiative transfer theory view point, a plane-parallel medium.

In Chapter 2 we discussed the various integrals of the radiance $L(z, \hat{\xi}, \lambda)$ propagating in the direction $\hat{\xi}$, in plane-parallel media. The *downwelling irradiance* $E_d(z, \lambda)$ and the *upwelling irradiance* $E_u(z, \lambda)$ were defined according to

$$E_d(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z > 0} |\hat{\xi} \bullet \hat{e}_z| L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (6.1)$$

and

$$E_u(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z < 0} |\hat{\xi} \bullet \hat{e}_z| L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (6.2)$$

respectively. These can be combined to form the *vector irradiance* $E_d - E_u$:²

$$E_d(z, \lambda) - E_u(z, \lambda) = \int_{\text{All } \hat{\xi}} \hat{\xi} \bullet \hat{e}_z L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}). \quad (6.3)$$

In addition, the downwelling (E_{0d}) and upwelling (E_{0u}) and total (E_0) *scalar irradiances* are defined as above but without the $\hat{\xi} \bullet \hat{e}_z$ factor:

$$E_{0d}(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z > 0} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (6.4)$$

$$E_{0u}(z, \lambda) \triangleq \int_{\hat{\xi} \bullet \hat{e}_z < 0} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}), \quad (6.5)$$

²The reader may wonder why $E_d - E_u$ is called the *vector irradiance*. A more general definition of the vector irradiance $\vec{E}_V(\vec{r})$ is

$$\vec{E}_V(\vec{r}) \triangleq \int_{\text{All } \hat{\xi}} \hat{\xi} L(\vec{r}, \hat{\xi}) d\Omega(\hat{\xi}).$$

The vector \vec{E}_V usually points in (or close to) the direction of propagation of the maximum value of the radiance at the point \vec{r} . Equation (6.3) is the z component of the vector irradiance, i.e.,

$$E_d(\vec{r}) - E_u(\vec{r}) = \hat{e}_z \bullet \vec{E}_V(\vec{r}).$$

It is a simple matter to start with the radiative transfer equation and derive a generalized Gershun's equation given by

$$\nabla \bullet \vec{E}_V(\vec{r}) = -a(\vec{r})E_0(\vec{r}) + \int_{\text{All } \hat{\xi}} Q(\vec{r}, \hat{\xi}) d\Omega(\hat{\xi}).$$

$$E_0(z, \lambda) \triangleq E_{0d}(z, \lambda) + E_{0u}(z, \lambda) = \int_{\text{All } \hat{\xi}} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi}). \quad (6.6)$$

The irradiances E_d and E_u are the energy fluxes (in the interval $\Delta\lambda$ and divided by $\Delta\lambda$) across a horizontal surface at z in the downward and upward directions, respectively. The vector irradiance is the *net* downward flux. The scalar irradiance was shown in Chapter 2 to be proportional to the energy density in the electromagnetic field, also in the interval $\Delta\lambda$ and divided by $\Delta\lambda$.

These radiometric quantities — the radiance and the irradiances derived from it — within a water body are strongly dependent on the the illumination from the Sun and sky. Consider an extreme case of illumination variation: the direct Sun is blocked by a passing cloud causing the downward irradiance at the water's surface and within the water to fall by an order of magnitude or more. When the value of a radiometric quantity can change by an order of magnitude within a few seconds, it is impossible that a given value of the radiometric quantity within the water can be of any use in a description of the water body, or of conditions within the water body. However, when a change in illumination such as described above occurs, it occurs at all depths in the medium (nearly) simultaneously. Thus, the ratio of various radiometric quantities (e.g., $E_d(z_2)/E_d(z_1)$, where z_1 and z_2 are two different depths within the medium, or $E_u(z)/E_d(z)$) remains unchanged (when they are measured simultaneously, and except for any change caused by differences in the angular distribution of the radiance). It would then seem that such ratios might be useful in the description of the water body, particularly if they depend only weakly on depth within the water. All of the apparent optical properties are related to the ratios of various radiometric quantities at a given depth, or at two different, depths. They are defined in the following section.³

6.2.2 Apparent Optical Properties: Definitions

We shall see in the examples provided below that E_d decreases in an almost exponential manner with increasing depth, so it is reasonable to determine the exponential decay coefficient. The decay coefficients of $E_d(z, \lambda)$, $E_u(z, \lambda)$ and $E_0(z, \lambda)$ are called the *irradiance attenuation coefficients* and are defined through

$$K_x(z, \lambda) \triangleq -\frac{1}{E_x(z, \lambda)} \frac{dE_x(z, \lambda)}{dz} = -\frac{d\ln[E_x(z, \lambda)]}{dz},$$

where $x = d, u$, or 0 . The downward irradiance attenuation coefficient $K_d(z, \lambda)$ is often referred to as the *diffuse attenuation coefficient*.

³A further requirement for an AOP to be useful in characterizing a water body is that it be relatively stable. For example, at a given time each day, e.g., noon, an AOP in the open ocean will be more-or-less constant from day to day and could be used to characterize a particular oceanic region.

Another AOP formed from these irradiances is the *irradiance reflectance* or *irradiance ratio*:

$$R(z) \triangleq \frac{E_u(z, \lambda)}{E_d(z, \lambda)}.$$

Most modern instrumentation for measuring aspects of the in-water light field do not measure E_u (although some do), but measure the radiance propagating upward in the water, i.e., in the direction $-\hat{e}_z$. This is called the upwelling radiance $L_u(z, \lambda)$. The radiance exiting the water body propagating toward the zenith is proportional to $L_u(0, \lambda)$, and is the signal measured by a remote sensor aimed toward the nadir. In the spirit of the definitions above, we define

$$K_L(z, \lambda) \triangleq -\frac{1}{L_u(z, \lambda)} \frac{dL_u(z, \lambda)}{dz} = -\frac{d\ln[L_u(z, \lambda)]}{dz},$$

and

$$R_L(z) \triangleq \frac{L_u(z, \lambda)}{E_d(z, \lambda)}.$$

The quantity $R_L(0)$ is often referred to as the “*remote sensing ratio*” (*RSR*). It is *the most important AOP for understanding and interpreting remotely-sensed radiance exiting the water surface*. If the upwelling radiance distribution were uniform (independent of $\hat{\xi}$) then $E_u(z, \lambda) = \pi L_u(z, \lambda)$ and $R(z, \lambda) = \pi R_L(z, \lambda)$. A quantity that is a measure of the deviation of the upwelling light field from uniformity is called the *Q-factor*, defined according to

$$Q(z, \lambda) \triangleq \frac{E_u(z, \lambda)}{L_u(z, \lambda)},$$

so $R_L(z, \lambda) = R(z, \lambda)/Q(z, \lambda)$. The *remote sensing reflectance* is an additional quantity often used in remote sensing and is defined according to

$$R_{rs}(\lambda) = \frac{L_u(0, \lambda)}{E_d^+(0, \lambda)} t_L,$$

where t_L is the transmittance of $L_u(0, \lambda)$ through the air-water interface (water to air) and $E_d^+(0, \lambda)$ is the downwelling irradiance incident just *above* the water surface.⁴ For a

⁴If we write the upwelling radiance just below the water surface propagating in the direction $\hat{\xi}$ as $L_u(0, \hat{\xi}, \lambda)$, then we can extend the definition of the remote sensing reflectance to that for radiance exiting the water propagating in a direction $\hat{\xi}'$ above the surface, i.e.,

$$R_{rs}(\hat{\xi}', \lambda) = \frac{L_u(0, \hat{\xi}, \lambda)}{E_d^+(0, \lambda)} t_L(\hat{\xi} \rightarrow \hat{\xi}'),$$

where $\hat{\xi}$ and $\hat{\xi}'$ are related by Snell's law. The quantity $L_u(0, \hat{\xi}, \lambda) t_L(\hat{\xi} \rightarrow \hat{\xi}')$ is called the *water-leaving radiance* and is usually written $L_w(\hat{\xi}', \lambda)$. It is the radiance backscattered *out* of the water.

flat interface, the transmittance t_L is just the Fresnel transmittance for normal incidence divided by m_w^2 (Chapter 2).

Finally, a rough characterization of the directionality of the in-water light field is the *average cosine* ($\langle\mu\rangle$) of the radiance:

$$\langle\mu\rangle \triangleq \langle\hat{\xi} \bullet \hat{e}_z\rangle \triangleq \frac{\int_{\text{All } \hat{\xi}} (\hat{\xi} \bullet \hat{e}_z) L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi})}{\int_{\text{All } \hat{\xi}} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi})} = \frac{E_d(z, \lambda) - E_u(z, \lambda)}{E_0(z, \lambda)}.$$

As above for irradiance, the average cosines of the downwelling and upwelling light fields are

$$\langle\mu_d\rangle \triangleq \frac{\int_{\hat{\xi} \bullet \hat{e}_z > 0} (\hat{\xi} \bullet \hat{e}_z) L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi})}{\int_{\hat{\xi} \bullet \hat{e}_z > 0} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi})} = \frac{E_d(z, \lambda)}{E_{0d}(z, \lambda)},$$

and

$$\langle\mu_u\rangle \triangleq -\frac{\int_{\hat{\xi} \bullet \hat{e}_z < 0} (\hat{\xi} \bullet \hat{e}_z) L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi})}{\int_{\hat{\xi} \bullet \hat{e}_z < 0} L(z, \hat{\xi}, \lambda) d\Omega(\hat{\xi})} = \frac{E_u(z, \lambda)}{E_{0u}(z, \lambda)},$$

respectively. The average cosines are sometimes inverted to define the *distribution function*: $D_d = 1/\langle\mu_d\rangle$ and $D_u = 1/\langle\mu_u\rangle$. The collection of AOPs, K 's, R 's and $\langle\mu\rangle$'s, are useful to the extent that they depend only weakly on depth in a homogeneous medium.

One of the biological areas in which oceanic optics makes a direct contribution is primary productivity by marine phytoplankton. This productivity is effected through photosynthesis, which of course requires light. Photosynthesis is a quantum process, e.g., for algae to undergo photosynthesis a particular energy state in a chlorophyll a molecule must be excited. This excitation can be achieved through absorption of a photon of higher energy by the chlorophyll molecule (or an accessory pigment molecule) than the required excitation energy, and subsequent non-radiative transfer to the required state (the excess energy being degraded to heat). Thus, a photon of virtually any energy in the visible spectrum can (when absorbed by an algal particle) be available for photosynthesis. Because of this, the relevant quantity associated with the visible light field is not the total irradiance (energy flux) in the spectrum, but the total photon flux in the visible spectrum. Since the energy associated with a photon of wave length λ is hc/λ . The total number of quanta in the visible spectrum, associated with the irradiance $E_d(z, \lambda)$, is

$$E_{Qd}(z) = \frac{1}{hc} \int_{400 \text{ nm}}^{700 \text{ nm}} \lambda E_d(z, \lambda) d\lambda.$$

The units are easily seen to be quanta/m²s. The quantity $E_{Qd}(z)$ is called the downwelling *quantum irradiance* at depth z . Likewise the upwelling and scalar quantum irradiances ($E_{Qu}(z)$ and $E_{Q0}(z)$) are defined by replacing E_d in the above equation with E_u and

E_0 , respectively. As might be expected, as for all the irradiances, there are K -values defined for quantum irradiance, e.g.,

$$K_{Q0}(z) \triangleq -\frac{1}{E_{Q0}(z)} \frac{dE_{Q0}(z)}{dz};$$

however, because of the strong variation of water absorption with wavelength, the K 's associated with the quantum irradiances display a strong variation with depth, especially at the surface.

Since the direction from which the photon is absorbed is of no consequence to an individual algal particle, the scalar quantum irradiance is the most important of the three and is often just called the quantum irradiance. A unit for quanta often used by biologists is the Einstein (\mathcal{E}), which is a *mole* of quanta, i.e., 6.023×10^{23} photons. If the lower limit on the integral above is extended to 350 nm, then $E_{Q0}(z)$ is usually referred to as “*photosynthetically available radiation*” or “*PAR(z)*,” and the associated decay coefficient as “ $K_{PAR}(z)$ ”.⁵ In many studies, $E_d(z)$ is used rather than $E_{Q0}(z)$ to estimate $PAR(z)$; however, the two can differ by tens of a percent, depending on the time of day, the depth, etc.

For an idea of the magnitude of E_{Q0} , consider the direct solar beam at the top of the atmosphere. For that $E_0(\lambda) = F_0(\lambda)$, where F_0 is the extraterrestrial solar irradiance. Substituting $F_0(\lambda)$ from Chapter 8 into the above integral for E_{Q0} yields 1.44×10^{21} quanta/m²s or $2.4 \times 10^{-3} \mathcal{E}/\text{m}^2\text{s}$. This is often conveyed using a hybrid set of units which replace meters by centimeters: $E_{Q0} = 0.24 \mu\mathcal{E}/\text{cm}^2\text{s}$, where $\mu\mathcal{E} = 10^{-6} \mathcal{E}$, a micro Einstein.

Finally, the *euphotic depth* is defined as the depth at which $E_{Q0}(z)$ falls to 1% of its value at the surface, when the Sun is near the zenith (and not obscured by clouds). Note that the euphotic depth is a fixed number for a given water body on a given day. Photosynthesis only takes place (for the most part) *above* the euphotic depth (if it takes place at all, e.g., it does not at night). So the euphotic depth is really only a measure of the *maximum* thickness of the surface layer within which photosynthesis can take place.

6.3 AOP Examples

Examples of AOPs are now provided. These examples are of two flavors: (1) derived from experimental measurements of irradiance and radiance carried out in natural waters; and (2) derived from solutions of the radiative transfer equation for specified IOPs similar to those found in natural waters.

⁵Many authors define PAR as E_{Q0} for the spectral range 400–700 nm rather than 350–700 nm.

6.3.1 AOPs Derived from Experimental Measurements

Figure 6.1 provides profiles of E_d and L_u at several wavelengths measured in the clear waters off Lanai, Hawaii. In this example, the IOPs are slowly increasing with depth, with a and b increasing by about 50 and 20%, respectively, from the surface to 100 m. Note that, in contrast to L_u , the E_d profile is very “noisy” near the surface. This noise is the result of fluctuations in the downward propagating light field due to the surface waves (like the bright and dark striations moving on the bottom of a swimming pool in sunlight, due to the waves on the surface). The effect of the surface waves decreases with depth and is small for $z \gtrsim 15$ m. Even in the presence of the noise, it is clear that the irradiance (and the radiance) decays with depth in approximately an exponential manner, with a nearly constant decay coefficient.

From these data we can estimate the profiles of K_d and R_L shown in Figure 6.2. As expected, K_d is very noisy near the surface, but for $z \gtrsim 20$ m the noise is greatly reduced. The value of K_d was computed using a depth interval of 2 m for Δz after the irradiance data were averaged over 11 m, i.e.,⁶

$$\langle E_d(z_j) \rangle = \frac{1}{11} \sum_{i=-5 \text{ m}}^{i=+5 \text{ m}} E_d(z_j + i)$$

The slow increase in K_d between 20 and 70 m follows the associated slow increase in the absorption and scattering coefficients. As there is no differentiation involved, the profile of R_L is much smoother than K_d , except near the surface where E_d is noisy. Thus, the variation of R_L with depth probably better reflects the small-scale variation in the aggregate IOPs than does K_d . The slow decrease of R with depth reflects the slow increase of the absorption coefficient.

Figures 6.3 and 6.4 provide the variation of the attenuation coefficients and R_L with wavelength. The irradiances at 2 and 20 m and the radiances at 2 and 6 m were used to compute K_d and K_L , respectively, and so these K s are really averages of $K(z)$ over these depths. The remote sensing ratio, R_L , was computed at 2 m depth. Both K_d and K_L follow the spectral shape of the absorption coefficient of pure water and have similar values to each other. The divergence of K_L from K_d for $\lambda \gtrsim 500$ nm is due to the increasing influence of Raman scattering on the upwelling radiance as the wavelength increases. The effects of Raman scattering on the in-water light field will be discussed in detail later in this chapter. The radiance reflectance R_L is accompanied by πR_L in Figure 6.4 and is more closely associated with remote sensing. The quantity πR_L is similar to the irradiance reflectance, R .⁷ Due to the high reflectance near 400 nm, the water at this location appears

⁶The values of K_d were virtually unchanged when $\ell n(E_d)$ replaced E_d in this equation.

⁷If it is assumed that the upward-propagating radiance is independent of the direction of propagation,

to be a deep blue (the maximum reflectance is actually in the ultraviolet near 389 nm). These spectra are characteristic of clear natural waters, that is, waters containing minimal concentrations of dissolved and suspended material. The historical average concentration of Chlorophyll *a* at this location is approximately 0.1 mg/m³.

Figures 6.5 and 6.6 provide the typical spectral variation of R and K_d for Case 1 waters⁸ with a range of pigment concentrations provided in Table 6.1. For these data, spectra labeled “D” are for Sargasso Sea and Caribbean waters with one in the Pacific near the Panama Canal (D15). Spectra labeled “C” are from off the coast of West Africa near Mauritania. The R spectra show a decrease in the reflectance in the blue and an increase in the green with increasing pigment concentration. The opposite effect is seen in K_d . The reflectance maximum near 685 nm is due to the natural fluorescence of chlorophyll *a* and will be discussed later.

Table 6.1: Pigment concentrations (mg/m³) at the various stations in Figures 6.5, 6.6 and 6.7. The ordering of the stations conforms to the order in the legends to the figures.

Station	Concentration
D23	0.028
D21	0.027
D18	0.039
D19	0.041
D15	0.180
C70	2.4
C92	9.0
C35	2.1

The AOPs of Case 2 waters are much more complicated and variable than those shown above for Case 1 waters. We show only one example here: spectra for sediment-dominated Case 2 waters, i.e., waters for which mineral particles have a strong influence on the IOPs. This example is shown in Figure 6.7 for two locations with nearly the same pigment concentration (2.1 mg/m³ for C35 and 2.4 mg/m³ for C70) but vastly different suspended particle concentrations.

i.e., completely diffuse, then $R = \pi R_L$.

⁸Recall from Chapter 5 that Case 1 waters are those for which the pigment concentration determines the IOPs, i.e., the optical effects of all of the constituents can be linked to the pigment concentration. In Case 2 waters this is not the case: the optical effect of the constituents is not solely determined by the pigment concentration.

6.3.2 AOPs Derived from Computed Light Fields

The most-often measured AOPs are K_d , R , and R_L . It is rare to have measurements of the other AOPs, e.g., $\langle\mu\rangle$, etc. Also it is rare to have measurements in a truly homogeneous medium with an absolutely flat surface. To provide examples of the rarely-measured AOPs and the effects of the wind-roughened surface and the atmosphere on the AOPs, it is helpful to look at AOPs derived from solutions of the radiative transfer equation under idealized conditions. This of course also allows isolation of the individual effects of Sun angle, surface roughness, etc. from one another. Here we provide examples computed at two wavelengths of interest in ocean color remote sensing: 440 and 550 nm. In the examples, the water IOPs are taken from Chapter 5. The particles are chosen to be weakly absorbing ($\omega_0 = 0.93$), and the spectral variation of the particle scattering coefficient chosen to be approximately independent of wavelength. The values of absorption coefficient a , the scattering coefficient b and the beam attenuation coefficient (extinction coefficient) c of the medium are $a(440) = 0.0171 \text{ m}^{-1}$, $b(440) = 0.148 \text{ m}^{-1}$, $c(440) = 0.165 \text{ m}^{-1}$, $a(550) = 0.067 \text{ m}^{-1}$, $b(550) = 0.144 \text{ m}^{-1}$, and $c(550) = 0.211 \text{ m}^{-1}$. The scattering phase function for the particles is that labeled “Petzold Phase-T” in Chapter 5.⁹ The polarization of light and Raman scattering are ignored.

Figure 6.8 provides depth profiles of the irradiance reflectance $R(z)$ with the upper graphs for 440 nm and the lower for 550 nm. Plotted on each panel are the results for two solar zenith angles, for a total of four solar zenith angles for each wavelength. The results are shown for a flat surface and one that has a surface roughness characterized by the Cox-Munk distribution with a surface slope variance, σ^2 (Chapter 7), of zero, and a wind-roughened surface characterized by $\sigma = 0.2$, corresponding to a wind speed of approximately 3.7 m/s.¹⁰ For each solar zenith angle (θ_0) there are two cases: one for the sun in a black sky, i.e., no atmosphere; and one with an atmosphere containing both molecular and aerosol scattering. At 440 nm, the atmosphere’s Rayleigh (τ_r) and aerosol (τ_a) optical depths are 0.252 and 0.295, respectively, while at 550 nm, the corresponding optical depths are 0.099 and 0.250. These aerosol optical depths would be considered to be quite large in a maritime setting (Chapter 4). Figure 6.8 shows that at 440 nm (top panels) there is a strong dependence of $R(z)$ on the solar zenith angle near the surface for $\theta_0 \leq 60^\circ$, with a much less dependence on the surface roughness or the presence or absence of the atmosphere. For $\theta_0 = 80^\circ$ there is a large increase in reflectance when the atmosphere is absent and even larger when the surface is roughened. The roughened surface causes some photons to enter the water at large angles relative to the surface normal, and this increases the probability that they will be scattered in upward directions. When the atmosphere is

⁹This phase function is also called “Petzold,” or just “Phase-T” later in this chapter.

¹⁰Note that in Chapter 7 we have mostly used $\tilde{\sigma}$ rather than σ for the surface slope variance, where $\sigma^2 = 2\tilde{\sigma}^2$.

added, however, there is very little difference between results for $\theta_0 = 80^\circ$ and 60° . The results are similar at 550 nm, with two exceptions: first, the overall reflectance is much smaller, which is associated with the larger absorption coefficient compared to 440 nm; and the effect of the atmosphere is smaller due to the smaller value of τ_r at 550 nm. One of the most striking features is the fact that all of the curves seem to be approaching an asymptotic regime in which $R(z)$ becomes virtually independent of depth. The existence of this regime was shown theoretically in Chapter 2 (Section 2.7.2). Note also that even in the cases with the most extreme variation in depth, the variation in R with z is *usually* less than ± 10 to 20% from the surface to 100 m — hence the notion of its being an *apparent* optical property of the medium.

Figure 6.9 shows trends in K_d with the variables that are similar to those for R , with the exception of the fact that K_d at 550 is larger than that at 440, i.e., the reverse of that for R . This of course suggests that R varies inversely with the absorption coefficient, while K_d varies directly with the absorption coefficient. The asymptotic behavior of K_d with depth is also evident, as it is with two other AOPs $\langle \mu_d \rangle$ and $\langle \mu \rangle$ shown in Figure 6.10. The various average cosines are related, i.e., it is easy to verify from the definitions that

$$R \frac{\langle \mu_d \rangle}{\langle \mu_u \rangle} = (1 - R) \frac{\langle \mu_d \rangle}{\langle \mu \rangle} - 1,$$

so given $\langle \mu_d \rangle$, $\langle \mu \rangle$ and R one can find $\langle \mu_u \rangle$. Since the left-hand-side of this equation must be positive, this shows that $(1 - R)\langle \mu_d \rangle / \langle \mu \rangle \geq 1$ or $\langle \mu_d \rangle \geq \langle \mu \rangle$, with the equality holding when $R = 0$, i.e., no upwelling radiance at all. In ocean optics, the $\langle \mu \rangle$'s are important in the application of Gershun's equation for estimating the absorption coefficient. If inelastic processes are unimportant, the Gershun's equation states (Chapter 2),

$$a(z) = -\frac{1}{E_0(z)} \frac{d}{dz} [E_d(z) - E_u(z)].$$

This can be reduced to

$$a = \langle \mu \rangle K_V,$$

where $K_V \triangleq -\partial \ln[E_d - E_u] / \partial z$ is the attenuation coefficient for vector irradiance. Thus, given the irradiances, we could determine the IOP $a(z)$ knowing the vertical profile of $\langle \mu \rangle$. This equation can also be manipulated to read

$$a = \langle \mu_d \rangle K_d \left[\frac{1 - RK_u/K_d}{1 + R\langle \mu_d \rangle / \langle \mu_u \rangle} \right],$$

which shows that when $R \ll 1$, $a \approx \langle \mu_d \rangle K_d$. This is important because under these conditions, and close to the surface, $\langle \mu_d \rangle$ can be approximated by $\cos \theta_{0w}$, where θ_{0w} is the zenith angle of the refracted solar beam in the water. Because R is usually $\ll 1$, this relationship ($a \approx \langle \mu_d \rangle K_d$) is very robust, and allows estimation of the near-surface a from

K_d . We shall return to the question of estimation the IOPs from the AOPs later in an appendix to this chapter.

Two quantities that are of prime interest in remote sensing of water properties are $Q = E_u/L_u$, and especially, $L_u/E_d = R/Q$. These quantities are provided in Figures 6.11 and 6.12, respectively. Note that the “noise” in these figures is due entirely to statistical fluctuations in the Monte Carlo(Chapter 2) solution to the radiative transfer equation that was used here. The fluctuations increase with increasing depth and are much larger in the upwelling quantities than the downwelling quantities and are particularly severe in L_u . As with the other AOPs, the solar zenith angle has the dominate influence on Q for a given set of AOP’s; however, when combined with R to form R/Q the effect of the sun angle is greatly reduced (Figure 6.12). In fact, near the surface, the depth of most interest in remote sensing, R/Q for these cases varies by only 5-6% as the solar zenith angle varies from 0 to 60°. Since the remotely sensed signal for any near nadir viewing sensor is proportional to L_u , and E_d can be estimated based on the optical properties of the atmosphere, it is of prime importance to be able to relate $R_L = R/Q$ just beneath the surface to the IOPs.

Finally, in Figure 6.13 we provide profiles of the various K_x ’s. Note that there is significant differences in the values of the various K ’s near the surface (up to almost a factor of 2); however, of interest here is the observation that K_L and K_d are close in value near the surface. This of course is the source of the weak variation in $R/Q = L_u/E_d$ with depth seen in Figure 6.12 compared to either R or Q . This nearness of K_L and K_d holds for all of the cases examined in the figures, i.e., over all of the cases examined, at a depth of 6 m, $0.951 \leq K_L/K_d \leq 1.002$ while, in contrast, $0.761 \leq K_u/K_d \leq 1.014$.

6.4 Relationships between AOPs and IOPs

The main difficulty establishing relationships between AOPs and IOPs is that they are connected through the difficult-to-solve radiative transfer equation (difficult in terms of closed-form expressions). Although codes for solving the radiative transfer equation in a marine environment for any set of IOPs are now readily available, and must be used if high-accuracy results are needed, it is instructive to try to develop simple models of the multiple scattering process to isolate the more important IOPs and to provide simple parameterizations of the AOPs in terms of the IOPs. Here we develop such a model using the quasi-single scattering approximation (QSSA) as the basis for an approximate solution of the full radiative transfer equation. We then examine the full range of validity of the model for the various AOPs through comparison with “exact” numerical solutions of the radiative transfer equation. Such models are essential both for understanding the causes of variability in the AOPs seen in natural waters, and for the retrieval of the IOPs given

the AOPs – one of the central problems of ocean color remote sensing.

We begin by developing the QSSA solution for a semi-infinite homogeneous medium illuminated by the sun in a black sky and using it to determine the AOPs. Then we methodically examine complicating factors: a flat refracting interface at the surface of the medium, a rough refracting interface at the surface, an atmosphere above the medium, etc. Finally, we consider vertically stratified water bodies and the effects of inelastic processes: fluorescence and Raman scattering.

6.4.1 The QSSA – A Simple Model to Develop AOP–IOP Relationships

Consider a water body modeled as a semi-infinite medium (surface at $z = 0$) of uniform IOPs and illuminated from above by the sun and sky. For the moment we assume that the refractive index of the water is unity in order to remove the complications arising from refraction and reflection at the air-water interface. The medium is assumed to be free of internal sources such as inelastic scattering. The scalar radiative transfer equation governing the propagation of radiance in the $\hat{\xi}$ direction at a depth z within such a medium is

$$\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial z} L(z, \hat{\xi}, \lambda) = -c(z, \lambda) L(z, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(z, \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(z, \hat{\xi}', \lambda) d\Omega(\hat{\xi}'),$$

or, dropping the explicit reference to the wavelength λ , noting the the IOPs are independent of z , writing $\tau = cz$, $\omega_0 = b/c$, and replacing the volume scattering function by the phase function, we have

$$\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L(\tau, \hat{\xi}) = -L(\tau, \hat{\xi}) + \frac{\omega_0}{4\pi} \int_{\text{All } \hat{\xi}'} P(\hat{\xi}' \rightarrow \hat{\xi}) L(\tau, \hat{\xi}') d\Omega(\hat{\xi}'). \quad (6.7)$$

The solution requires the incident radiance on the boundary $L(0, \hat{\xi})$. Because the phase function is sharply peaked in the forward direction (Figure 5.15), i.e., $\hat{\xi}' = \hat{\xi}$, the integral term can be quite large and cannot be ignored. However, some simplification can be achieved by replacing the phase function by a Dirac-delta function at $\hat{\xi}' = \hat{\xi}$, i.e., $\delta^{(2)}(\hat{\xi}' - \hat{\xi})$, plus a more slowly varying (with scattering angle) function $P_r(\hat{\xi}' \rightarrow \hat{\xi})$, where the subscript “ r ” stands for “remaining,” i.e.,

$$P(\hat{\xi}' \rightarrow \hat{\xi}) = A_\delta \delta^{(2)}(\hat{\xi}' - \hat{\xi}) + A_r P_r(\hat{\xi}' \rightarrow \hat{\xi}). \quad (6.8)$$

We can relate A_δ and A_r by integrating this phase function over all solid angles and over those for which $\hat{\xi}' \bullet \hat{\xi} < 0$ (i.e., all scattering angles $\Theta > 90^\circ$):

$$\int_{\text{All } \hat{\xi}} P(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}) = 4\pi = A_\delta + A_r \int_{\text{All } \hat{\xi}} P_r(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi});$$

and

$$\int_{\hat{\xi}' \bullet \hat{\xi} < 0} P(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}) = 4\pi B = A_r \int_{\hat{\xi}' \bullet \hat{\xi} < 0} P_r(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}),$$

where we employ the symbol “ B ” for the backscattering probability rather than the more cumbersome \tilde{b}_b usually used in the literature. Substituting Eq. (6.8) into Eq. (6.7) we obtain

$$\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau} L(\tau, \hat{\xi}) = -L(\tau, \hat{\xi}) \left(1 - \frac{\omega_0 A_\delta}{4\pi} \right) + \frac{\omega_0 A_r}{4\pi} \int_{\text{All } \hat{\xi}'} P_r(\hat{\xi}' \rightarrow \hat{\xi}) L(\tau, \hat{\xi}') d\Omega(\hat{\xi}'),$$

or

$$\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial \tau'} L(\tau', \hat{\xi}) = -L(\tau', \hat{\xi}) + \frac{\omega'_0}{4\pi} \int_{\text{All } \hat{\xi}'} P_r(\hat{\xi}' \rightarrow \hat{\xi}) L(\tau', \hat{\xi}') d\Omega(\hat{\xi}'), \quad (6.9)$$

which is the same as Eq. 6.7 but with

$$\begin{aligned} \tau \rightarrow \tau' &= \tau \left(1 - \frac{\omega_0 A_\delta}{4\pi} \right) \\ \omega_0 \rightarrow \omega'_0 &= \frac{\omega_0 A_r}{(1 - \omega_0 A_\delta / 4\pi)}. \end{aligned} \quad (6.10)$$

Since P_r lacks the strong forward peak (the δ -function removed it) the integral term in Eq. (6.9) is much smaller than in Eq. (6.7) which suggests that a good approximation to the radiance might actually be possible by solving Eq. (6.9) in the single-scattering approximation.

In Chapter 2 we developed the single scattering solution to Eq. 6.7 for a source of illumination $L_0(u, \phi)$, $u > 0$, incident on a slab of optical thickness $\tau_1 = cz_1$ from above:

$$\begin{aligned} L_d(\tau, u, \phi) &= L_0(u, \phi) \exp[-\tau/u] \\ &+ \frac{\omega_0}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi) L_0(u', \phi')}{1 - u/u'} \\ &\times (\exp[-\tau/u'] - \exp[-\tau/u]); \quad u > 0, \end{aligned} \quad (6.11)$$

and

$$\begin{aligned} L_u(\tau, u, \phi) &= \frac{\omega_0}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi) L_0(u', \phi')}{1 - u/u'} \\ &\times (\exp[-\tau/u'] - \exp[-\tau/u] \exp[-\tau_1(1/u' - 1/u)]); \quad u < 0, \end{aligned} \quad (6.12)$$

If the slab is infinitely thick, $\tau_1 \rightarrow \infty$. It is often convenient to replace u by $-\mu$ when $u < 0$. This is equivalent to measuring the polar angle of propagation of the upward radiance from the *outward* normal contrary to u , which is measured relative to the *downward* normal, i.e.,

into the medium, so $\mu = \cos \theta_{up}$. If we apply this solution to the reduced radiative transfer equation Eq. (6.9) and make the indicated changes,

$$\begin{aligned} L_d(\tau', u, \phi) &= L_0(u, \phi) \exp[-\tau'/u] \\ &+ \frac{\omega'_0}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' \frac{P_r(u' \rightarrow u, \phi' \rightarrow \phi) L_0(u', \phi')}{1 - u/u'} \\ &\times (\exp[-\tau'/u'] - \exp[-\tau'/u]); \quad u > 0, \end{aligned}$$

and

$$\begin{aligned} L_u(\tau', \mu, \phi) &= \frac{\omega'_0}{4\pi} \int_0^1 du' \int_0^{2\pi} d\phi' \frac{P_r(u' \rightarrow -\mu, \phi' \rightarrow \phi) L_0(u', \phi')}{1 + \mu/u'} \\ &\times (\exp[-\tau'/u']); \quad \mu > 0, \end{aligned}$$

Now if the incident radiance is in the form of a parallel beam propagating in the direction $\hat{\xi}_0 = (u_0, \phi_0)$, e.g., from the Sun in a black sky, then

$$L_0(u, \phi) = F_0 \delta(u - u_0) \delta(\phi - \phi_0),$$

and the first order solutions become

$$\begin{aligned} L_d(\tau', u, \phi) &= F_0 \delta(u - u_0) \delta(\phi - \phi_0) \exp[-\tau'/u] \\ &+ \frac{\omega'_0 F_0}{4\pi} \frac{P_r(u_0 \rightarrow u, \phi_0 \rightarrow \phi)}{1 - u/u_0} \times (\exp[-\tau'/u_0] - \exp[-\tau'/u]); \quad u > 0, \end{aligned} \quad (6.13)$$

and

$$L_u(\tau', \mu, \phi) = \frac{\omega'_0 F_0}{4\pi} \frac{P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{1 + \mu/u_0} \times (\exp[-\tau'/u_0]); \quad \mu > 0, \quad (6.14)$$

From these expressions, we can compute all of the AOPs in this approximation. Start with the downwelling irradiance:

$$\begin{aligned} E_d(\tau') &= \int_0^{2\pi} d\phi \int_0^1 u L_d(\tau', u, \phi) du = F_0 u_0 \exp[-\tau'/u_0] \\ &+ \frac{\omega'_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \frac{u P_r(u_0 \rightarrow u, \phi_0 \rightarrow \phi)}{1 - u/u_0} \times (\exp[-\tau'/u_0] - \exp[-\tau'/u]) du. \end{aligned} \quad (6.15)$$

Using this we can find the downwelling irradiance attenuation coefficient K_d . First,

$$\begin{aligned} -\frac{dE_d(\tau')}{d\tau'} &= F_0 \exp[-\tau'/u_0] \\ &+ \frac{\omega'_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \frac{u P_r(u_0 \rightarrow u, \phi_0 \rightarrow \phi)}{1 - u/u_0} \times \left(\frac{\exp[-\tau'/u_0]}{u_0} - \frac{\exp[-\tau'/u]}{u} \right) du. \end{aligned}$$

The integral above can only be carried out numerically; however, if we let $\tau' \rightarrow 0$, there is considerable simplification:

$$-\lim_{\tau' \rightarrow 0} \frac{dE_d(\tau')}{d\tau'} \rightarrow F_0 - \frac{\omega'_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P_r(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du.$$

so

$$-\lim_{\tau' \rightarrow 0} \frac{dE_d(\tau')}{E_d(\tau')d\tau'} \rightarrow \frac{1}{u_0} \left(1 - \frac{\omega'_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P_r(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du \right).$$

Now, using Eq. (6.10) and defining

$$P_{rd}(u_0, \phi_0) \triangleq \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P_r(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du,$$

we find

$$\lim_{z' \rightarrow 0} \frac{K_d(z')}{c} \triangleq \frac{K_d(0)}{c} = \frac{1}{u_0} \left(1 - \frac{\omega_0 A_\delta}{4\pi} \right) \left(1 - \frac{\omega_0 A_r}{(1 - \omega_0 A_\delta/4\pi)} P_{rd}(u_0, \phi_0) \right). \quad (6.16)$$

The upwelling irradiance is

$$\begin{aligned} E_u(\tau') &= \int_0^{2\pi} d\phi \int_0^1 \mu L(\tau', \mu, \phi) d\mu \\ &= \frac{\omega'_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{1 + \mu/u_0} \times (\exp[-\tau'/u_0]) d\mu, \end{aligned} \quad (6.17)$$

from which it is trivial to show that

$$-\frac{dE_u(\tau')}{E_u(\tau')d\tau'} = \frac{1}{u_0},$$

so

$$\frac{K_u(z)}{c} = \frac{1}{u_0} \left(1 - \frac{\omega_0 A_\delta}{4\pi} \right). \quad (6.18)$$

Equation 6.14 shows that in this approximation, $K_L(z) = K_u(z)$.

The irradiance reflectance (or irradiance ratio) is given by $R(z) = E_u(z)/E_d(z)$ and is also easy to evaluate just beneath the surface

$$\lim_{z \rightarrow 0} R(z) \rightarrow \frac{\omega'_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} d\mu$$

or

$$R(0) = \frac{\omega_0 A_r}{4\pi (1 - \omega_0 A_\delta/4\pi)} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} d\mu. \quad (6.19)$$

The downwelling average cosine (E_{0d}/E_d) evaluated at $z = 0$ is easily seen to be u_0 .

Finally, the quantity of most interest in remote sensing, the radiance exiting the medium, $L_u(0, \mu, \phi)$, is just

$$\frac{L_u(0, \mu, \phi)}{E_d(0)} = \frac{\omega_0 A_r}{4\pi(1 - \omega_0 A_\delta/4\pi)} \frac{P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu}, \quad (6.20)$$

with $R_L = L_u(0, 1, \phi)/E_d(0) = R/Q$.

The approximations used in this section fall under what is commonly termed the quasi-single scattering approximation — QSSA. The next order of business is to determine the range of validity, if any, of these approximations.

6.4.2 The Range of Validity of the QSSA Model for AOPs.

To understand the range of validity of the QSSA, we compare its predictions with exact solutions of the radiative transfer equation in a variety of situations.¹¹ We begin by employing a specific phase function replacement in Eq. (6.8). We shall assume that $P_r = P$ for a range of scattering angles that can scatter the incident solar beam, propagating along $\hat{\xi}_0 = (\theta_0, \phi_0)$, into the upward hemisphere, i.e., toward the surface. We further assume that $P_r = 0$ for all smaller scattering angles (see Figure 6.14). Then, $A_r = 1$ and

$$A_\delta = 4\pi - \int_{\text{All } \hat{\xi}} P_r(\hat{\xi}_0 \rightarrow \hat{\xi}) d\Omega(\hat{\xi})$$

where the integral in A_δ is easily seen to be

$$\int_{\text{All } \hat{\xi}} P_r(\hat{\xi}_0 \rightarrow \hat{\xi}) d\Omega(\hat{\xi}) = 2\pi \int_{(\pi/2) - \theta_0}^{\pi} P_r(\Theta) \sin \Theta d\Theta \triangleq 4\pi B'.$$

The quantity B' is the probability of scattering through an angle greater than $(\pi/2) - \theta_0$, and when $\theta_0 = 0$, $B' = B$, where B is the backscattering probability. Thus,

$$A_\delta = 4\pi(1 - B').$$

Defining

$$P_{ru}(u_0, \phi_0) \triangleq \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi) d\mu,$$

¹¹Recall, the solutions are numerical, but are “exact” to the extent that they can be made as accurate as desired.

one sees that

$$P_{ru}(u_0, \phi_0) + P_{rd}(u_0, \phi_0) = B',$$

i.e., $P_{ru}(u_0, \phi_0)$ is the vertically striped volume in Figure 6.14 and $P_{rd}(u_0, \phi_0)$ is the horizontally striped volume. Inserting these relationships into Eqs. (6.16)–(6.20) provides the QSSA-estimated AOPs in terms of the IOPs:

$$\begin{aligned} \frac{K_d(0)}{c} &= \frac{1}{u_0} [1 - \omega_0(1 - P_{ru})] \triangleq \frac{QSSA_K(u_0)}{u_0}, \\ \frac{K_u(z)}{c} &= \frac{1}{u_0} [1 - \omega_0(1 - B')], \\ \frac{L_u(0, \mu, \phi)}{E_d(0)} &= \frac{\omega_0}{4\pi[1 - \omega_0(1 - B')]} \frac{P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} \triangleq QSSA_L(u_0, \mu), \\ R(0) &= \frac{\omega_0}{4\pi[1 - \omega_0(1 - B')]} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} d\mu \triangleq QSSA_R(u_0). \end{aligned} \quad (6.21)$$

Note that in the limit $\omega_0 \ll 1$, these equations reduce to the single scattering approximation, i.e., the terms introduced to try to correct the single scattering approximation (SSA) for multiple scattering, $\omega_0(1 - B')$ and $\omega_0(1 - P_{ru})$, become negligible.

One observation that can be immediately drawn from Eqs. (6.21) is that, since $B' \geq P_{ru}$, $K_d(0) \leq K_u(0) = K_L(0)$; however, Figures 6.3 and 6.13, show that in the examples provided $K_d(0) > K_L(0) > K_u(0)$. So, which of the results are better, those for K_u and K_L or that for K_d ? To answer this question, we compare the prediction with exact computations. We use the Phase-T phase function from Chapter 5 and carry out computations for a range of values for ω_0 and θ_0 . As in the QSSA, we assume for the moment that the refractive index of the medium is unity. Figures 6.15 and 6.16 provide the values of P_{ru} and B' , respectively, as a function of the solar zenith angle for this phase function. One sees that for $\theta_0 = 60^\circ$, B' is more than double P_{ru} . Figure 6.17 compares the exact value of $K_d(0)/c$ (y axis) with the QSSA prediction (x axis) for $\theta_0 = 0$ and 60° with $0.1 \leq \omega_0 \leq 0.9$. The solid line is the linear regression of these two quantities. The slope of the regression line is 0.9999 and the R^2 value is 0.9999. Clearly, the QSSA provides an excellent approximation to $K_d(0)$. This implies that the approximation for $K_u(z)$, and possibly the other upwelling quantities, may not be as accurate.

We continue the examination of the range of validity of the QSSA by looking at $R(0)$. To compute the QSSA approximation to $R(0)$, we need the integral

$$\frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} d\mu \triangleq I_R(u_0).$$

Actually, later we will find it more useful to have $I'_R(u_0) \triangleq I_R(u_0)/B$, where $B = B'$ with $u_0 = 1$, the backscattering probability. This is provided in Figure 6.18 for Phase-T (and for

later use, for the water itself – Rayleigh scattering). The comparison between the QSSA and “exact” computations of $R(0)$ is provided in Figure 6.19 for $\theta_0 = 0$ and 60° . Also included on Figure 6.19 is the single scattering approximation (SSA) for the two θ_0 ’s. We see that for both Sun angles the QSSA and exact computations agree well for ω_0 up to about 0.5, after which departures up to a factor of 2 are seen. In contrast, the SSA shows significant error even for $\omega_0 = 0.1$ (4.4% and 11.1% for $\theta_0 = 0$ and 60° , respectively).

We have also included in Figure 6.19 the results of replacing $B'(\theta_0)$ with B in Eq. 6.21 for $R(0)$, i.e.,

$$R(0) = \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} d\mu \triangleq QSSA'_R(u_0).$$

The quantities B' and B are the same for $\theta_0 = 0$, but $B' \approx 6.6B$ for $\theta_0 = 60^\circ$. With this replacement there is a considerable improvement between the QSSA and the exact computations. Figure 6.20 shows this relationship plotted in another way and suggests an almost linear dependence of $R(0)$ on $QSSA'_R(u_0)$:

$$R(0) \approx 1.12 QSSA'_R(u_0). \quad (6.22)$$

We now turn to the quantity of most interest in remote sensing: $L_u(0, \mu, \phi)$. Since this quantity depends on u_0 and ϕ_0 in addition to μ and ϕ , it is useful in this section to indicate this explicitly in the notation by the following replacement:

$$L_u(0, \mu, \phi) \longrightarrow L_u(0, \mu, \phi; u_0, \phi_0).$$

Let us start by examining $L_u(0, \mu = 1, \phi; u_0, \phi_0)$, the radiance at the surface propagating toward the zenith, given by

$$\frac{L_u(0, \mu = 1, \phi; u_0, \phi_0)}{E_d(0)} = QSSA_L(\mu = 1, u_0) = \frac{\omega_0}{4\pi[1 - \omega_0(1 - B'(u_0))]} \frac{P_r(u_0 \rightarrow -1, \phi_0 \rightarrow \phi)}{u_0 + 1} \quad (6.23)$$

where we have explicitly indicated that B' is a function of u_0 . Note that in the above equation there is really no dependence on ϕ or ϕ_0 because for this geometry the scattering angle is independent of ϕ and ϕ_0 . Based on our experience with $R(0)$, we expect that the quality of this equation in predicting the correct L_u/E_d will be improved by replacing $B'(u_0)$ by B . Let’s see if this is the case.

Consider Figure 6.21 for the Petzmas phase function. In the left panel the QSSA formula uses the actual $B'(u_0)$ to make the prediction, i.e., Eq. (6.23). Note that at larger values of $QSSA_L(1, u_0)$, in addition to deviating from the prediction the comparison shows a strong, systematic, dependence on θ_0 (or u_0). For a given ω_0 , as u_0 is varied the points fall on arcs

with the largest values of L_u/E_d corresponding to $\theta_0 = 60^\circ$ and the smallest to $\theta_0 = 0$. This strong dependence on θ_0 cannot be caused by the phase function, which only varies by about 10% over the relevant range of scattering angles (Figure 5.17). The only other candidate for this behavior is $B'(u_0)$. The right panel shows the effect of replacing $B'(u_0)$ by $B'(1) = B$ in Eq. (6.23). In this case the points cluster better and the deviation from the QSSA prediction is much smaller and more systematic. Thus, as with $R(0)$, we see that replacing $B'(u_0)$ with B improves the agreement with exact calculations, yielding a near-linear relationship between $L_u(0, \mu = 1, \phi; u_0, \phi_0)/E_d(0, u_0, \phi_0)$ and $QSSA'_L(\mu, \phi; u_0, \phi_0)$ defined through

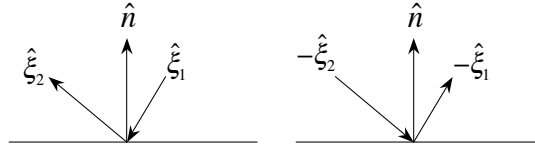
$$QSSA'_L(\mu, \phi; u_0, \phi_0) \triangleq \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \frac{P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu}, \quad (6.24)$$

with $\mu = 1$.

There is an additional (and favorable) consequence of replacing B' by B . It was shown earlier that the reciprocity principle (Chapter 2, Eq. 2.35) applied the boundary of a medium without internal sources and with unit refractive index requires

$$\frac{L_u(\hat{\xi}_1 \rightarrow \hat{\xi}_2)}{E_d(\hat{\xi}_1)} = \frac{L_u(-\hat{\xi}_2 \rightarrow -\hat{\xi}_1)}{E_d(-\hat{\xi}_2)},$$

where in this equation the incident direction is to the left of the arrows and the reflected direction to the right of the arrows. The left figure below shows a schematic of the processes associated with the left-hand-side of this equation, and the right figure the right-hand-side. (\hat{n} is the surface normal.) In the present notation, where $\mu = -u$,



$$\frac{L_u(0, \mu = 1, \phi; u_0, \phi_0)}{E_d(0, u_0, \phi_0)} = \frac{L_u(0, -u_0, -\phi_0; u = 1, -\phi)}{E_d(0, u = 1, -\phi)},$$

where the irradiances explicitly indicate the direction of the incident beam, i.e., $E_d(0, u_0, \phi_0)$ is the downward irradiance when the incident beam is propagating toward (u_0, ϕ_0) . Now $P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi) = P_r(\mu \rightarrow -u_0, \phi \rightarrow \phi_0)$ and, as the scattering angle depends on ϕ and ϕ_0 only through $|\phi - \phi_0|$, we see that satisfying reciprocity principle requires that B' in Eq. (6.23) cannot depend on u_0 .

An example of how well the modified QSSA reproduces the variation of $L_u(0, \mu = 1, \phi; u_0, \phi_0)$ with u_0 , we computed “exact” values for this quantity using two scattering phase functions: Petzold Phase-T and Petzmas (Chapter 5). The results of the computation are provided in Figure 6.22 for several values of ω_0 along with the $QSSA'_L(1, u_0)$ prediction. It shows that for small ω_0 (e.g., $\omega_0 \lesssim 0.5$) the departures from the QSSA prediction are small, but they increase with increasing ω_0 . As the Petzmas phase function is now thought to be more representative of the scattering by particles in natural waters for scattering angles in the backscattering hemisphere, one sees that the departure of $L_u(0, \mu, \phi; 1, \phi_0)$ from the behavior predicted by $QSSA'_L(1, u_0)$ is at most expected to be about 10% and with the QSSA *always* yielding a smaller value for all solar zenith angles.

In sum, the modified QSSA (by replacing B' with B) appears to yield relatively accurate results for $\omega_0 \lesssim 0.5$ and is a considerable improvement on the single scattering approximation, which for similar accuracy requires a much smaller upper limit on ω_0 . However, there are also inconsistencies, e.g., the prediction that $K_d < K_u$ is not borne out by exact calculations. Still, the approximation is useful, providing relatively simple formulae for writing the AOPs as functions of the IOPs, and for investigating the variation of the AOPs resulting from changes in the IOPs. We shall see later, they also provide a framework estimating IOPs from measurements of the AOPs.

6.4.3 Decomposition of Scattering into Water and Particle Components

In the last section we modified the QSSA by replacing $QSSA_R$ and $QSSA_L$ by $QSSA'_R$ and $QSSA'_L$, respectively. This considerably improved the agreement between the relationships provided in Eq. (6.21) and $R(0)$ and L_u/E_d (nadir viewing). Noting that the scattering by the water itself can make an important contribution to the volume scattering function, especially in the backscattering directions, we now modify the AOP-IOP formulas to separate the water and particle scattering. If we write $b = b_p + b_w$, $c = c_p + c_w$ and $\beta = \beta_p + \beta_w$, where the subscripts “ p ” and “ w ” stand for “particles” and “water”, respectively, then

$$\begin{aligned}\omega_0 &= \frac{b_p + b_w}{c_p + c_w}, \\ P(\Theta) &= \frac{P_p(\Theta)b_p + P_w(\Theta)b_w}{b_p + b_w},\end{aligned}\tag{6.25}$$

where Θ is the scattering angle, and the unscripted quantities refer to water *plus* particles. To include water in the QSSA or modified QSSA formulation we need quantities such as $P_{ru}(u_0)$, $B'(u_0)$, and $I_R(u_0)$ in Eq. (6.21). The Rayleigh phase function is symmetric around $\Theta = 90^\circ$, therefore $P_{ww}(u_0) = 0.5$, $B'_w(u_0) = 0.5$, and $B_w = 0.5$. The calculation

of $I_R(u_0)$ is just straightforward integration, and is shown in Figure 6.18.¹² Thus, for example

$$P_u(u_0) = \frac{P_{up}(u_0)b_p + P_{uw}(u_0)b_w}{b_p + b_w},$$

and

$$I_R(u_0) = \frac{I_{Rp}(u_0)b_p + I_{Rw}(u_0)b_w}{b_p + b_w},$$

etc., and we can then apply the QSSA or modified QSSA to problems including both particle and water scattering given b_p and b_w (see Chapter 5 for the variation of b_p and b_w with wavelength, Chlorophyll *a*, etc.).

Since the addition of water has a considerable effect on the overall phase function of the medium, especially for $\Theta \gtrsim 90^\circ$, we need to see if this changes the efficacy of the modified QSSA. Furthermore, in Chapter 5 we introduced a model for the particle phase function that depended on the concentration of Chlorophyll *a*. The end members of this model were modified Petzmas phase functions with backscattering probabilities $B_p = 0.0019$ and 0.014 . (Figure 5.23 in Chapter 5 shows these phase functions.) Thus, we need to assess how well the modified QSSA works when water is added to a particle phase function similar to Petzold's Phase-T or to one much more strongly forward scattering. This assessment is accomplished here by carrying out exact computations with water added to the Petzold Phase-T in amounts required to render the total backscattering probability,

$$B = \frac{B_p b_p + B_w b_w}{b_p + b_w},$$

equal to $0.02(0.02)0.12$.¹³ In addition, Phase-T was modified by enhancing the scattering in the range $0 < \Theta < 2^\circ$ by a factor of approximately 33, which resulted in a B_p of 0.002. Because they have the same angular dependence in the backward scattering directions, the Phase-T and modified Phase-T phase functions have the same $I_{Rp}(u_0)/B_p$ (Figure 6.18). Water was also added to this model to again render $B = 0.02(0.02)0.12$ as well as smaller values of B . This encompasses much of the variation of B and ω_0 for the IOP-Chlorophyll *a* relationship model described in Chapter 5 (Figure 6.23). The resulting phase functions (particles *plus* water) are shown in Figure 6.24 for selected values of B . Note that the biggest difference between the two sets is the stronger forward scattering for the modified Phase-T ($B_p = 0.002$). We would expect the AOPs would follow the QSSA relationships

¹²In the case where the depolarization factor for water is ignored, i.e., $P(\Theta) = 3(1 + \cos^2 \Theta)/4$, $I_{Rw}(u_0)$ is easily found to be

$$I_{Rw}(u_0) = \frac{3}{8} \left[\left(\frac{3}{2} - \frac{1}{2}u_0 \right) \left(1 - \ell n \left(\frac{1+u_0}{u_0} \right) \right) + \left(\frac{3}{2}u_0^2 - \frac{1}{2} \right) \left(\frac{1}{3} - \frac{u_0}{2} + u_0^2 - u_0^3 \ell n \left(\frac{1+u_0}{u_0} \right) \right) \right].$$

¹³The notation $i(j)k$ means from i to k in increments of j .

better for the modified Phase-T, as it more resembles a Dirac delta function at small angles. Also, note that for $B = 0.10$ and $\Theta \gtrsim 90^\circ$ there is little difference between the phase functions for the water *plus* particles.

Figure 6.25 compares $R(0)$ with $QSSA'_R(u_0)$ for the two particle phase functions for $\theta_0 = 10^\circ, 37^\circ$ and 60° , with $\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85$, and 0.9 , and $B = 0.020$ to 0.120 for $B_p = 0.0181$, and $B = 0.003$ to 0.12 for $B_p = 0.002$. Again, the modified QSSA works better for the $B_p = 0.002$ case, although in both cases the QSSA-predicted $R(0)$'s are usually too low. The average (absolute) error in the modified QSSA is 1.59% for $B_p = 0.002$ and 6.27% for $B_p = 0.0181$. The largest errors are usually associated with the largest values of ω_0 (more multiple scattering). Linear regressions

$$R(0) = C_R QSSA'_R(u_0) \quad (6.26)$$

yielded $C_R = 1.0321$ and 1.09800 for $B_p = 0.002$ and 0.181 , respectively. These regressions do *not* change the average absolute error significantly, but *do* remove most of the bias.

Of more interest is the upwelling radiance L_u . Figure 6.26 compares exact values of L_u/E_d (nadir-viewing) with the modified QSSA prediction, $QSSA'_L(u_0)$. Not surprising, the comparisons of L_u/E_d and $QSSA'_L(u_0)$ are similar to those for $R(0)$ and $QSSA'_R(u_0)$. The average (absolute) error in $QSSA'_L$ is 1.57% for $B_p = 0.002$ and 6.45% for $B_p = 0.0181$, and in

$$\frac{L_u}{E_d} = C_L QSSA'_L(u_0), \quad (6.27)$$

the values of C_L are 1.0363 ($B_p = 0.002$) and 1.1179 ($B_p = 0.0181$). Additional simulations show that much of the bias in the QSSA can be removed by taking $C_L \approx 0.99 + 6.9B_p$.

For K_d , Figure 6.27 provides results for IOP parameters that are the same as those in Figures 6.25 and 6.26, and shows that $QSSA_K(u_0)$ provides an excellent estimate of $K_d(0)$ in the presence of water as well as particle scattering.

Clearly, the QSSA expressions $QSSA'_L$, $QSSA'_R$, and $QSSA_K$ reproduce their associated AOPs quite well. As expected, the accuracy is higher the more strongly the particles scatter in the near-forward direction.

6.4.4 Influence of the Environmental Setting on QSSA Validity

Thus far we have ignored the fact that the water surface is refracting, and is usually ruffled by the wind. In addition, we have ignored the atmosphere above the water surface. Here we examine any changes in the QSSA model that may be required to accommodate these omitted features of the water-atmosphere system.

6.4.4.1 A Flat Refracting Interface

We have assumed the refractive index of water (m_w) was unity rather than about 1.333. The air-water refracting interface will influence the light field in the medium in two important ways: (1) the solar beam will be refracted so as to make a smaller angle θ_{0w} in the medium given by

$$\sin \theta_{0w} = \frac{1}{m_w} \sin \theta_0;$$

and (2) some upwelling radiance in the water will be reflected at the interface and rejoin the downwelling light field. Of course, fewer photons will enter the water because some will be Fresnel-reflected from the surface.

In spite of these complications, one might reasonably expect Eqs. (6.21) should still be applicable if u_0 is replaced by $u_{0w} = \cos \theta_{0w}$, and $L_u(0, \mu, \phi; u_0, \phi_0)/E_d(0)$ and $R(0)$ are evaluated *just beneath the interface*. This is to be expected because for $\omega_0 \lesssim 0.5$, the upwelling field is small, so its reflection from the interface should provide very little enhancement of the downwelling light field, and there will be very little deviation from illumination with a parallel beam having zenith angle θ_{0w} , but *without* the interface. However, as ω_0 increases, the illumination (downward radiance just beneath the surface) will become less and less like a beam source. To examine the validity of this conjecture, we compare solutions to the radiative transfer equation in the absence and presence of the refracting interface. Such comparisons are made in Figure 6.28, which provides the ratio of L_u/E_d (nadir-viewing), $R(0)$, and $K_d(0)$ for medium refractive indices of 1.333 and 1.000 (i.e., L_u/E_d for $m = 1.333$ divided by L_u/E_d for $m = 1.000$, etc.). To test our hypothesis, we use $\theta_0 = 40^\circ$ for $m = 1.000$, and $\theta_0 = 58.96^\circ$ for $m = 1.333$, so that the θ_{0w} for $m = 1.333$ is equal to θ_0 for $m = 1.000$. The phase function is Petzold's Phase-T. The ratios are noisy because the quantities are computed from radiative transfer simulations using the Monte Carlo method (Chapter 2). The radiances (L_u) are estimated to be accurate to within $\pm 0.8\%$, the reflectance to within $\pm 0.2\%$, and K_d to within about $\pm 0.1\%$. The ratios show that the reflecting interface has virtually no effect on L_u/E_d , i.e., the ratio is unity to within the accuracy of the simulations. In contrast, for R and K_d we see an ever increasing effect with ω_0 , with R increasing 5.3% and K_d increasing 8% as one increases m from 1.000 to 1.333 with $\omega_0 = 0.9$. This behavior of R and K_d is due to the internal reflection of upwelling light from the interface. Most internally-reflected photons will have values of θ greater than the critical angle (48.6°) and be more strongly attenuated with depth when they join the downwelling stream than photons in the refracted solar beam. This will result in an increased K_d . Internally-reflected photons also have a higher probability of being scattered back into the upwelling field than photons in the direct solar beam, and this increases R . Thus, results seen in Figures 6.25 and 6.26 are easily explained by internal reflection of the upwelling light field.

In the case of K_d , Figure 6.28 suggests that an expansion in increasing powers of $QSSA_K$ might be useful. However, it is more appropriate to perform the regression in the following manner:

$$u_{0w} \frac{K_d}{c} - 1 = C_{K1}[\omega_0(1 - P_{ru})] + C_{K2}[\omega_0(1 - P_{ru})]^2 + C_{K3}[\omega_0(1 - P_{ru})]^3; \quad (6.28)$$

as this ensures that K_d reduces to the correct value for $\omega_0 = 0$, i.e., $K_d = a/u_{0w}$. Note that in the above equation P_{ru} is evaluated at u_{0w} . Interestingly, the best fit is obtained with $C_{K2} = C_{K3} = 0$. The quality of that fit (with $C_{K1} = 0.9892$) is shown in Figure 6.29, where the average unsigned error in K_d/c is 0.25% and the maximum error is slightly less than 1% (for $\omega_0 = 0.9$ and $\theta_0 = 0$). Thus, K_d can be well represented by Eq. (6.28) with $C_{K2} = C_{K3} = 0$, i.e.,

$$u_{0w} \frac{K_d}{c} = 1 + C_{K1}[\omega_0(1 - P_{ru})]. \quad (6.29)$$

In the earlier simulations L_u/E_d was evaluated by using the known $E_d = u_0 F_0$. But in the presence of the refracting interface, and in the absence of simulations, we only know E_d^+ , the irradiance incident on the water surface from above. So, if we wanted to use Eq. (6.27) to obtain L_u alone from L_u/E_d in the presence of the water surface, we would need to be able to determine E_d (which is unknown) from E_d^+ , which, for the solar beam, is just $u_0 F_0$. This can be effected with high accuracy in the following manner. When a beam of irradiance E_d^+ is incident on the surface, the Fresnel-transmitted irradiance is $E_d^t = t_f(u_0 \rightarrow u_{0w})E_d^+$, where t_f is the Fresnel transmittance. Some of this downwelling irradiance, $R(u_{0w})E_d^t$, is backscattered (by the medium below) into the upwelling field. A fraction, r , of this backscattered irradiance is reflected from the interface, augmenting the downwelling irradiance by $rR(u_{0w})E_d^t$. Again, a portion of this is reflected upward and back downward augmenting the irradiance by $rR(u_{0w})r'R'E_d^t$, where r' and R' are used for the second reflection because the radiance distribution of the light fields change from the first to the second reflection. Continuing the process, we have¹⁴

$$E_d = t_f E_d^+ (1 + rR(u_{0w}) + rR(u_{0w})r'R' + rR(u_{0w})r'R''R'' + \dots).$$

Now, noting that the values of R are usually < 0.1 and $r \sim 0.5$, the series will converge rapidly, i.e., the terms in R' and R'' will not be very important. Thus, we simplify the result by letting $R' = R'' = R(u_{0w})$, etc., and $r' = r'' = r$, etc. Then the series is just a geometric series and

$$E_d = \frac{t_f E_d^+}{1 - rR(u_{0w})}. \quad (6.30)$$

¹⁴It might seem troubling that $E_d > t_f E_d^+$, i.e., greater than the transmitted irradiance entering the water! However, the photons in each term of the series entered the water at *times earlier* than the photons contribution to E_d^+ .

If we further approximate the upwelling field as totally diffuse, i.e., the upward radiance is independent of direction, then

$$r = \frac{\int_0^1 \mu(1 - t_f(\mu)) d\mu}{\int_0^1 \mu d\mu},$$

which for $m = 1.333$ results in $r \approx 0.48$. Comparison with E_d from Eq. (6.30) with that obtained from “exact” radiative transfer simulations show that Eq. (6.30) is satisfied to within a small fraction of one percent. Thus, the error in E_d determined from E_d^+ is not a significant source of error in estimating L_u from the IOPs using

$$L_u = \frac{t_f E_d^+}{1 - rR(u_{0w})} \left(\frac{L_u}{E_d} \right), \quad (6.31)$$

where the quantity in the parenthesis is determined from Eq. 6.27 with u_{0w} replacing u_0 .

6.4.4.2 A Wind-Roughened Refracting Interface

When the water surface is roughened by the wind, the solar beam is no longer unidirectional in the water, i.e., it is no longer propagating in the direction specified by a single value of u_{0w} . Photons are propagating in a range of directions close to u_{0w} with a spread determined by the Cox-Munk surface slope distribution and variance σ^2 (Chapter 7). This deviation of the incident radiance (beneath the surface) is clearly expected to have some influence on the AOPs, however, examination of Figures 6.8-6.12 suggest that the influence is small except at very large solar zenith angles. A similar effect is produced by the presence of the atmosphere above the interface, i.e., the presence of sky light, as well as a wind-roughened surface, results in the incident radiance (beneath the surface) propagating mostly in a range of directions close to u_{0w} . The atmospheric effect is discussed in the next section.

To examine that surface roughness effect, we have computed the AOPs R , K_d , and L_u/E_d for $\theta_{0w} = 0, 10^\circ, 20^\circ, 30^\circ$, and 40° , ($\theta_0 = 0, 13.38^\circ, 27.12^\circ, 41.68^\circ, 58.96^\circ$), with $0.1 \leq \omega_0 \leq 0.9$ with $\sigma = 0.2$ (wind speed of about 3.7 m/s) rather than zero, using the rough-surface modification to the Monte Carlo procedure developed in Chapter 7. In the case of L_u/E_d and R , the average ratio of the quantity with $\sigma = 0.2$ to that with $\sigma = 0$ (flat surface) differs from unity by less than 0.2% and 0.7%, respectively (the *maximum* difference from unity is 1.2% and 2.9%, respectively, for $\theta_0 = 58.9^\circ$). Thus, these two AOPs are almost unaffected by surface roughness.

In the case of K_d , there are more significant differences between the flat- and the rough-surface cases; however, they are easily related to the structure of the in-water light field. We have seen in the last Section that $u_{0w}K_d/c$ is directly related to the IOPs (Eq. (6.28)).

We need to find a modification that relates K_d to the IOPs in the presence of a rough surface. The obvious candidate is a replacement for u_{0w} . The simplest description of the geometric structure of the *incident* light field is its average cosine, i.e., the ratio E_d/E_{0d} , where both quantities refer only to photons that are *transmitted* through the surface. We call this quantity $\langle u_{0w} \rangle_0$, where the second subscript “0” means it is for a medium with $\omega_0 = 0$, i.e., a totally absorbing water body.¹⁵ If the surface is flat, this reduces to u_{0w} . The computation of the effect of roughness on $\langle u_{0w} \rangle_0$ is carried out by simulating the radiative transfer and setting the IOPs so $\omega_0 = 0$. Figure 6.30 provides $D_0 = 1/\langle u_{0w} \rangle_0$ as a function of θ_0 and σ resulting from such simulations. For each value of θ_0 we have fit the corresponding value of D_0 to a polynomial in σ :

$$D_0 = C_{D0} + C_{D1}\sigma + C_{D2}\sigma^2, \quad (6.32)$$

and the results are provided in Table 6.2. In turn, the coefficients in Table 6.2 can be

Table 6.2: Expansion coefficients in Eq. (6.32) as a function of θ_0 derived from simulations.

θ_0	C_{D0}	C_{D1}	C_{D2}
0	1.000	0	0.0300
13.38	1.015	0.0005	0.0350
27.12	1.064	0.0004	0.0625
41.68	1.154	0.0022	0.1625
58.96	1.305	0.0049	0.6525
70.49	1.415	0.0885	1.4150

reasonable well fit to functions of u_{0w} for $\theta_0 \leq 60^\circ$:

$$\begin{aligned} C_{D0} &= 1/u_{0w}, \\ C_{D1} &\approx 0.0155(C_{D0} - 1), \\ C_{D2} &\approx 0.03 - 0.1831(C_{D0} - 1) + 7.268(C_{D0} - 1)^2. \end{aligned} \quad (6.33)$$

The average ratio of $\langle u_{0w} \rangle_0 K_d/c$ with $\sigma = 0.2$ to that with $\sigma = 0$ computed using these equations for D_0 is 0.9957, and the largest deviation from unity is 0.9935 for the simulations described in this section. Thus, Eq. (6.29) with u_{0w} replaced by $\langle u_{0w} \rangle_0$ remains valid to a high degree of accuracy, utilizing Eq. (6.32) and Eq. (6.33) to estimate $\langle u_{0w} \rangle_0$.

The E_d used in these simulations was the actual value beneath the surface, determined from the Monte Carlo results. Thus, one might also ask if a formula similar to Eq. (6.30) can be used to estimate E_d from E_d^+ . The answer is yes: the ratio E_d/E_d^+ for $\sigma = 0$ and

¹⁵When the waterbody is totally absorbing, i.e., there is no scattering, the geometrical structure of the light field *just beneath* the surface is that of the *transmitted* radiance alone.

$\sigma = 0.2$, for all of the cases examined in this section, differed by an average of 0.15% and a maximum of 0.47%, with the difference generally increasing with increasing θ_0 .

In sum, we see that the influence of surface roughness (due to the wind) is small or easily taken into account for $\theta_0 \lesssim 60^\circ$. For larger solar zenith angles the effects of roughness become larger (Figures 6.8-6.12) and the effects may not be simply explained in a quantitative manner. It should be noted that the modifications to the AOP-IOP relationships required a model of the surface roughness, e.g., the Cox-Munk model used here to relate the surface slope distribution to the wind speed. Large solar zenith angles depend on such models being accurate for very large surface slopes, which is highly questionable. For this reason we have limited our discussion to $\theta_0 < 60^\circ$ (with the exception of Figure 6.30 and Table 6.2).

6.4.4.3 An Atmosphere Above a Refracting Interface

To examine the influence of the atmosphere on L_u/E_d , we used an aerosol model typical of that for a clear maritime aerosol. The aerosol optical thicknesses $\tau_a(\lambda)$ were 0.126, 0.107, and 0.094, respectively at, $\lambda = 400, 500$, and 600 nm (0.10 at 550 nm). At these wavelengths, the Rayleigh optical thicknesses $\tau_r(\lambda)$ were 0.364, 0.145, and 0.069. Radiative transfer simulations were run for all combinations of $\theta_{0w} = 0, 10^\circ, 20^\circ, 30^\circ$, and 40° ($\theta_0 = 0, 13.38^\circ, 27.12^\circ, 41.68^\circ, 58.96^\circ$), with $\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9$, the Petzmas phase function, and $\lambda = 400, 500$, and 600 nm. Surface roughness was characterized by $\sigma = 0$ and 0.2 . In addition, similar computations, but *without* the atmosphere, were carried out as well. The values of the ratio $(L_u/E_d)_{\text{Atm}} \div (L_u/E_d)_{\text{NoAtm}}$, where the subscripts “Atm” and “NoAtm” mean, respectively *with* and *without* the atmosphere (both *with* a refracting water surface), are shown in Figure 6.31 for $\lambda = 400$ nm.. We see that this ratio is always between 3% above unity and 1.5% below unity depending on ω_0, θ_0 and σ . It deviates more from unity at 400 nm than for longer wavelengths as shown in Table 6.3, which also provides information on similar ratios formed with R and E_d/E_d^+ . One notes that the average effect of the atmosphere is to increase L_u/E_d slightly above unity and to decrease E_d/E_d^+ slightly below unity. This suggests that the effect of the atmosphere on L_u/E_d^+ should be less than the effect on either of the above quantities. In fact, the average value of this ratio over those cases in Table 6.3 is 0.9997 at 400 nm. Since the remote sensing reflectance is given by $R_{rs} = (L_u/E_d^+)t_L$, and t_L is not affected by the presence of the atmosphere, one expects that the absence or presence of the atmosphere has very little effect on R_{rs} .

Table 6.3: Values for the ratios (with the atmosphere \div without the atmosphere) of the indicated quantities for all combinations of $\theta_{0w} = 0, 10^\circ, 20^\circ, 30^\circ$, and 40° , with $\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9$, the Petzmas phase function, $\sigma = 0$ and 0.2 , and $\lambda = 400, 500$, and 600 nm. “Ave” is the average deviation from unity and “Max” is the maximum deviation from unity.

λ (nm)	L_u/E_d Ave	L_u/E_d Max	R Ave	R Max	E_d/E_d^+ Ave	E_d/E_d^+ Max
400	1.012	1.029	1.042	1.082	0.988	0.982
500	1.004	1.016	1.021	1.039	0.993	0.986
600	1.004	1.010	1.013	1.039	0.995	0.990

6.4.4.4 Estimation of E_d^+ and $\langle u_{0w} \rangle_0$ with a Rough Surface, Sun and Sky

If we want to use the QSSA to evaluate L_u by itself, as discussed before, we need E_d . The estimate of E_d can be made utilizing Eq. (6.30). Table 6.3 shows that to within about 1% or less, $(E_d/E_d^+)_{\text{Atm}} = (E_d/E_d^+)_{\text{No Atm}}$. Thus, the problem of estimating E_d in the presence of the atmosphere and rough refracting surface reduces to that of estimating E_d^+ . Also, we will need to estimate $\langle u_{0w} \rangle_0$, as in the spirit of the QSSA, we will replace $QSSA'_L(u_{0w})$ by $QSSA'_L(\langle u_{0w} \rangle_0)$. In addition, we will need both of these quantities for further application of the QSSA in the computation of $K_d(0)$ in the presence of the surface, Sun, and sky, described at the end of this section.

The problem of finding E_d^+ , the solar irradiance transmitted by the atmosphere and incident on the water surface is easily effected by applying the single-scattering approximation to the atmosphere. The first-order solution for a atmospheric layer of optical thickness τ is given by an equation similar to Eq. (6.15).

$$E_d^+(\tau) = \int_0^{2\pi} d\phi \int_0^1 u L_d(\tau, u, \phi) du = F_0 u_0 \exp[-\tau/u_0] + \frac{\omega_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \frac{u P(u_0 \rightarrow u, \phi_0 \rightarrow \phi)}{1 - u/u_0} \times (\exp[-\tau/u_0] - \exp[-\tau/u]) du.$$

Here P is the phase function for the atmosphere, composed of Rayleigh and aerosol scattering. When $\tau \ll 1$, the exponentials can be expanded we have (the single-scattering approximation)¹⁶

$$E_d^+(\tau) \approx F_0 u_0 (1 - \tau/u_0) + \frac{\omega_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \tau P(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du.$$

¹⁶This ignores Fresnel reflection from the interface. A derivation including (flat) surface Fresnel reflection is provided in Chapter 9, Section 9.5, Eq. 9.37.

Note that

$$\omega_0 P\tau = \omega_0 Pcz = \omega_r P_r c_r z + \omega_a P_a c_a z = \omega_r P_r \tau_r + \omega_a P_a \tau_a$$

where the subscripts “ r ” and “ a ” refer to “Rayleigh” and “aerosol” scattering by the atmosphere, respectively. Inserting this into the integral and recalling that P_r is symmetric in the forward and backward directions,¹⁷ we have

$$E_d^+(\tau) \approx F_0 u_0 (1 - \tau/u_0) + \omega_r F_0 \tau_r / 2 + \frac{\omega_0 F_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 \tau_a P_a(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du.$$

Writing $\int_0^{2\pi} d\phi \int_0^1 P_a(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du \triangleq 4\pi I_a(u_0)$ and noting that $\omega_r = 1$,¹⁸

$$E_d^+(\tau) \approx F_0 u_0 (1 - \tau/u_0) + \omega_r F_0 \tau_r / 2 + \omega_0 F_0 \tau_a I_a = F_0 u_0 (1 - \tau_r / 2u_0 - \tau_a (1 - \omega_a I_a) / u_0).$$

We can usually drop the aerosol term in this expression because typically, $\tau_a \lesssim 0.1$, $\omega_a \approx 1$ and I_a is close to 1 (aerosols scatter strongly in the forward direction). Then,

$$E_d^+(\tau) \approx F_0 u_0 \left(1 - \frac{\tau_r}{2u_0}\right) \approx F_0 u_0 \exp\left(-\frac{\tau_r}{2u_0}\right).$$

However, ignoring the aerosol altogether can lead to error of a few % in E_d^+ . We have found, more or less by trial and error, that an acceptable compromise when the aerosol optical thickness at 550 nm is ≤ 0.1 is

$$E_d^+(\lambda) = F_0 u_0(\lambda) \exp\left(-\frac{\tau_r(\lambda)}{x(\lambda)u_0}\right), \quad (6.34)$$

where the factor $x(\lambda) = 2.3, 2.2$, and 2.1 for $\lambda = 400, 500$, and 600 , respectively. The error in Eq. (6.34) for the simulations carried out in the last section (6.4.4.3) is generally $< 0.5\%$, except for 400 nm with $\theta_0 \approx 60^\circ$ for which the error was ≈ -1.8 to -2.3% depending on ω_0 in the water (larger error was for larger ω_0 , as photons scattered *out* of the water can backscatter in the atmosphere and contribute to E_d^+).

To estimate $\langle u_{0w} \rangle_0$ we note that there are two sources of irradiance incident on the water surface: direct sunlight and sky light. For simplicity, we assume the sky light is totally diffuse (sky radiance is assumed independent of viewing direction from the surface). As in an earlier section, we prefer to work with $D_0 = 1/\langle u_{0w} \rangle_0$. If there are N sources of irradiance incident on the surface (each indexed by “ i ”), then

$$D_0 = \frac{\sum_{i=1}^N (E_{0d})^i}{\sum_{i=1}^N (E_d)^i} = \frac{\sum_{i=1}^N D_0^i (E_d)^i}{\sum_{i=1}^N (E_d)^i} = \frac{\sum_{i=1}^N D_0^i (E_d)^i}{E_d^{\text{Total}}} = \sum_{i=1}^N f_i D_0^i,$$

¹⁷Remember in this section P_r is the Rayleigh phase function, while in the rest of this chapter it is the “remaining” part of the phase function after the Dirac delta function has been subtracted for the QSSA.

¹⁸ $I_a(u_0)$ is the same as the quantity $p_a^{--}(|u|)$ in Chapter 9 and P_d in Chapter 2. An example of its variation with θ_0 is shown in Figure 9.15 (with θ_0 replacing θ_{ex}).

where f_i is the fraction of the incident irradiance produced by source i . In our case (two sources),

$$D_0 = fD_0^S + (1 - f)D_0^\oplus, \quad (6.35)$$

where the superscripts S and \oplus refer to the sky and the sun, respectively. Now, D_0^S and D_0^\oplus can be determined once-and-for-all using a model for the surface roughness, as D_0^S depends only on the roughness (σ) and D_0^\oplus depends on θ_0 and σ (see Section 6.4.4.2 for the D_0 dependence on θ_0 and σ). In the case of D_0^S , Table 6.4 provides the dependence on σ . Regressions using the values provided in Table 4 yield,

$$\begin{aligned} D_0^S &\approx 1.1876 + 0.1505\sigma + 0.675\sigma^2 \\ T^S &\approx 0.933 + 0.04\sigma, \end{aligned} \quad (6.36)$$

which can be used for intermediate values of σ . If we could find f in Eq. (6.35) we

Table 6.4: Simulated values of D_0^S and the irradiance transmittance of the water surface T^S for totally diffuse sky light as a function of the surface roughness parameter σ .

σ	D_0^S	T^S
0.0	1.188	0.933
0.1	1.208	0.936
0.2	1.246	0.943
0.3	1.293	0.949

would be able to estimate D_0 , and thus, $\langle u_{0w} \rangle_0$. The quantity f is the sky fraction of the irradiance just *beneath* the interface. Its estimate requires the sky fraction just *above* the interface. Call this latter fraction g , i.e., $g \triangleq E_d^{S+}/E_d^+$, $1 - g = E_d^{\oplus+}/E_d^+$. Now earlier in this section we found E_d^+ could be estimated from

$$E_d^+ = u_0 F_0 \exp \left[-\frac{\tau_r(\lambda)}{x(\lambda)u_0} \right].$$

But what about $E_d^{\oplus+}$? This is the solar irradiance modified by the *direct* transmittance of the atmosphere: $\exp[-(\tau_r + \tau_a)/u_0]$. However, as most photons scattered by the aerosol are scattered through small angles, and thus may be considered to be still part of the collimated solar beam, the τ_a term in the direct transmittance effectively causes the attenuation to be over estimated and should be ignored. Therefore, in the same spirit as leading to Eq. (6.34), we ignore the aerosol and take

$$E_d^{\oplus+} = u_0 F_0 \exp \left[-\frac{\tau_r(\lambda)}{u_0} \right],$$

so

$$g(\lambda) = 1 - \exp \left[-\frac{\tau_r(\lambda)}{u_0} \left(1 - \frac{1}{x(\lambda)} \right) \right]. \quad (6.37)$$

Note that this will be adequate only for small values of τ_a . Now,

$$f = \frac{E_d^S}{E_d} = \frac{E_d^{S+} T^S}{E_d^{S+} T^S + E_d^{\oplus+} T^{\oplus}} = \frac{g T^S}{g T^S + (1 - g) T^{\oplus}}, \quad (6.38)$$

where T^{\oplus} is the irradiance transmittance of the direct solar beam. Recalling that $T^{\oplus} \approx t_f(\theta_0)$, where t_f is the Fresnel reflectance for a *flat* surface in the absence of the atmosphere, these relationships along with Tables 6.2 and 6.4 (or the associated regressions) provide estimates of f and D_0 .

The validity of these estimates of D_0 is tested with Monte Carlo simulations of D_0 obtained for $\theta_0 = 0, 13.38^\circ, 27.12^\circ, 41.68^\circ$ and 58.96° , $\lambda = 400, 500$, and 600 nm, $\tau_a(550) = 0, 0.1$, and 0.2 , and $\sigma = 0, 0.1, 0.2$, and 0.3 . The results are provided in Figure 6.32. All of the estimated values are too low, but the average error is only -0.23% and the maximum error -1.23% . Thus, D_0 , and the associated $\langle u_{0w} \rangle_0$ can be very robustly estimated in the presence of the atmosphere and a wind-roughened surface using Eqs. (6.35)-(6.38).

This completes computation of all of the ingredients required to estimate L_u in the presence of the atmosphere (and a wind-roughened water surface) using Eq. (6.27) with u_{0w} replaced by $\langle u_{0w} \rangle_0$. But how well can we compute L_u/E_d in this more realistic setting? Figure 6.33 compares exact computations of L_u/E_d with predictions of the QSSA in the presence of a flat water surface and the atmosphere using Eq. (6.27) with u_{0w} replaced by $\langle u_{0w} \rangle_0$ and $C_L = 1$. These computations are for a wavelength of 400 nm (for the maximum atmospheric effect), an aerosol optical thickness of 0.10 at 550 nm, and solar zenith angles of $10^\circ, 37^\circ$, and 60° . The particle scattering phase function is either Petzold ($B_p = 0.0181$) or modified Petzold with $B_p = 0.0020$. The average (unsigned) error is 2.25% for $B_p = 0.0020$ and 7.27 % for $B_p = 0.0181$.¹⁹ Computations for more values of B_p shows that much of the bias can be removed by replacing $C_L = 1$ by $C_L = 0.99 + 6.9B_p$. The results are similar to those in Figure 6.26, which has neither a water surface nor an atmosphere. This suggests that, at least as far as L_u/E_d is concerned, we have adequately accounted for the presence of the water surface and the atmosphere and, as we can now compute E_d , we can estimate L_u using these simple relationships between L_u and the IOPs.

What about the effects of the environment on $K_d(0)$? Using the same simulations including the atmosphere as in the discussion of D_0 above, we computed the value of $K_d(0)$ and plotted the results in a manner similar to Figure 6.29. This is provided in Figure 6.34 in which u_{0w} from Figure 6.29 is replaced by $\langle u_{0w} \rangle_0$. The line on the figure is Eq. (6.28) with the same values of C_{Ki} as in Figure 6.29, i.e., the linear fit. The average (unsigned) error in the linear expansion is 2.31% and the maximum difference is 3.69%.

¹⁹We also examined the quality of the QSSA if u_{0w} is not replaced by $\langle u_{0w} \rangle_0$. On the scale of Figure 6.33 the results are almost indistinguishable; however they are slightly better with the average error of 1.57% for $B_p = 0.0020$ and 6.25 % for $B_p = 0.0181$.

(If a new linear regression is carried out the result is $C_{K1} = 1.0139$, for which the average error is reduced to 0.44% and the maximum difference reduced to 2.36%.) Thus, it would seem that the main effect of the environmental setting on the QSSA expression for $K_d(0)$ is the replacement of u_{0w} with $\langle u_{0w} \rangle_0$.

6.4.4.5 Summary of Environmental Effects on QSSA

We have shown that factors affecting the radiative transfer in the ocean-atmosphere system, but were not included in the original QSSA, namely a flat or rough refracting water surface and an atmosphere above the water, can be handled with the simple replacement of u_0 by $\langle u_{0w} \rangle_0$. When this is done, the accuracy of the QSSA is virtually the same as in their absence. If needed, the computation of E_d from E_d^+ can be effected with high accuracy as well. These conclusions apply to the QSSA-computation of both L_u/E_d and $K_d(0)$. For convenience, we provide the complete set of equations necessary to effect the computation of L_u/E_d and L_u in Figure 6.35, with $C_L = 1$. (The estimate can be improved by using $C_L = 0.99 + 6.9B_p$.)

6.4.5 Depth-Averaged $K_d(z)$

One of the important uses of the QSSA is the estimate of the IOP $[1 - \omega_0(1 - P_{ru})]$ from K_d . As we see from Figure 6.15, P_{ru} is small compared to unity, so $\langle u_{0w} \rangle_0 K_d \approx c(1 - \omega_0) = c - b = a$, and the estimated IOP is close to the absorption coefficient. However, although we have the very accurate formula (Eq. (6.29)) relating $K_d(z)$, determined just beneath the surface ($z = 0$), to the IOPs, it is often useless in practice because $K_d(0)$ is nearly impossible to determine experimentally due to the presence of strong fluctuations in $E_d(z)$ resulting from the motion of the wind-roughened surface. Fortunately, these fluctuations decrease rapidly with increasing z . We have seen earlier that $E_d(0)$, i.e., just beneath the surface, can be accurately estimated from $E_d^+(0)$, the downwelling irradiance just *above* the surface (Eq. (6.30)). So, given experimental measurement of $E_d^+(0)$, it is possible to compute an average K_d between the surface and some depth z' , where the wave-induced irradiance fluctuations are unimportant, i.e.,

$$\langle K_d(0 \rightarrow z') \rangle \triangleq -\ell n \left[\frac{E_d(z')}{E_d(0)} \right] = \frac{1}{z'} \int_0^{z'} K_d(z) dz.$$

Can this be related in a simple manner to the IOPs? The obvious choice for such a relationship would be to try to relate $\langle K_d(0 \rightarrow z') \rangle$ to $\omega_0(1 - P_{ru})$ in the manner of Eq. (6.28). Figure 6.36 compares computations of $\langle K_d(0 \rightarrow z') \rangle$, for z' just beneath the surface and z' equal to the depth at which E_d falls to $\exp(-3)$ of its value just beneath

the surface, to $\omega_0(1 - P_{ru})$ for a water body with a *flat* refracting interface in the *absence* of the atmosphere. Clearly, the strong linear relationship between $K_d(0)$ and $\omega_0(1 - P_{ru})$ still holds for $\langle K_d(0 \rightarrow z') \rangle$, although the slope of the relationship is not the same. Thus, we seek the least-squares fit of the computations to determine $\langle C_K \rangle_X$ in

$$u_{0w} \frac{\langle K_d \rangle_X}{c} = 1 - \langle C_K \rangle_X \omega_0(1 - P_{ru}), \quad (6.39)$$

where $\langle K_d(0 \rightarrow z') \rangle \triangleq \langle K_d \rangle_X$ and z' is the depth at which $E_d(z')/E_d(0) = \exp(-X)$. Table 6.5 provides $\langle C_K \rangle_X$ for various values of X , along with the resulting average (unsigned) and maximum % errors in $\langle K_d \rangle_X$. In all cases, when $X \geq 1$ the maximum error is for $\omega_0 = 0.9$. In general the departure of the computed $\langle K_d \rangle_X$ from its true value increases with increasing ω_0 and with increasing X . However, one should note that (for a given value of c) larger values of ω_0 result in *smaller* values of $K_d(z)$, i.e., the irradiance decays at a smaller rate. Thus, in field applications one could expect the use a smaller value of X for larger ω_0 's to avoid the surface-induced fluctuations. For example, for irradiance data off Hawaii (Figure 6.1) if one assumes that the wave-induced fluctuations are negligible below $z = 20$ m, then one could take $X \sim 0.5$ in the blue, ~ 1.5 near 580 nm, and ~ 4 near 600 nm, to avoid such fluctuations. In this case, to retrieve the IOPs — $c\omega_0(1 - P_{ru}) = b(1 - P_{ru})$ — from the irradiance measurements using Eq. (6.39) the more-accurate formulas would

Table 6.5: Value of $\langle C_K \rangle_X$ for various values of X , and the average (unsigned) and maximum error in $\langle K_d \rangle_X$ determined using Eq. (6.39).

X	$\langle C_K \rangle_X$	Ave (%)	Max (%)
0	0.9892	0.25	0.95
1	0.9639	0.44	4.04
2	0.9519	1.07	9.60
3	0.9447	1.72	14.22
4	0.9402	2.29	17.57

be used with the smaller values of K_d and the less-accurate for the larger values of K_d , but the less-accurate formulas are actually quite accurate for larger values of K_d .

These formulas result from a simple extension of $QSSA_K$ at the surface to the water column. However, the surface was flat and there was no atmosphere. In either case, the irradiance entering the water will be propagating in a range of directions around the direction specified by θ_{0w} . Thus the generalization of Eq. (6.39) to more realistic settings is

$$\langle u_{0w} \rangle_0 \frac{\langle K_d \rangle_X}{c} = 1 - \langle C_K \rangle_X \omega_0(1 - P_{ru}), \quad (6.40)$$

where P_{ru} is evaluated using $\langle u_{0w} \rangle_0$ in place of u_{0w} . We extended the radiative transfer computations of $K_d(0 \rightarrow z')$ used in computing Table 6.5 to three other situations: (1)

a rough surface characterized by the Cox-Munk omnidirectional wind with $\sigma = 0.2$, but no atmosphere; (2) a flat surface but with an atmosphere characteristic of a maritime atmosphere at $\lambda = 400$ nm; and (3) the combination of (1) and (2). In each set of simulations we used all combinations of $\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85$, and 0.9 , and $\theta_{0w} = 0, 10^\circ, 20^\circ, 30^\circ$, and 40° . In the case of the atmosphere, the aerosol optical depth is 0.1 at 550 nm which is typical; however, at the wavelength used in these calculations (400 nm) the sky radiance is dominated by Rayleigh scattering. These simulations are pooled

Table 6.6: Value of $\langle C_K \rangle_X$ for various values of X , and the average (unsigned) and maximum error in $\langle K_d \rangle_X$ determined using Eq. (6.40).

X	$\langle C_K \rangle_X$	Ave (%)	Max (%)
0	0.9858	0.65	3.73
1	0.9655	0.49	4.48
2	0.9550	1.16	9.56
3	0.9491	1.80	13.38
4	0.9455	2.34	16.02

with those used for Table 6.5 and again used with Eq. (6.40) to determine $\langle C_K \rangle_X$ via least-squares analysis. The results of the analysis are presented in Table 6.6. Comparison of these results with those in Table 6.5 suggests that the presence of the atmosphere and/or surface waves does not significantly affect the accuracy of Eq. (6.39) as long as $\langle u_{0w} \rangle_0$ replaces u_{0w} .

How does the inclusion of Rayleigh scattering by the water itself influence these results? We examined full Monte Carlo simulations yielding $1 - \langle u_{0w} \rangle_0 \langle K_d \rangle_X / c$ as a function of $\omega_0(1 - P_{ru})$ for media in which the particle phase function was either the Petzold Phase-T ($B_p = 0.0181$) or the modified Petzold ($B_p = 0.0020$). In these computations $\theta_0 = 10^\circ, 37^\circ$ and 60° , with $\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85$, and 0.9 . The total backscattering probability varied from 0.0030 to 0.12 . As before, the atmosphere was maritime with an aerosol optical thickness of 0.1 at 550 nm. The water surface was flat and the computations were carried out at a wavelength of 400 nm. The results are displayed on Figure 6.37, and show that for the smaller B_p the results fit Eq. (6.40) considerably better than those for the larger, which show systematic departures from $\langle C_K \rangle_X = 1$. Table 6.7 provides the linear-regression estimate of $\langle C_K \rangle_X$. Note the closeness of $\langle C_K \rangle_X$ to unity and the near independence on X for $B_p = 0.0020$, and the closeness of the $\langle C_K \rangle_X$ values for $B_p = 0.0181$ to those in Table 6.6, for which there is no contribution to scattering by the water itself. The average (absolute) error in $1 - \langle u_{0w} \rangle_0 \langle K_d \rangle_X / c$ for these regressions is of the order of 1.5 – 2.5% with the maximum error $\sim 10\%$. The errors are of similar magnitude for both values of B_p . These examples again underscore the fact that the QSSA becomes more accurate as the phase function becomes more peaked in the forward direction.

6.4.6 Angular Distribution of L_u (BRDF Effect)

The discussion of upwelling radiance thus far has been limited to that propagating toward the zenith. However, in a general remote sensing setting, the sensor views the radiance exiting the water propagating toward the sensor in almost any conceivable viewing direction (θ_v, ϕ_v) except those close to the position of specular reflection of direct solar radiation (i.e., in the Sun glitter pattern, described in Chapter 7). Because of internal reflection,

Table 6.7: Value of $\langle C_K \rangle_X$ for various values of X when water scattering is added to particle scattering. B_p is the particle backscattering coefficient. As in Table 6.6, “Ave” and “Max” are the average (unsigned) and maximum error in $\langle K_d \rangle_X$ determined using Eq. (6.40).

X	$B_p = 0.0020$	Ave (%)	Max(%)	$B_p = 0.0181$	Ave (%)	Max(%)
0	1.0059	1.88	11.28	0.9778	1.79	8.33
1	1.0029	1.94	11.88	0.9521	1.82	10.62
2	1.0059	1.36	9.02	0.9449	1.69	12.77
3	1.0083	1.06	7.16	0.9420	1.71	14.38
4	1.0103	1.03	5.95	0.9404	1.97	15.55

only $L_u(0, \theta_{vw}, \phi_{vw})$, where $\sin \theta_v = m_w \sin \theta_{vw}$ and $\phi_v = \phi_{vw}$, with $0 \leq \theta_{vw} \leq 48.6^\circ$ (for a flat water surface) can be observed from above the water. (Note: θ_v and θ_{vw} are measured relative to the *upward* normal to the water surface.) Thus, rather than $L_u(0, 0, \phi_{vw})/E_d(0)$ in terms of the IOPs, in general we need the full radiance distribution $L_u(0, \theta_{vw}, \phi_{vw})/E_d(0)$. The variation of L_u with θ_{vw} and ϕ_{vw} is called the “BRDF Effect.”²⁰ The success of Eq. (6.27) in relating $L_u(0, 0, \phi_{vw})/E_d(0)$ to the IOPs, and its extension to the case of a refracting interface, suggests its utility could be extended to the full radiance distribution by replacing $QSSA'_L(u_0)$ by

$$QSSA'_L(u_{0w}, \mu_{vw}, \phi_{vw}) \triangleq \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \frac{P_r(u_{0w} \rightarrow -\mu_{vw}, \phi_{0w} \rightarrow \phi_{vw})}{u_{0w} + \mu_{vw}}, \quad (6.41)$$

where $u_{0w} = \cos \theta_{0w}$ and $\mu_{vw} = \cos \theta_{vw}$.

Testing this relationship over the full range of parameter variation is well beyond the scope of this work; however, we have already provided an example that sheds some light on the accuracy that can be expected using Eq. (6.41) to determine the dependence of L_u on viewing angle. In Figure 6.22 we provided the normalized L_u/E_d (nadir viewing) as a function of the solar zenith angle (for the case of no atmosphere and water refractive index of unity). Reciprocity can be used to relate this to the angular distribution of L_u when the sun is at the zenith. Let $L_u(\hat{\xi}_1 \rightarrow \hat{\xi}_2)$ be the radiance leaving the medium in a

²⁰See Chapter 2 for a discussion of the BRDF.

direction $\hat{\xi}_2$ when the surface is illuminated by the solar beam (in a black sky) propagating in a direction $\hat{\xi}_1$, and $E_d(\hat{\xi}_1)$ be the irradiance into the medium (by the beam) under the same circumstances. Then reciprocity requires

$$\frac{L_u(\hat{\xi}_1 \rightarrow \hat{\xi}_2)}{E_d(\hat{\xi}_1)} = \frac{L_u(-\hat{\xi}_2 \rightarrow -\hat{\xi}_1)}{E_d(-\hat{\xi}_2)}.$$

It is easy to verify that Eq. (6.41) satisfies this relationship by simply interchanging the symbols $u_{0w} \leftrightarrow \mu_{0v}$ and $\phi_{0w} \leftrightarrow -\phi_{0v}$.²¹ Since $E_d(\hat{\xi}_a) = |\hat{\xi}_a \bullet \hat{n}| F_0$, where \hat{n} is the surface normal (directed *out* of the medium),

$$\frac{L_u(\hat{\xi}_1 \rightarrow \hat{\xi}_2)}{|\hat{\xi}_1 \bullet \hat{n}|} = \frac{L_u(-\hat{\xi}_2 \rightarrow -\hat{\xi}_1)}{|\hat{\xi}_2 \bullet \hat{n}|}.$$

So, if $\hat{\xi}_2$ is in the direction of the zenith (the direction for the computations in Figure 6.22), i.e., $-\hat{\xi}_2 = \hat{n}$,

$$\frac{L_u(\hat{\xi}_1 \rightarrow \hat{n})}{|\hat{\xi}_1 \bullet \hat{n}|} = L_u(-\hat{n} \rightarrow -\hat{\xi}_1).$$

The radiance on the left-hand-side of this equation is that which is plotted in Figure 6.22, but in the figure it is normalized by dividing by $L_u(-\hat{n} \rightarrow \hat{n})$. Therefore,

$$\frac{L_u(-\hat{n} \rightarrow -\hat{\xi}_1)}{L_u(-\hat{n} \rightarrow \hat{n})} = \frac{1}{|\hat{\xi}_1 \bullet \hat{n}|} \left[\frac{L_u(\hat{\xi}_1 \rightarrow \hat{n})}{L_u(-\hat{n} \rightarrow \hat{n})} \right].$$

The term in the square brackets is the quantity actually plotted in Figure 6.22. This figure can therefore be converted to a plot to the normalized angular distribution of the radiance exiting the medium when the sun is at the zenith (solar radiation propagating in the direction $-\hat{n}$), simply by replacing the θ_0 by the viewing angle (θ_v) and dividing each point by $|\hat{\xi}_1 \bullet \hat{n}| = \cos \theta_v = \mu$. The result is provided in Figure 6.38. It suggests that Eq. (6.41) will reproduce the variation of L_u/E_d with viewing angle – sun angle variations to within $\sim 20\%$ for the Petzold phase function and $\sim 10\%$ for the Petzmas phase function. The variation prediction is best for smaller values of ω_0 , but the performance of Eq. (6.41) is surprisingly good of larger ω_0 values. Since scattering is a diffusive process, the more multiple scattering the less the shape of the phase function in the backward directions affects the shape of the radiance distribution. The QSSA-radiance distribution always exactly reflects the actual phase function (Eq. (6.41)), so it will become progressively less accurate for increasing multiple scattering. Because the shape of the Petzmas phase function is thought to be more representative of the scattering by ocean particles for $\Theta \gtrsim 90^\circ$, and when the scattering by water is added to that of particles the variation in the particle phase function at large Θ is suppressed in the total phase function (Figure 6.24), it is to be

²¹Note that P_r depends on ϕ_{0w} and ϕ_{0v} only through $|\phi_{0w} - \phi_{0v}|$.

expected that Eq. (6.41) will reproduce actual variations L_u/E_d even better than shown in Figure 6.38.

For another example of the ability of Eq. (6.41) to reproduce the angular variation in L_u/E_d we have redrawn Figure 6.26 (left panel) stratifying the results according to θ_0 . The redrawn graphs are shown in Figure 6.39. Some sets of three points are connected by horizontal lines. These points have the same IOPs but different values of θ_0 . The larger values of L_u/E_d correspond to larger values of ω_0 and B . One sees the tendency of connected points (which are for larger values of ω_0 and B) to have roughly the same value of L_u/E_d but different values of $QSSA'_L$. In contrast, for smaller L_u/E_d (smaller ω_0) the tendency of the values is to fall *along* the one-to-one line, and QSSA predicts the correct angular distribution of radiance.²²

6.4.7 Degree of Polarization of L_u

Nearly all of the radiative transfer simulations used in this chapter were carried out using a code that includes the polarization of the radiant light field. However, since most ocean color sensors do not measure polarization (although they may in the future) we have not focussed attention on this aspect of the upwelling radiance. However, for completeness we provide an example of the degree of polarization (DOP) and how it is influenced by multiple scattering. Figure 6.40 shows the DOP of $L_u(0, \theta_{vw}, \phi_{vw})$ for water body with $\omega_0 = 0.2$, i.e., near the single scattering limit, and one with significant multiple scattering: $\omega_0 = 0.9$. In these simulations $\theta_0 = 40^\circ$, the water surface was flat, and the atmosphere included Rayleigh scattering (~ 500 nm) but was aerosol free. The normalized Voss-Fry scattering phase Mueller matrix (Chapter 5) was used along with the Petzold Phase-T phase function to describe the polarized scattering. The general features of the L_u DOP can be understood from the DOP associated with the Voss-Fry phase matrix, i.e., DOP small at scattering angles near $\Theta = 0$ and 180° , and large at Θ near 90° (similar to Rayleigh scattering). The figure shows as expected that for small ω_0 the DOP is a maximum near where the scattering angle Θ (in single scattering) would be 90° , e.g., for $\theta_0 = 40^\circ$, $\theta_{0w} \approx 30^\circ$, so at $\phi_{vw} = 0$ the maximum DOP is at $\theta_{vw} \approx 60^\circ$. As ϕ_{vw} increases to 90° , the maximum DOP shifts to $\theta_{vw} \sim 90^\circ$ preserving $\Theta \approx 90^\circ$. For $\phi_{vw} > 90^\circ$, the scattering angle $\Theta > 90^\circ$ reaching 180° at $\theta_{vw} = 30^\circ$ with $\phi_{vw} = 180^\circ$, where the DOP is very low. For the large value of ω_0 the DOP is decreased in all directions by multiple scattering; however, the general features (maxima and minima) remain in roughly the same angular positions. The maximum DOP decreases from about 0.6 to 0.4 as ω_0 increases from 0.2 to 0.9; however, viewed from above the surface the maxima would not be observed in the positions shown

²²Remember that, although the figures are for nadir radiance at different sun angles, the results can be directly applicable to the distribution of upward radiance with viewing angle when the sun is at the zenith.

in the figure because radiance from these directions would not escape the water. This example provides, at least, some idea of the magnitude of the DOP one might expect in natural waters. This will be of interest when we discuss the vicarious calibration of ocean color sensors.

6.4.8 Summary of QSSA

The AOPs are related to the IOPs through the radiative transfer equation. In general this equation can be solved only numerically, except within the approximation that photons scatter only once: the single scattering approximation (SSA). The QSSA is an approximation derived by replacing the forward portions of the volume scattering function by a Dirac delta function. Its utility is that it can provide analytical expressions for the AOPs that are similar to those derived from the SSA, but in addition it can account (approximately) for the effects of multiple scattering. In fact, we have shown earlier that the QSSA-estimate of R or L_u/E_d can be reasonably accurate for ω_0 as high as 0.7–0.8, and the average error in the QSSA-estimated L_u/E_d can be as low as a few % over a wide range of IOPs. Noting that in the QSSA, the average number of scatterings²³ in a homogeneous medium is $1/(1 - \omega_0 F)$, we see that the QSSA can account for the effects of 3–5, or even more, scatterings quite effectively. It also can provide a good approximation to the shape of the radiance distribution just beneath the sea surface (as long as $\omega_0 \lesssim 0.5$) which directly provides the water-leaving radiance distribution. In addition, it can be modified in a relatively simple manner to account for environmental factors such as a rough water surface and the presence of an atmosphere above the water surface, e.g., through $u_0 \rightarrow u_{0w} \rightarrow \langle u_{0w} \rangle_0$, etc. We shall see in the next section that, because it identified the relevant parameters that determine the AOPs, i.e., a , b_b , and P_r for scattering angles greater than 50° , the QSSA can lead to further analytical/empirical models that produce the AOPs even better.

6.4.9 Alternate Formulations of AOP-IOP Relationships

The QSSA approach to the relationships between AOPs and IOPs has proven reasonably successful in estimating the AOPs of interest in remote sensing given the IOPs (Figures 6.34–6.37), and also in isolating the relevant parameters. Is it possible to improve these estimates short of actually solving the radiative transfer equation? In studies in the early 1970's relating AOPs to IOPs it was found that both the reflectance $R(0)$ and L_u/E_d could

²³Recall from Chapter 2 that the mean number of scatterings $\langle n \rangle$ for L_u is given by $\langle n \rangle = (\omega_0/L)(dL/d\omega_0)$. Application of this to Eq. (6.27) yields the stated result.

be expanded in a power series in the parameter $b_b/(a + b_b)$ or b_b/a , i.e.,

$$R(0) = \sum_{n=1}^3 r_n \left[\frac{b_b}{a + b_b} \right]^n = \sum_{n=1}^3 r_n G^n \quad (6.42)$$

and

$$\frac{L_u}{E_d} = \sum_{n=1}^2 g_n \left[\frac{b_b}{a + b_b} \right]^n = \sum_{n=1}^2 g_n G^n \quad (6.43)$$

and $K_d(0)$ could be expanded in a series

$$\frac{K_d(0)}{D_0} = \sum_{n=1}^3 k_n [a + b_b]^n \quad \text{or} \quad \frac{K_d(0)}{cD_0} = \sum_{n=1}^3 k'_n [1 - \omega_0 F]^n, \quad (6.44)$$

where $F = 1 - B$.²⁴ Note that the solar zenith angle only appears in the equation for K_d , i.e., in D_0 and does *not* occur in the other two equations at all. An example of the fits of exact computations to relationships such as Eqs. (6.42) to (6.44) is provided in Figure 6.41. (Note: the *water surface and the atmosphere at 400 nm*, as described in Section 6.4.4.3, are included in all the results presented in this section.) We will now combine these relationships with the QSSA to develop better AOP-IOP relationships.

6.4.9.1 Hybrid Model for L_u/E_d

Because it is central to remote sensing, we will discuss an alternate formulation for L_u/E_d first. Figure 6.39 shows that for larger ω_0 and/or B , the exact value of L_u/E_d does not depend much on u_0 (although the QSSA value does); however, for small ω_0 , the dependence of L_u/E_d on u_0 can be significant. This suggests the possibility of developing a more accurate *hybrid* model for L_u/E_d as a function of the IOPs: for small ω_0 use the QSSA, and for large ω_0 use an expansion similar to Eq. (6.43). We have tried this using $\omega_0 = 0.7$ as a cutoff, i.e., for $\omega_0 < 0.7$ we use the standard QSSA, and for $\omega_0 \geq 0.7$ we use Eq. (6.43):

$$\begin{aligned} \frac{L_u}{E_d} &= QSSA'_L; & \omega_0 < 0.7, \\ \frac{L_u}{E_d} &= g_1 G + g_2 G^2; & \omega_0 \geq 0.7. \end{aligned} \quad (6.45)$$

Note that for the QSSA portion of Eqs. (6.45), $C_L = 1$, i.e., there is no *fitting* of the model to the computations for $\omega_0 \leq 0.7$. The performance of this hybrid model is provided in Figure 6.42 for the same cases as in Figure 6.33. For this figure, $g_1 = 0.1054$ and

²⁴Indeed, these parameterizations were developed based on guidance provided by the QSSA.

$g_2 = 0.0057$. One should note the considerable improvement over the QSSA alone afforded by this hybrid model. In these cases, the average (absolute) error in the hybrid model was 3.64% for $B_p = 0.0181$ and 3.00% for $B_p = 0.0020$, i.e., it is more uniformly distributed with respect to B_p than the QSSA alone. This hybrid model has the additional advantage of better predicting the angular distribution of the upwelling radiance, i.e., strong dependence on angle for small ω_0 related to the phase function and little dependence at large ω_0 .

6.4.9.2 K_d

We now turn to K_d . In this case, most of the dependence on the solar zenith angle is accounted for by $\langle u_{0w} \rangle_0 = 1/D_0$ in the QSSA, with the dependence on P_{ru} being of minor. This is why Eq. (6.44) was successful in predicting K_d as a function of AOPs. When $P_{ru}(u_0, \phi_0)$ is replaced by B in Eqs. (6.29) or (6.39) as suggested by Eq. (6.44), we have

$$\langle u_{0w} \rangle_0 \frac{\langle K_d \rangle_X}{c} = 1 - \langle C'_K \rangle_X \omega_0 (1 - B) = 1 - \langle C'_K \rangle_X \omega_0 F. \quad (6.46)$$

Figure 6.43 shows how well this relationship is satisfied for the same computations as in Figure 6.37, and Table 6.8 provides the fit to Eq. (6.46). It is interesting to note that the

Table 6.8: Value of $\langle C'_K \rangle_X$ in Eq. (6.46) for various values of X when water scattering is added to particle scattering. B_p is the particle backscattering coefficient. As in Table 6.6, “Ave” and “Max” are the average (unsigned) and maximum error in $\langle K_d \rangle_X$ determined using Eq. (6.46).

X	$B_p = 0.0020$	Ave (%)	Max(%)	$B_p = 0.0181$	Ave (%)	Max(%)
0	0.9869	1.46	6.15	0.9747	1.76	8.43
1	0.9838	1.80	9.24	0.9490	1.76	8.99
2	0.9865	1.24	6.58	0.9417	1.59	11.26
3	0.9889	0.89	4.67	0.9388	1.54	12.88
4	0.9908	0.76	3.50	0.9373	1.80	13.99

fits are similar to, and of similar quality to, those shown in Figure 6.37 and Table 6.7.

6.4.9.3 $R(0)$

For $R(0)$ hybrid models do not seem to be an improvement over the QSSA, however, one is tempted here to try to make similar expansions similar to Eq. (6.42) but with $QSSA'_R$, i.e.,

$$R(0) = C_{R1} QSSA'_R(u_{0w}) + C'_{R2} [QSSA'_R(u_{0w})]^2 + C'_{R3} [QSSA'_R(u_{0w})]^3, \quad (6.47)$$

where

$$QSSA'_R(u_{0w}) = \frac{\omega_0 B}{1 - \omega_0 F} \frac{I_R(u_{0w})}{B} = \frac{\omega_0 B}{1 - \omega_0 F} I'_R(u_{0w}) = \frac{b_b}{a + b_b} I'_R(u_{0w}) = G I'_R(u_{0w}). \quad (6.48)$$

To facilitate using Eq. (6.48) it is useful to reformulate I'_R . Recall that

$$I'_R = \frac{I'_{Rp} b_{bp} + I'_{Rw} b_{bw}}{b_{bp} + b_{bw}}, \quad (6.49)$$

where the subscripts “ w ” and “ p ” stand for water and particles, respectively. Figure 6.18 provides I'_{Rp} and I'_{Rw} as a function of the θ_0 (for the case $m_w = 1.333$) when the particle phase function is either Petzold Phase-T, modified Petzold, or Petzmas. However, a simplification appears if the graph is redrawn using $D_0 = 1/u_{0w}$ as the independent variable. This is provided in Figure 6.44, which shows that both I'_{Rp} and I'_{Rw} are nearly linear functions of D_0 . They can be accurately represented by

$$\begin{aligned} I'_{Rw} &= +0.2098 + 0.1229 D_0, \\ I'_{Rp} &= -0.1988 + 0.4790 D_0. \end{aligned} \quad (6.50)$$

Figure 6.45 provides the fit for $C_{R1} = 1.029$, $C'_{R2} = 1.2526$ and $C'_{R3} = -0.2304$ to the same simulations presented in Figure 6.26, but including the water surface and the atmosphere as in Figures 6.39, 6.41, and 6.42. In this case the results for $B_p = 0.0020$ and 0.0181 were pooled to derive the expansion coefficients. The average (absolute) error in the fit is 3.68%; however, as can be seen in the figure, the maximum error can be large, $\sim 28\%$ in this case. The maximum error here is always for $B \approx B_p$ for the Petzold (or Petzmas) phase function. For a particle phase function with *less* forward scattering than Petzold, we can expect the error in the modified QSSA to be even larger.

6.4.9.4 An Important Observation

These alternate formulations of the AOPs as a function of the IOPs, and the nearness of $\langle C_K \rangle_X$ in Eq. (6.46) to unity, i.e.,

$$\langle u_{0w} \rangle_0 \langle K_d \rangle_X \approx c(1 - \omega_0 F) = a + b_b,$$

suggest that the AOP's K_d , L_u/E_d and R depend principally on a and b_b and in only a minor way on b (or c). Thus, in strongly forward-scattering media such as natural waters, the IOPs b and c *cannot* be derived from measurements of the AOPs. In contrast, a and b_b *can* be estimated from AOP measurements. Such estimation is the subject of Appendix 1 to this chapter (Section 6.6).

6.4.10 The Penetration Depth

As we have mentioned earlier, from a remote sensing perspective, the quantity of most interest is $L_u(z = 0, \mu, \phi)$, as it provides the water's contribution to the radiance measured at the sensor. It is natural to want to understand the range of depths over which photons contributing to L_u are backscattered, i.e., how deeply can the remote sensor “see” into the water? If the IOPs of the water body are independent of depth over this depth range, then the water body can be considered to be a homogeneous medium and the formulae developed in this chapter are applicable. In contrast, if there is significant vertical structure in the IOPs, the formulae would seem to be of little value (however, see the next section). So, in a homogeneous medium, what is the depth over which $L_u(z = 0, \mu, \phi)$ is determined? We will use the QSSA to estimate this depth.

Equations (6.11) and (6.12) provide the single-scattered radiance for a layer of optical thickness τ_1 . In particular, for $\tau = 0$ and $L_0(u', \phi') = F_0 \delta(u' - u_0) \delta(\phi' - \phi_0)$,

$$L_u(\tau = 0, u, \phi, \tau_1, u_0, \phi_0) = \frac{\omega_0 F_0 u_0}{4\pi} \frac{P(u_0 \rightarrow u, \phi_0 \rightarrow \phi)}{u_0 - u} \left\{ 1 - \exp \left[-\tau_1 \left(\frac{1}{u_0} - \frac{1}{u} \right) \right] \right\},$$

where the full dependence of L_u on the angles and τ_1 is provided in the argument list. The term in front of the curly brackets is just $L_u(\tau = 0, u, \phi, \tau_1 = \infty, u_0, \phi_0)$. Replacing u by $-\mu$, we have

$$\frac{L_u(\tau = 0, \mu, \phi, \tau_1, u_0, \phi_0)}{L_u(\tau = 0, \mu, \phi, \infty, u_0, \phi_0)} = 1 - \exp \left[-\tau_1 \left(\frac{1}{u_0} + \frac{1}{\mu} \right) \right].$$

Now to make the transition to QSSA, we replace τ_1 by

$$\tau'_1 = \tau_1 \left(1 - \frac{\omega_0 A_\delta}{4\pi} \right) \rightarrow cz_1 [1 - \omega_0 (1 - B'(u_0))]$$

Therefore, within the context of the QSSA,

$$\frac{L_u(\tau = 0, \mu, \phi, \tau_1, u_0, \phi_0)}{L_u(\tau = 0, \mu, \phi, \infty, u_0, \phi_0)} = 1 - \exp \left[-\{c[1 - \omega_0 (1 - B'(u_0))]\} \left(\frac{1}{u_0} + \frac{1}{\mu} \right) z_1 \right]. \quad (6.51)$$

The quantity in the curly brackets is recognized as $u_0 K_u(z) = u_0 K_L(z)$ in the QSSA (Eq. (6.21)); however, we note that the QSSA approximations to K_u and K_L are not as accurate as those for the other AOPs, e.g., it gives the wrong relationship between the magnitudes of $K_d(0)$ and $K_u(0)$. In any event, according to the QSSA,

$$\frac{L_u(\tau = 0, \mu, \phi, \tau_1, u_0, \phi_0)}{L_u(\tau = 0, \mu, \phi, \infty, u_0, \phi_0)} = 1 - \exp \left[-K_L z_1 \left(1 + \frac{u_0}{\mu} \right) \right]. \quad (6.52)$$

There is another way to derive a formula similar to Eq. (6.52). Consider a semi-infinite homogeneous medium, The radiance reflected from such a medium is

$$L_u(\tau = 0, \mu, \phi, \infty, u_0, \phi_0) = \frac{R}{Q} E_d(u_0) = \frac{R}{Q} F_0 u_0 \quad (6.53)$$

If a thin layer of thickness z_1 with the same IOPs is added to the top of this medium, then the total exiting radiance is

$$(L_u)_{\text{Total}} = (L_u)_{\text{Layer}} + (L_u)_{\text{Semi-infinite}},$$

where the first term is due to the layer and the second to the medium below it. It is easy to approximate the second term: it is just Eq. (6.53) with $E_d(u_0)$ reduced by a factor $\exp[-K_d(u_0)z_1]$ and then L_u reduced by an additional factor $\exp[-K_L(\mu)z_1]$. Thus,

$$(L_u)_{\text{Total}} = (L_u)_{\text{Layer}} + \frac{R}{Q} F_0 u_0 \exp[-z_1(K_d(u_0) + K_L(\mu))].$$

But, the addition of this layer simply redefines the zero of depth, it doesn't change the value of the reflected radiance (i.e., $(L_u)_{\text{Total}}$ is still $(R/Q)F_0 u_0$). Thus,²⁵

$$\begin{aligned} (L_u(z = 0, \mu, \phi, z_1, u_0, \phi_0))_{\text{Layer}} &= L_u(z = 0, \mu, \phi, \infty, u_0, \phi_0) \\ &\times (1 - \exp[-z_1(K_d(u_0) + K_L(\mu))]) \end{aligned} \quad (6.54)$$

Note the similarity to Eq. (6.52). This equation is not exact, but it is reasonably accurate for small z_1 . The main approximation in this equation is that the radiance entering the semi-infinite layer *after* the thin layer is added is still in the form of a collimated beam, which of course is not the case, but for sufficiently small z_1 it is a good approximation. It has also been assumed that K_d and K_L are constant within the layer which, as shown in Figures 6.9 and 6.13, is not valid. Thus, it is more appropriate that K_d and K_L be replaced by $\langle K_d \rangle$ and $\langle K_L \rangle$ with the averages taken over $0 \rightarrow z_1$.

To try to understand the efficacy of these formulae, we computed $L_u(z = 0, \mu, \phi, z_1, u_0, \phi_0)$ for a layer of thickness z_1 and IOPs identical to those in Figure 6.9 (flat interface, no atmosphere) for $\theta_0 = 0$ and 60° , $\lambda = 440$ nm and 550 nm, and formed the ratio

$$\frac{L_u(z = 0, \mu, \phi, z_1, u_0, \phi_0)}{L_u(z = 0, \mu, \phi, \infty, u_0, \phi_0)}$$

for $\mu = 1$. We have no accurate estimate for K_L from the QSSA; however, it is computed in the simulation and therefore available. Since the K 's are depth dependent, we use their values averaged over z_1 . Figure 6.46 shows the ratio $L_u(z_1)/L_u(\infty)$ as a function of $\langle K_d + K_L \rangle z_1$ for the cases considered, along with the prediction of Eq. (6.54) with the K 's

²⁵This is an example of the idea of invariant imbedding discussed in Chapter 2, Section 2.7.4.

Table 6.9: Values of $\langle K_d(z) \rangle$ (in m^{-1}) near $\langle K_d(z) \rangle z_1 = 1$ for the simulation parameters used in Figures 6.46 and 6.47. The value of $1/\langle K_d(z) \rangle$ (in m) is provided in parenthesis.

θ_0	$\lambda = 440 \text{ nm}$	$\lambda = 550 \text{ nm}$
0	0.027 (37)	0.078 (13)
60°	0.036 (28)	0.100 (10)

replaced by their averages. Note that Eq. (6.54) follows the exact computation reasonably well for $\theta_0 = 0$, but overestimates $L_u(z_1)/L_u(\infty)$ for $\theta_0 = 60^\circ$. Also, when $\langle K_d + K_L \rangle z_1 = 2$, $L_u(z_1)$ has reached between 79 and 85% of its value for an infinitely deep water body, and between 86 and 91% for $\langle K_d + K_L \rangle z_1 = 2.5$. Thus, it is reasonable to define the remote sensing *penetration depth* (z_{Pd}) as the depth $z_{Pd} = 2/\langle K_d + K_L \rangle$ to $2.5/\langle K_d + K_L \rangle$. This is the depth over which 85-90% of the remote sensing signal (L_u) originates. Unfortunately, the QSSA performs poorly in estimating K_L ; however, accurate estimates of $K_d(0)$ and $\langle K_d \rangle$ are available. Figure 6.13 suggests that $K_L(z) \approx K_d(z)$ and $\langle K_L(z) \rangle \approx \langle K_d(z) \rangle$, so it is reasonable to investigate the accuracy of

$$\frac{L_u(z = 0, \mu, \phi, z_1, u_0, \phi_0)}{L_u(z = 0, \mu, \phi, \infty, u_0, \phi_0)} = 1 - \exp[-2\langle K_d(u_0) \rangle z_1] \quad (6.55)$$

This is provided in Figure 6.47. There is virtually no difference between this equation and Eq. (6.54), and an equally good estimate of (z_{Pd}) is

$$z_{Pd} = \frac{1}{\langle K_d(z) \rangle} \quad \text{to} \quad \frac{1.25}{\langle K_d(z) \rangle} \quad (6.56)$$

For the cases considered, $\langle K_d(z) \rangle$ at depths near $\langle K_d(z) \rangle z_1 \approx 1$ are given in Table 6.9 as can be roughly verified by examination of Figure 6.9. Note that at 440 nm with $\theta_0 = 60^\circ$ only about 30% of L_u is reflected from the depth over which 80% is reflected at 550 nm. At $\theta_0 = 0$ the corresponding numbers are 40 and 80%.

As a “real-world” example of penetration depths, Figure 6.48 provides an estimate of the penetration depth for the data presented in Figures 6.5 and 6.6 for Case 1 waters. Note that in the blue, z_{Pd} varies from 2 to 50 m for these data. Also note the strong dependence on wavelength with $z_{Pd}(440)$ usually, but not always, much greater than $z_{Pd}(550)$. In the open oceans, the water is usually well mixed near the surface down to some depth (the top of the thermocline) due to solar heating and wind-induced mixing. This mixed layer can extend from a few meters to over 100 m, depending on geophysical factors. In the mixed layer the IOPs are nearly independent of depth and the layer can be considered optically homogeneous. Below the mixed layer the IOPs can in some cases depend strongly on depth. It is immediately apparent that for the water to be considered homogeneous throughout the spectrum for the purposes of remote sensing, the mixed layer must extend

beyond the depth of the greatest z_{Pd} . Interpretation of L_u in terms of the IOPs of the medium becomes an almost intractable problem if the mixed layer depth is $<$ the maximum z_{Pd} . A common algorithm in remote sensing relates $L_u(440)/L_u(550)$ to the concentration of Chlorophyll a in the water column. Such an algorithm could be clearly frustrated at relative low Chlorophyll concentrations were the mixed layer depth less than about 15 m and the water strongly stratified below the mixed layer. Finally, if asked how deep the remote sensor can “see” into the water, one answer is about 50 m in the blue in clear ocean water.

6.4.11 QSSA Estimate of the Influence of Vertical Stratification

The IOPs of most water bodies are stratified in the vertical (below the mixed layer); however, our discussion thus far has been limited to media with depth-independent IOPs. Here we briefly apply the QSSA to a vertically stratified medium. From Chapter 2 (Appendix 1), the first-order radiative transfer equation for $L^{(1)}$ in a medium of optical depth τ_1 is given by

$$\begin{aligned} u \frac{\partial}{\partial \tau} L^{(0)}(\tau, u, \phi) + L^{(0)}(\tau, u, \phi) &= 0 \\ u \frac{\partial}{\partial \tau} L^{(1)}(\tau, u, \phi) + L^{(1)}(\tau, u, \phi) &= \frac{1}{4\pi} \int_{-1}^1 du' \int_0^{2\pi} d\phi' \omega_0(\tau) P(\tau, u' \rightarrow u, \phi' \rightarrow \phi) L^{(0)}(\tau, u', \phi') \end{aligned}$$

with the boundary conditions

$$\begin{aligned} L^{(0)}(0, u, \phi) &= L(0, u, \phi) & u > 0, \\ L^{(0)}(\tau_1, u, \phi) &= 0 & u < 0, \end{aligned}$$

and

$$\begin{aligned} L^{(1)}(0, u, \phi) &= 0, & u > 0, \\ L^{(1)}(\tau_1, u, \phi) &= 0, & u < 0. \end{aligned}$$

Here, the dependence of the IOPs on depth is explicitly given by the τ in the argument list of ω_0 and P . The solution for $L^{(0)}$ is $L^{(0)} = L(0, u, \phi) \exp(-\tau/\mu)$. If $L(0, u, \phi)$ is a beam source directed toward the (u_0, ϕ_0) direction, i.e., $L(0, u, \phi) = F_0 \delta(u - u_0) \delta(\phi - \phi_0)$, then $L^{(1)}$ satisfies

$$u \frac{\partial}{\partial \tau} L^{(1)}(\tau, u, \phi) + L^{(1)}(\tau, u, \phi) = \frac{\omega_0(\tau) F_0}{4\pi} P(\tau, u_0 \rightarrow u, \phi_0 \rightarrow \phi) \exp(-\tau/u_0).$$

To solve this, we first simplify the notation and consider a single direction (u, ϕ) . Then

$$u \frac{dL^{(1)}}{d\tau} + L^{(1)} = F_0 A(\tau) \exp(-\tau/u_0),$$

where $A(\tau) = \omega_0(\tau)P(\tau, u_0 \rightarrow u, \phi_0 \rightarrow \phi)/4\pi$. Multiplying this by $\exp(\tau/u)$ and simplifying the left-hand-side yields

$$u \frac{d}{d\tau} \left(L^{(1)} \exp \left[\frac{\tau}{u} \right] \right) = F_0 A(\tau) \exp \left[\tau \left(\frac{1}{u} - \frac{1}{u_0} \right) \right],$$

which can be immediately integrated between two depths τ_a and τ_b :

$$L^{(1)}(\tau_b) \exp \left[\frac{\tau_b}{u} \right] - L^{(1)}(\tau_a) \exp \left[\frac{\tau_a}{u} \right] = \frac{1}{u} \int_{\tau_a}^{\tau_b} F_0 A(\tau) \exp \left[\tau \left(\frac{1}{u} - \frac{1}{u_0} \right) \right] d\tau.$$

Now, if $\tau_b = \infty$, we have for the upward radiance ($u = -\mu$),

$$L^{(1)}(\tau_a, \mu, \phi) = \frac{\exp[\tau_a/\mu]}{\mu} \int_{\tau_a}^{\infty} F_0 A(\tau) \exp \left[-\tau \left(\frac{1}{u_0} + \frac{1}{\mu} \right) \right] d\tau.$$

Furthermore, if $A(\tau) = A$, independent of τ (i.e., a homogeneous medium), the integration results in

$$L^{(1)}(\tau_a, \mu, \phi) = \frac{u_0 F_0 \exp(-\tau_a/u_0)}{u_0 + \mu} A \triangleq L_H^{(1)}(\tau_a, \mu, \phi),$$

where $L_H^{(1)}(\tau_a, \mu, \phi)$ is the first-order solution for the upward radiance in a homogeneous medium, i.e., the subscript “ H ” stands for “homogeneous.” Recognizing that to this order,

$$E_d^{(1)}(\tau_a) = u_0 F_0 \exp(-\tau_a/u_0),$$

we can write the radiance reflectance

$$R_{LH}^{(1)}(\tau_a, \mu, \phi) \triangleq \frac{L_H^{(1)}(\tau_a, \mu, \phi)}{E_d^{(1)}(\tau_a)} = \frac{A}{u_0 + \mu},$$

but, since A is independent of τ_a in a homogeneous medium, this is the same as $R_{LH}(0, \mu, \phi)$. In a similar manner, for the inhomogeneous case A is dependent on τ , and we identify

$$A(\tau) = (u_0 + \mu) R_{LH}^{(1)}(\tau, \mu, \phi),$$

where $R_{LH}^{(1)}(\tau, \mu, \phi)$ is the radiance reflectance of a *homogeneous medium* with IOPs equal to those in the *stratified medium* at the optical depth τ , i.e., $R_{LH}^{(1)}(\tau, \mu, \phi) = R_{LH}^{(1)}(0, \mu, \phi)$,

where the latter is the reflectance of a homogeneous medium with IOPs that exist in the stratified medium at the optical depth τ . Letting

$$R_L^{(1)}(\tau_a, \mu, \phi) = \frac{L^{(1)}(\tau_a, \mu, \phi)}{E_d^{(1)}(\tau_a)},$$

we have

$$R_L^{(1)}(\tau_a, \mu, \phi) = \int_{\tau_a}^{\infty} \frac{u_0 + \mu}{u_0 \mu} R_{LH}^{(1)}(\tau, \mu, \phi) \exp \left[-(\tau - \tau_a) \left(\frac{1}{u_0} + \frac{1}{\mu} \right) \right] d\tau.$$

Since

$$\frac{d}{d\tau} \exp \left[-(\tau - \tau_a) \left(\frac{1}{u_0} + \frac{1}{\mu} \right) \right] = -\frac{u_0 \mu}{u_0 + \mu} \exp \left[-(\tau - \tau_a) \left(\frac{1}{u_0} + \frac{1}{\mu} \right) \right],$$

letting

$$M(u_0, \mu) \triangleq \frac{1}{u_0} + \frac{1}{\mu},$$

the final result we obtain is

$$R_L^{(1)}(\tau_a, \mu, \phi) = - \int_{\tau_a}^{\infty} R_{LH}^{(1)}(\tau, \mu, \phi) \frac{d}{d\tau} \exp [-(\tau - \tau_a) M(u_0, \mu)] d\tau. \quad (6.57)$$

At the surface Eq. (6.57) reduces to

$$R_L^{(1)}(0, \mu, \phi) = - \int_0^{\infty} R_{LH}^{(1)}(\tau, \mu, \phi) \frac{d}{d\tau} \exp [-\tau M(u_0, \mu)] d\tau. \quad (6.58)$$

If we approximate the vertical structure of the medium by a set of layers, each with homogeneous IOPs between τ_i and τ_{i+1} , then the radiance ratio at the surface ($\tau_a = 0$) can be written

$$R_L^{(1)}(\tau = 0, \mu, \phi) = \sum_{i=0}^{\infty} R_{LH}^{(1)}(\tau_i \rightarrow \tau_{i+1}, \mu, \phi) \{ \exp [-M(u_0, \mu) \tau_i] - \exp [-M(u_0, \mu) \tau_{i+1}] \}, \quad (6.59)$$

where $R_{LH}^{(1)}(\tau_i \rightarrow \tau_{i+1}, \mu, \phi)$ is the value of $R_{LH}^{(1)}$ just beneath the surface of a homogeneous water body with the same IOPs as extant between τ_i and τ_{i+1} .

The transition from the first-order approximation to the QSSA is made as usual by replacing τ with τ' , where

$$d\tau' = d\tau \left(1 - \frac{\omega_0 A_\delta}{4\pi} \right) \rightarrow c dz [1 - \omega_0 (1 - B'(u_0))].$$

But, as we have noted earlier, the QSSA formulas work better (and satisfy reciprocity) if $B'(u_0)$ is replaced by B in which case $d\tau' = c dz [1 - \omega_0 F]$, and since $1 - \omega_0 F \approx u_0 K_d / c$, we have $d\tau' \approx u_0 K_d dz$, so²⁶

$$\tau' = \int_0^z u_0 K_d(z) dz \triangleq u_0 \langle K_d \rangle_{0 \rightarrow z} z.$$

Thus, the final QSSA result for the radiance reflectance of a stratified water body is²⁷

$$\begin{aligned} R_L(z=0, \mu, \phi) &= \sum_{i=0}^{\infty} R_{LH}(z_i \rightarrow z_{i+1}, \mu, \phi) \\ &\times \left\{ \exp[-M(u_0, \mu) u_0 \langle K_d \rangle_{0 \rightarrow z_i} z_i] - \exp[-M(u_0, \mu) u_0 \langle K_d \rangle_{0 \rightarrow z_{i+1}} z_{i+1}] \right\}. \end{aligned} \quad (6.60)$$

Note that the superscripts “(1)” have been removed from the equation, i.e., it is approximately a *multiple scattering* result in the QSSA.

We will not provide an exhaustive investigation into the accuracy of Eq. (6.60). Instead we will examine two specific examples: a 2-layer medium and a 3-layer medium for two values of θ_0 . For both we carry out exact radiative transfer simulations using the Petzold phase function for the medium, including the refracting water surface, but excluding the atmosphere. In the 2-layer case, the upper layer has $\omega_0 = 0.7$, while ω_0 for the lower layer varies from 0 to 0.9, and extends from $\tau = 0.5$ to ∞ . Thus, we have a moderately scattering upper layer over a semi-infinite lower layer with IOPs varying from low to high scattering relative to absorption. Figure 6.49 compares the result of applying Eq. (6.60) with the exact computations. With the exception of $\omega_0(\text{Lower}) = 0$ and $\theta_0 = 60^\circ$, the maximum error is 6% and usually much less. In the 3-layer case, the same phase function and θ_0 values are used. The top layer extends from $\tau = 0$ to $\tau = 0.5$ and has $\omega_0 = 0.7$. The bottom layer extends from $\tau = 1$ to ∞ and also has $\omega_0 = 0.7$. In the middle layer ω_0 varies from 0 to 0.9. The comparison of Eq. (6.60) with exact computations is presented in Figure 6.50. Here, with the exception of $\omega_0(\text{Middle}) = 0$ or 0.9, the maximum error is less than 5%. These comparisons suggest that Eq. (6.60) is remarkably accurate considering the approximations involved in its derivation. It is certainly sufficiently accurate to be used as a guide for understanding the influence of vertical stratification on R_L .

²⁶Since $c(1 - \omega_0 F) = a + b_b$, $c dz \rightarrow (a + b_b) dz$, and a simple “rule of thumb,” that SSA \rightarrow QSSA when $c \rightarrow a + b_b$, will yield excellent results (although not precisely QSSA, e.g., Section 6.4.9).

²⁷This equation can be written in an alternate form. If we label the layers starting from the top calling R_{LH} for the first layer R_1 , $\langle K_d \rangle$ for the first layer K_1 , and similarly for the second layer R_2 and K_2 , etc., then

$$R_L = R_1 + (R_2 - R_1) \exp(-M\tau'_1) + (R_3 - R_2) \exp(-M\tau'_2) \cdots$$

where

$$\tau'_1 = u_0 K_1 z_1, \quad \tau'_2 = u_0 [K_1 z_1 + K_2 (z_2 - z_1)], \quad \text{etc.}$$

We emphasize again that the R 's here refer to the radiance reflectance (just beneath the surface) of a *homogeneous* medium with the same IOPs as extant for the given layer in the stratified medium,

It is interesting to look at the structure of the K 's in a stratified medium. A simple example will shed some light on their behavior. Consider the two extreme cases in the simulations used in Figure 6.50, i.e., $\omega_0(\text{Middle}) = 0$ and $\omega_0(\text{Middle}) = 0.9$. Figure 6.51 provides the depth profiles of K_d and K_L from the radiative transfer simulations for these two cases. One sees immediately that in each of the three individual regions, K_d is essentially independent of depth, and independent of the IOPs in the adjoining region. In the region $0.5 \leq \tau \leq 1.0$, $K_d/c \approx 1.02$ and 0.125 for $\omega_0(\text{Middle}) = 0$ and 0.9 , respectively. These should be compared to 1.00 and about 0.128 for a homogeneous media with $\omega_0 = 0$ and 0.9 , respectively, over a similar depth range. Thus, the appropriate value of K_d for a given layer is that of a homogeneous medium with the IOPs of the layer in question, i.e., K_d is a *local property* of the medium: at a given depth z it depends mostly on the IOP's at z .

The behavior of K_L is not as simple as that of K_d , but nevertheless is easy to understand with the help of Figure 6.52, which provides the zenith propagating radiance (and $R_L(\tau)$) as a function of depth for the three values of $\omega_0(\text{Middle})$. First consider $\omega_0(\text{Middle}) = 0$ and start in the bottom layer. In that layer, K_L is nearly identical to that of a homogeneous medium with the same IOPs. When radiance scattered by the bottom layer propagates toward the zenith (L_u) it encounters the middle layer, in which there is no scattering, so it is simply absorbed as it goes from $\tau = 1.0$ to 0.5 , i.e., the radiance decreases with decreasing depth so K_L is *negative*. The value of -1 for K_L/c follows from the fact that here $K_L = -a$. Proceeding into the top layer, new upward radiance is scattered from within this layer, with the amount increasing with decreasing depth. Thus, K_L is large near $\tau = 0.5$, but decreases in magnitude as one approaches the surface; however, because of the lack of radiance generated in the middle layer, K_L does not reach the value obtained in a homogeneous medium. For $\omega_0(\text{Middle}) = 0.9$ we again start with the bottom layer and see that K_L is the same as that for a homogeneous medium. In the middle layer K_L decreases with depth in a manner similar to the top layer in the $\omega_0(\text{Middle}) = 0$ case, and for the same reason: additional radiance is generated by the middle layer. In the top layer K_L is less than that for a homogeneous medium because of the additional radiance generated by the middle layer. In contrast to K_d , K_L at a given τ is a strong function of the structure of the IOPs for the medium at depths $> \tau$, while K_d depends mostly on the IOPs at τ .

6.4.12 The Influence of Inelastic Processes on the AOPs

In inelastic processes a photon is absorbed at one wavelength and subsequently a photon is emitted at a different (usually longer) wavelength. Thus, there is an internal source of light at the emission wavelength, the propagation of which depends on the IOPs at

the emission wavelength. The theory of radiative transfer in the presence of inelastic processes is most easily developed by considering such processes to be scattering of an incident photon *plus* a change in wavelength (energy) of the scattered photon, hence the term *inelastic scattering*. In natural waters, there are two important inelastic processes: fluorescence by suspended or dissolved constituents in the water, and Raman scattering by the water molecules. The most important fluorescence process (from a remote sensing view point) is the natural fluorescence of Chlorophyll *a* contained in phytoplankton. Photons absorbed by Chlorophyll *a*, and by the accessory pigments within phytoplankton, induce fluorescent emission near 683 nm. An example of this is the strong peak in the reflectance *R* extending from about 670 to 700 nm seen in Station C92 in Figure 6.5. Its importance depends primarily on the concentration of Chlorophyll *a*, and secondarily on the light exposure history of the phytoplankton population. A secondary process (of less importance to remote sensing) is the fluorescence of dissolved organic material. Raman scattering, resulting from the interaction of light with the water molecule (Chapters 3 and 5), is an inelastic process that is always present in natural waters, and as such its importance must always be ascertained. In this section, we examine the influence of these two processes on the AOPs of the water. We start by solving the radiative transfer equation in the lowest order that incorporates inelastic processes, and then specifically apply the solution to fluorescence and Raman scattering.

6.4.12.1 Radiative Transfer and Inelastic Processes.

The radiative transfer equation including inelastic processes was provided in Chapter 2.

$$\begin{aligned} \hat{\xi} \bullet \nabla L(\vec{r}, \hat{\xi}, \lambda) = & -c(\vec{r}, \lambda)L(\vec{r}, \hat{\xi}, \lambda) \\ & + \int_{\text{All } \hat{\xi}'} \beta(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda)L(\vec{r}, \hat{\xi}', \lambda) d\Omega(\hat{\xi}') \\ & + \int_{\lambda_E < \lambda} \int_{\text{All } \hat{\xi}'} \beta_{\text{In}}(\vec{r}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda_E \rightarrow \lambda)L(\vec{r}, \hat{\xi}', \lambda_E) d\Omega(\hat{\xi}') d\lambda_E. \end{aligned}$$

The last term represents the contribution of inelastic scattering. We consider a 1-D medium, i.e., one in which the IOPs depend only on depth and the incident illumination is independent of horizontal position. Then this reduces to

$$\begin{aligned} u \frac{dL(z, u, \phi, \lambda)}{dz} + c(z, \lambda)L(z, u, \phi, \lambda) = & \int_0^{2\pi} d\phi' \int_{-1}^1 \beta(z, u' \rightarrow u, \phi' \rightarrow \phi, \lambda)L(z, u', \phi', \lambda) du' \\ & + \int_{\lambda_E < \lambda} d\lambda_E \int_0^{2\pi} d\phi' \int_{-1}^1 \beta_{\text{In}}(z, u' \rightarrow u; \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda)L(z, u', \phi', \lambda_E) du'. \end{aligned} \quad (6.61)$$

We write the last term as

$$Q(z, u, \phi, \lambda) = \int_{\lambda_E < \lambda} d\lambda_E J(z, u, \phi, \lambda, \lambda_E), \quad (6.62)$$

where

$$J(z, u, \phi, \lambda, \lambda_E) = \int_0^{2\pi} d\phi' \int_{-1}^1 \beta_{\text{In}}(z, u' \rightarrow u; \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda) L(z, u', \phi', \lambda_E) du'. \quad (6.63)$$

We now make the transition to dimensionless variables $d\tau = c(z, \lambda) dz$, $P = \beta/4\pi b$ and $\omega_0(z, \lambda) = b(z, \lambda)/c(z, \lambda)$, where b is the scattering coefficient at λ . Then,

$$\mathcal{L}L = \left(u \frac{d}{d\tau} + 1\right) L = \frac{\omega_0}{4\pi} \int_0^{2\pi} d\phi' \int_{-1}^1 PL du' + \frac{Q}{c}.$$

We assume that the radiance at λ_E is known, so $Q = Q(z, u, \phi, \lambda)$ is a known function of its arguments. The solution of this equation can be developed in a manner similar to the successive order of scattering (Chapter 2) by expanding the radiance in increasing powers of ω_0 . (This is most straightforwardly done by initially assuming that ω_0 is independent of depth.)

$$L = L^{(0)} + \omega_0 L^{(1)} + \omega_0^2 L^{(2)} + \dots,$$

yielding a set of differential equations;

$$\begin{aligned} \mathcal{L}L^{(0)}(\tau, u, \phi, \lambda) &= \frac{Q(\tau, u, \phi, \lambda)}{c(\tau, \lambda)}, \\ \mathcal{L}L^{(1)}(\tau, u, \phi, \lambda) &= \frac{1}{4\pi} \int_0^{2\pi} d\phi' \int_{-1}^1 P(\tau, u' \rightarrow u, \phi' \rightarrow \phi, \lambda) L^{(0)}(\tau, u', \phi', \lambda) du', \\ \mathcal{L}L^{(2)}(\tau, u, \phi, \lambda) &= \frac{1}{4\pi} \int_0^{2\pi} d\phi' \int_{-1}^1 P(\tau, u' \rightarrow u, \phi' \rightarrow \phi, \lambda) L^{(1)}(\tau, u', \phi', \lambda) du', \\ &\vdots \end{aligned} \quad (6.64)$$

where it must be remembered that τ is the optical depth at λ . The first term is the result of the inelastic scattering in the direction (u, ϕ) directly transmitted from its position of generation to the depth z , while the second corresponds to inelastically-generated radiance in the direction (u', ϕ') being scattered in the direction (u, ϕ) before reaching the depth z , etc. Each of these differential equations is of the form

$$\mathcal{L}L^{(n)} = \left(u \frac{d}{d\tau} + 1\right) L^{(n)} = f^{(n)}(\tau),$$

where $f^{(n)}(\tau)$ is a known function of τ , derived from Q or $L^{(n-1)}$. As we have seen in earlier sections (e.g., 6.4.11), if we multiply this by $\exp[\tau/u]$, the equation can be rewritten

$$u \frac{d}{d\tau} [L^{(n)} \exp(\tau/u)] = [\exp(\tau/u)] f^{(n)}(\tau),$$

and the solution can be found by direct integration from τ_a to τ_b :

$$L^{(n)}(\tau_b) \exp(\tau_b/u) - L^{(n)}(\tau_a) \exp(\tau_a/u) = \frac{1}{u} \int_{\tau_a}^{\tau_b} f^{(n)}(\tau') \exp(\tau'/u) d\tau'.$$

If we first take u to be positive (downward propagating radiance) with $\tau_a = 0$ and no inelastically scattered radiance *entering* the medium, i.e., $L^{(n)}(0) = 0$, and then $u < 0$ (upward propagating radiance) with $\tau_b = \infty$ and $L^{(n)}(\infty) = 0$, we arrive at the solutions:²⁸

$$\begin{aligned} L^{(n)}(\tau, u, \phi) &= \frac{\exp(-\tau/u)}{u} \int_0^\tau f^{(n)}(\tau', u, \phi) \exp(\tau'/u) d\tau' & u > 0 \\ L^{(n)}(\tau, u, \phi) &= \frac{\exp(-\tau/u)}{u} \int_\infty^\tau f^{(n)}(\tau', u, \phi) \exp(\tau'/u) d\tau' & u < 0 \end{aligned} \quad (6.65)$$

The first of these equations ($n = 0$) yields

$$\begin{aligned} L^{(0)}(\tau, u, \phi, \lambda) &= \frac{\exp(-\tau/u)}{u} \int_0^\tau \frac{Q(\tau', u, \phi, \lambda)}{c(\tau', \lambda)} \exp(\tau'/u) d\tau' & u > 0 \\ L^{(0)}(\tau, u, \phi, \lambda) &= \frac{\exp(-\tau/u)}{u} \int_\infty^\tau \frac{Q(\tau', u, \phi, \lambda)}{c(\tau', \lambda)} \exp(\tau'/u) d\tau' & u < 0 \end{aligned} \quad (6.66)$$

and provides the lowest order estimate for the radiance resulting from inelastic scattering. We will see that it is often sufficient to meet remote sensing needs. If not, $L^{(0)}(\tau, u, \phi, \lambda)$ can be inserted into Eq. (6.64) for computation of $L^{(1)}(\tau, u, \phi, \lambda)$, etc.

6.4.12.2 Fluorescence

We now apply the formulas derived above to fluorescence. A common trait of fluorescent emission is that it is isotropic, that is $\beta_{\text{In}} \triangleq \beta_F$ is independent of the angle between $\hat{\xi}$ and $\hat{\xi}'$. This result is due to the time lag between the absorption at λ_E and the emission at λ . During this time, the fluorescent molecule makes many rotations and “forgets” the direction of the incident photon. Thus, the inelastic scattering function becomes

$$\beta_F(z, u' \rightarrow u, \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda) = \frac{b_F(z, \lambda_E \rightarrow \lambda)}{4\pi},$$

and inserting this into Eq. (6.63) yields

$$J(z, u, \phi, \lambda) = \frac{b_F(z, \lambda_E \rightarrow \lambda)}{4\pi} \int_0^{2\pi} d\phi' \int_{-1}^1 L(z, u', \phi', \lambda_E) du'.$$

²⁸For $u < 0$, it might appear that $L^{(n)}$ is negative because $u < 0$; however note that the limits on the integral are $\infty \rightarrow \tau$. so the integral is also negative, and $L^{(n)} > 0$.

The integral is just the scalar irradiance, $E_0(z, \lambda_E)$, at λ_E , so finally the source function for fluorescence is

$$Q_F(z, u, \phi, \lambda) = \int_{\lambda_E < \lambda} \frac{b_F(z, \lambda_E \rightarrow \lambda)}{4\pi} E_0(z, \lambda_E) d\lambda_E.$$

Combining this with Eqs. (6.66) provides the lowest order contribution of fluorescence to the radiance at λ :

$$\begin{aligned} L^{(0)}(\tau, u, \phi, \lambda) &= \frac{\exp(-\tau/u)}{4\pi u} \int_{\lambda_E < \lambda} d\lambda_E \\ &\times \int_0^\tau b_F(\tau', \lambda_E \rightarrow \lambda) E_0(\tau', \lambda_E) \exp(\tau'/u) \frac{d\tau'}{c(\tau', \lambda)}; \quad u > 0, \end{aligned}$$

with a similar expression for the upward radiance but with the integration from $\infty \rightarrow \tau$. Noting that $\tau(z) = \int_0^z c(z) dz$, we can rewrite this in terms of real depth; however, the equations are considerably simplified if we assume that $c(z, \lambda) \triangleq c$ is independent of z , i.e., $\tau = c(\lambda)z$. Then

$$\begin{aligned} L^{(0)}(z, u, \phi, \lambda) &= \frac{\exp(-cz/u)}{4\pi u} \int_{\lambda_E < \lambda} d\lambda_E \\ &\times \int_0^z b_F(z', \lambda_E \rightarrow \lambda) E_0(z', \lambda_E) \exp(cz'/u) dz'; \quad u > 0, \\ L^{(0)}(z, u, \phi, \lambda) &= \frac{\exp(-cz/u)}{4\pi u} \int_{\lambda_E < \lambda} d\lambda_E \\ &\times \int_\infty^z b_F(z', \lambda_E \rightarrow \lambda) E_0(z', \lambda_E) \exp(cz'/u) dz'; \quad u < 0. \end{aligned} \tag{6.67}$$

The only approximation used in these equations is that of a depth-independent $c(\lambda)$. If we further assume that b_F is independent of depth and that

$$E_0(z, \lambda_E) = E_0(0, \lambda_E) \exp[-K_0(\lambda_E)z]$$

with a constant scalar irradiance attenuation coefficient, we can carry out the z' integration and find

$$\begin{aligned} L_d^{(0)}(z, u, \phi, \lambda) &= \frac{1}{4\pi u} \int_{\lambda_E < \lambda} d\lambda_E \left[\frac{E_0(0, \lambda_E) b_F(\lambda_E \rightarrow \lambda)}{c(\lambda) - K_0(\lambda_E)u} \right] \\ &\times \{\exp[-K_0(\lambda_E)z] - \exp[-cz/u]\}; \quad u > 0, \\ L_u^{(0)}(z, \mu, \phi, \lambda) &= \frac{1}{4\pi \mu} \int_{\lambda_E < \lambda} d\lambda_E \left[\frac{E_0(0, \lambda_E) b_F(\lambda_E \rightarrow \lambda)}{c(\lambda) + K_0(\lambda_E)\mu} \right] \\ &\times \exp[-K_0(\lambda_E)z]; \quad \mu = -u > 0, \end{aligned}$$

where the subscripts “d” and “u” stand for “downward” and “upward,” respectively. These equations represent the lowest order estimate for the radiance generated by inelastic scattering through fluorescence. If more accuracy is desired, they can be used in Eq. (6.64) to determine $L^{(1)}$, etc. One can use these equations to determine the lowest order estimate of the upward and downward irradiances. For the upward case (of more interest to remote sensing) the simple integration ($E_u = \int_0^1 \mu d\mu \int_0^{2\pi} d\phi L$) yields

$$E_u^{(0)}(z) = \int_{\lambda_E < \lambda} d\lambda_E \left[\frac{E_0(0, \lambda_E) b_F(\lambda_E \rightarrow \lambda)}{2K_0(\lambda_E)} \right] \times \left\{ 1 - \frac{c(\lambda)}{K_0(\lambda_E)} \ell n \left(1 + \frac{K_0(\lambda_E)}{c(\lambda)} \right) \right\} \exp[-K_0(\lambda_E)z].$$

An important feature of these results is that the upward radiance and irradiance produced through fluorescence excited by a narrow band of wavelengths $\Delta\lambda_E$ decays with depth exponentially, but with a decay coefficient equal to $K_0(\lambda_E)$, i.e., from excitation within $\Delta\lambda_E$ around λ_E , the upward light field decays with depth with attenuation coefficients $K_L(\lambda) = K_u(\lambda) = K_0(\lambda_E)$. To make the transition to the QSSA for fluorescence, we need only replace $c(\lambda)$ by $a(\lambda) + b_b(\lambda)$.

It is often useful to replace the inelastic scattering coefficient by the *quantum efficiency of fluorescence*, $\eta(\lambda, \lambda_E)$. This quantity is defined as the rate of emission of photons between $\pm\Delta\lambda/2$ around λ , divided by the rate of absorption of photons between $\pm\Delta\lambda_E/2$ around λ_E , i.e.,

$$\eta(\lambda, \lambda_E) = \frac{\Delta N(\text{Emitted within } \Delta\lambda)/\Delta t}{\Delta N(\text{Absorbed within } \Delta\lambda_E)/\Delta t},$$

where N stands for number of photons and t is time. Now, $\Delta N/\Delta t = E_1^{-1} \Delta E/\Delta t$, where E_1 is the energy of a single photon: $E_1 = hc/\lambda$. Therefore $\Delta N/\Delta t = (\lambda/hc) \times$ (Power emitted or absorbed). Over a given path length $\Delta\ell$, the power absorbed by the fluorescent pigment is just $a_{\text{pig}}(\lambda_E)\Delta\ell$ times the incident power within $\Delta\lambda_E$, and the power emitted through fluorescence into narrow band of wavelengths $\Delta\lambda$ is $b_F(\lambda_E \rightarrow \lambda)\Delta\lambda\Delta\ell$ times the incident power within $\Delta\lambda_E$. Combining these, we have

$$\eta(\lambda, \lambda_E) = \frac{\lambda}{\lambda_E} \frac{b_F(\lambda_E \rightarrow \lambda)\Delta\lambda}{a_{\text{pig}}(\lambda_E)}.$$

The usefulness of the quantum efficiency resides in the fact that it is often nearly independent of λ_E , i.e., as long as a photon is absorbed (with $\lambda_E < \lambda$) it has a given probability of causing a fluorescent emission regardless of its wavelength. This happens because fluorescence is the emission of a photon resulting from a transition from a given energy state in the molecule to a given lower state. Above the upper state, there are many higher energy states (or even bands of closely spaced energy states) and the absorbed energy is cascaded to lower energy states through radiation-less transitions (producing heat) ending at the

upper state of the fluorescence. The molecule ends up in the upper state regardless of the energy at which the cascade started. Of course, the exciting photon must have more (or an equal amount of) energy than the emitted photon.

When $\eta(\lambda, \lambda_E) = \eta(\lambda)$, the equations we developed can be rewritten in terms of the absorption coefficient of the fluorescing material and the quantum efficiency. For example, the upwelling irradiance becomes

$$E_u^{(0)}(z, \lambda) = \frac{\eta(\lambda)}{\lambda \Delta \lambda} \int_{\lambda_E < \lambda} d\lambda_E \left[\frac{E_0(0, \lambda_E) \lambda_E a_{\text{Pig}}(\lambda_E)}{2K_0(\lambda_E)} \right] \times \left\{ 1 - \frac{c(\lambda)}{K_0(\lambda_E)} \ell n \left(1 + \frac{K_0(\lambda_E)}{c(\lambda)} \right) \right\} \exp[-K_0(\lambda_E)z].$$

In the case of phytoplankton fluorescence at 683 nm (an example of which is Station C92 in Figure 6.5), if $a_{\text{Pig}}(\lambda_E)$ is taken to be the absorption coefficient of phytoplankton (roughly proportional to the concentration of Chlorophyll *a* in the water) and $\Delta \lambda \approx 25$ nm, then $\eta(683)$ is generally in the range 0.005 to 0.01, and the upwelling radiance and irradiance at 683 nm is roughly proportional to the concentration of Chlorophyll *a* in the water (for Station C92 in Figure 6.5, the upwelling irradiance at 683 nm can be explained with $\eta(683) = 0.0066$). The upper state of the Chlorophyll *a* fluorescence transition is the same state that the Chlorophyll *a* molecule must be excited for photosynthesis to occur. Thus, photosynthesis and fluorescence are competing processes.

6.4.12.3 Raman Scattering

Raman scattering by water is typically a far more important inelastic process than fluorescence for two reasons: (1) it is *always* present in natural waters; and (2) except in specific regions of the spectrum (e.g., near 683 nm) it is almost always much larger than fluorescence. To evaluate the contribution of Raman scattering to the light field we need to evaluate the integral

$$Q(z, u, \phi, \lambda) = \int_{\lambda_E < \lambda} d\lambda_E \int_0^{2\pi} d\phi' \int_{-1}^1 \beta_R(z, u' \rightarrow u; \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda) L(z, u', \phi', \lambda_E) du',$$

where β_R is the inelastic volume scattering function and λ_E is the excitation wavelength. For Raman scattering the excitation is from a narrow band of wavelengths $\Delta \lambda_E$ so, assuming $L(z, u', \phi', \lambda_E)$ varies slowly with λ_E , it can be removed from the λ_E integration (i.e., it's assumed to be constant over λ_E), and

$$Q(z, u, \phi, \lambda) = \int_0^{2\pi} d\phi' \int_{-1}^1 L(z, u', \phi', \lambda_E) du' \int_{\Delta \lambda_E} \beta_R(z, u' \rightarrow u; \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda) d\lambda_E.$$

The λ_E integration was developed in Chapter 5 (Section 5.2.3) with the result

$$\int_{\Delta\lambda_E} \beta_R(z, u' \rightarrow u; \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda) d\lambda_E = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda)(1 + 0.55 \cos^2 \alpha)$$

where λ is the wavelength of the scattered photon and λ_E the wavelength at the center of the excitation band, are related by

$$\lambda_E = \frac{\lambda}{1 + 3.357 \times 10^{-4} \lambda},$$

with λ in nm. The angle α is the scattering angle:

$$\cos \alpha = u'u + \sqrt{1 - u'^2} \sqrt{1 - u^2} \cos(\phi' - \phi).$$

The quantity $b_R(\lambda_E \text{ Band} \rightarrow \lambda)$ is the Raman scattering coefficient for the entire Raman band (of width $\Delta\lambda_E$) into λ . It has units m^{-1} and is given by

$$b_R(\lambda_E \text{ Band} \rightarrow \lambda) = 2.61 \times 10^{-4} (589/\lambda)^{4.8}, \quad (6.68)$$

with λ in nm.

Let's begin by calculating the Raman contribution to the upward radiance propagating toward the zenith $L^{(0)}(u' \rightarrow u = -1, \phi' \rightarrow \phi, \lambda_E \rightarrow \lambda)$, for which $\beta_R = 0.0673 b_R(1 + 0.55u'^2)$. The source function is then

$$Q(z, u, \phi, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) \int_0^{2\pi} d\phi' \int_{-1}^1 (1 + 0.55u'^2) L(z, u', \phi', \lambda_E) du'.$$

If we define the average of the cosine squared of the light field at λ_E to be

$$\langle u^2 \rangle_E = \frac{\int_0^{2\pi} d\phi \int_{-1}^1 u^2 L(z, u, \phi, \lambda_E) du}{E_0(z, \lambda_E)},$$

then,

$$Q(z, u, \phi, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) [1 + 0.55 \langle u^2 \rangle_E] E_0(z, \lambda_E).$$

Given that $L(z, u, \phi, \lambda_E)$ is strongly peaked near the direction of propagation of the incident solar beam, we expect that $\langle u^2 \rangle_E \approx u_{0w}^2$.²⁹ As in the case of fluorescence, we assume the scalar irradiance can be written

$$E_0(z, \lambda_E) = E_0(0, \lambda_E) \exp[-K_0(\lambda_E) z]$$

²⁹Detailed Monte Carlo simulations suggest that an excellent approximation to $\langle u^2 \rangle_E$ is μ_d^2 at λ_E .

with a depth-independent $K_0(\lambda_E)$. Inserting this into Eq. (6.66) and carrying out the indicated integration yields

$$\begin{aligned}
 L_u^{(0)}(z, u = -1, \lambda) &= 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) (1 + 0.55 \langle u^2 \rangle_E) \\
 &\quad \times \int_z^\infty E_0(z', \lambda_E) \exp[-c(\lambda) z'] dz', \\
 &= \left[\frac{0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda)}{c(\lambda) + K_0(\lambda_E)} (1 + 0.55 \langle u^2 \rangle_E) \right] \\
 &\quad \times E_0(0, \lambda_E) \exp[-K_0(\lambda_E) z],
 \end{aligned} \tag{6.69}$$

where we have assumed $c(\lambda)$ and $\langle u^2 \rangle_E$ are independent of depth.

If we carry out the same analysis for the upwelling radiance in any direction $\mu = -u$, due to the solar beam directed toward (u_{0w}, ϕ_0) , we find

$$\begin{aligned}
 L_u^{(0)}(z, \mu, \phi, \lambda) &= 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) \left(\frac{\exp[cz/\mu]}{\mu} \right) \int_z^\infty dz' \exp[-cz'/\mu] \\
 &\quad \times \int_0^{2\pi} d\phi' \int_{-1}^1 du' (1 + 0.55 \cos^2 \alpha) L(z', u', \phi', \lambda_E),
 \end{aligned} \tag{6.70}$$

with $\cos \alpha = -u'\mu + \sqrt{1 - u'^2} \sqrt{1 - \mu^2} \cos(\phi' - \phi)$. Carrying out the solution requires knowledge of the complete radiance distribution, $L(z', u', \phi', \lambda_E)$ at the excitation wavelength rather than just the scalar irradiance; however, if we only wanted the upwelling irradiance, significant simplification occurs. The upwelling irradiance is

$$\begin{aligned}
 E_u^{(0)}(z, \lambda) &= \int_0^{2\pi} d\phi \int_0^1 \mu L_u^{(0)}(z, \mu, \phi, \lambda) d\mu \\
 &= 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) \int_0^{2\pi} d\phi \int_0^1 d\mu \int_z^\infty dz' \exp[-c(z' - z)/\mu] \\
 &\quad \times \int_0^{2\pi} d\phi' \int_{-1}^1 du' (1 + 0.55 \cos^2 \alpha) L(z', u', \phi', \lambda_E),
 \end{aligned}$$

To evaluate this, first write

$$\begin{aligned}
 \cos^2 \alpha &= u'^2 \mu^2 + \frac{1}{2} (1 - u'^2) (1 - \mu^2) \\
 &\quad - 2u'\mu \sqrt{1 - u'^2} \sqrt{1 - \mu^2} \cos(\phi' - \phi) \\
 &\quad + \frac{1}{2} (1 - u'^2) (1 - \mu^2) \cos 2(\phi' - \phi),
 \end{aligned}$$

then carry out the integration over ϕ . This results in

$$E_u^{(0)}(z, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) \int_0^1 d\mu \int_z^\infty dz' \exp[-c(z' - z)/\mu] \\ \times \int_0^{2\pi} d\phi' \int_{-1}^1 du' \left(1 + 0.55[u'^2 \mu^2 + \frac{1}{2}(1 - u'^2)(1 - \mu^2)] \right) L(z', u', \phi', \lambda_E).$$

Now recall the definition of the average cosine squared of the light field at λ_E , $\langle u^2 \rangle_E$, and this simplifies to

$$E_u^{(0)}(z, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) \int_0^1 d\mu \int_z^\infty dz' \exp[-c(z' - z)/\mu] \\ \times \left(1 + 0.55[\langle u^2 \rangle_E \mu^2 + \frac{1}{2}(1 - \langle u^2 \rangle_E)(1 - \mu^2)] \right) E_0(z', \lambda_E). \quad (6.71)$$

Finally, assuming $E_0(z, \lambda_E) = E_0(0, \lambda_E) \exp[-K_0(\lambda_E) z]$ and that $\langle u^2 \rangle_E$ is independent of z , we have

$$E_u^{(0)}(z, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) E_0(0, \lambda_E) \exp[-K_0(\lambda_E) z] \\ \times \left[\int_0^1 \mu d\mu \frac{1 + 0.55[\langle u^2 \rangle_E \mu^2 + \frac{1}{2}(1 - \langle u^2 \rangle_E)(1 - \mu^2)]}{c(\lambda) + \mu K_0(\lambda_E)} \right]. \quad (6.72)$$

The integral in this equation can be carried out in closed form.³⁰ A similar procedure gives

$$E_d^{(0)}(z, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) E_0(0, \lambda_E) \\ \times \int_0^1 u du \frac{1 + 0.55[\langle u^2 \rangle_E u^2 + \frac{1}{2}(1 - \langle u^2 \rangle_E)(1 - u^2)]}{c(\lambda) - u K_0(\lambda_E)} (\exp[-K_0(\lambda_E) z] - \exp[-cz/u]). \quad (6.73)$$

We test the efficacy of these equations, by comparing Raman-excited radiances and irradiances computed using them to the results of exact (Monte Carlo) computations in the following examples.

³⁰First, factor out $K_0(\lambda_E)$ and write the remaining integral as

$$I_E \triangleq \int_0^1 x dx \frac{1 + c_1[c_2^2 + \frac{1}{2}(1 - c_2^2)(1 - x^2)]}{c_3 + x},$$

where $c_1 = 0.55$, $c_2 = \langle u^2 \rangle_E$, and $c_3 = c(\lambda)/K_0(\lambda_E)$. Then,

$$I_E = 1 + c_1 \left[\frac{1}{3} + (1 - 3c_2^2) \left(\frac{c_3}{4} - \frac{c_3^2}{2} \right) \right] + c_3 \left[1 + (1 - c_3^2) \frac{c_1}{2} - (1 - 3c_3^2) \frac{c_2^2}{2} \right] \ell n \left(\frac{c_3}{1 + c_3} \right)$$

Finally,

$$E_u^{(0)}(z, \lambda) = 0.0673 b_R(\lambda_E \text{ Band} \rightarrow \lambda) E_0(0, \lambda_E) \exp[-K_0(\lambda_E) z] \frac{I_E}{K_0(\lambda_E)}.$$

Table 6.10: Exact values of the Raman scattering contribution to $L_u(0, \mu = 1, \lambda)$ in $\text{mW}/\text{cm}^2 \mu\text{mSr}$ and $E_u(0)$ in $\text{mW}/\text{cm}^2 \mu\text{m}$ at 550 nm for a pure sea water body, compared with values of $L_u^{(0)}(0, \mu = 1, \lambda)$ and $E_u^{(0)}(0)$ computed using Eqs. (6.69) and (6.72), respectively, with $\langle u^2 \rangle_E = \langle \mu_d(\lambda_E) \rangle^2$. The values labeled “Approx.” are computed by replacing K_0 with K_d and using Eq. (6.74) to estimate E_0 from $E_d(0)$ and $E_u(0)$ at λ_E , and $\langle u^2 \rangle_E = u_{0w}^2$.

θ_0	L_u (Exact)	L_u Eq. (6.69)	L_u (Approx.)	E_u (Exact)	E_u Eq. (6.72)	E_u (Approx.)
0°	0.1125	0.1114	0.1127	0.3223	0.3221	0.3159
60°	0.0533	0.0531	0.0520	0.1699	0.1697	0.1630

Example 1: Computation of Raman Component of $L_u(0)$ and $E_u(0)$ for Pure Seawater

Consider a water body consisting of pure sea water below an aerosol-free atmosphere. Let us compare the Raman contributions to L_u and E_u just beneath the surface to exact computations at 550 nm for which $\lambda_E = 464$ nm. We take the actual values of $K_0(\lambda_E)$ (actually, its depth mean) and $E_0(0, \lambda_E)$ along with $\langle u^2 \rangle_E = \langle \mu_d(\lambda_E) \rangle^2$ from exact Monte Carlo computations. Table 6.10 compares the exact values of $L_u^{(0)}(0, \mu = 1, \lambda)$ and $E_u^{(0)}(0)$ with those computed using Eqs. (6.69) and (6.72), respectively. Clearly, when the correct excitation parameters are used in the equations, excellent results are obtained for L_u and E_u ; however, typically if any experimental measurements are made at the excitation wavelength, they will be measurements of $E_d(z)$ and $E_u(z)$ (or $L_u(z)$), so usually $E_0(0)$ and K_0 will be unknown. In that case the scalar irradiance can be reasonably well predicted from

$$E_0(0) \approx D_0 E_d(0) + 2.5 E_u(0), \quad (6.74)$$

while if $L_u(0)$ is measured, $E_u(0)$ above is replaced by $\pi L_u(0)$, which assumes the upwelled radiance distribution at λ_E is totally diffuse. The value of $\mu_d(\lambda_E)$ is usually unknown, so we use the approximation $\langle u^2 \rangle_E = u_{0w}^2$. With these approximations, the columns labeled “Approx.” in Table 6.10 are obtained. This suggests that Eqs. (6.69) and (6.72) provide robust estimates of $L_u^{(0)}(0, \mu = 1, \lambda)$ and $E_u^{(0)}(0)$, respectively. Note that the depth dependence of these two quantities is given by the surface value multiplied by $\exp[-K_0(\lambda_E)z]$, i.e., pure exponential decay governed by the *excitation* wavelength.

Example 2: Efficacy of the Computation the Raman Component of $E_d(z, \lambda)$.

What about $E_d(z, \lambda)$? Equation (6.73) shows that the depth dependence of E_d is more complex than E_u or L_u . Is this borne out in exact computations? We note that this equation determines the contribution to E_d from the initial Raman scattering, and is thus zero at $z = 0$; however, the actual $E_d(0)$ at λ includes Raman-induced $E_u(0)$ reflected downward from the interface. For comparison of Eq. (6.73) with exact computations, this reflected component must be removed from the exact values of the Raman contribution to $E_d(z)$. How can this component of $E_d(z, \lambda)$ be estimated? We proceed as follows:

first, we assume that the upward Raman radiance $L_u^{(0)}(z, \mu, \phi, \lambda)$ at $z = 0$ is isotropic ($L^{(0)}$ independent of μ and ϕ);³¹ next, we compute the Fresnel-reflected component of this radiance; and finally, we determine the decay characteristics of this reflected component with depth.

For isotropic upward radiance L_0 , the reflected irradiance at $z = 0$ is

$$E_d(0) = 2\pi L_0 \int_0^1 ur_f(u) du = 2E_u(0) \int_0^1 ur_f(u) du,$$

where $E_u(0) = \pi L_0$ and r_f is the Fresnel reflectance of the air-water interface *from below*. The downward radiance $L_d(z, u)$ is

$$L_d(z, u) = L_0 r_f(u) \exp(-cz/u),$$

so the downward irradiance is

$$E_d(z) = 2\pi L_0 \int_0^1 ur_f(u) \exp(-cz/u) du,$$

and

$$\frac{E_d(z)}{E_d(0)} = \frac{\int_0^1 ur_f(u) \exp(-cz/u) du}{\int_0^1 ur_f(u) du} \triangleq \exp(-K_d^D z).$$

Figure 6.53 provides K_d^D/c as a function of the optical depth $\tau = cz$. Note that except very near the surface, or at large τ , $K_d^D \approx 2c$. The reflected fraction of $E_u(0)$, i.e., $\rho_{\text{Diffuse}} \triangleq E_d(0)/E_u(0)$ for the above assumptions is ~ 0.48 ; however, exact computations yield slightly larger values because $E_u^{(0)}$ is not totally diffuse. Using the above assumptions, the total downward irradiance at depth z is

$$[E_d^{(0)}(z, \lambda)]_{\text{Total}} = E_d^{(0)}(z, \lambda) + \rho_{\text{Diffuse}} E_u^{(0)}(0, \lambda) \exp[-K_d^D(\lambda) z]. \quad (6.75)$$

We can now examine the accuracy of the downwelling Raman irradiance for the same cases examined above, namely a particle-free water body and an aerosol-free atmosphere at 550 nm for $\theta_0 = 0$ and 60° . We compare the values of $E_d^{(0)}(z, \lambda)$, computed from the exact $[E_d(z, \lambda)]_{\text{Total}}$, ρ_{Diffuse} , and $E_u^{(0)}(0, \lambda)$ using Eq. (6.75) and Figure 6.53, with those from Eq. (6.73). The resulting “exact” Raman-induced $E_d(z)$ at 550 nm is provided in Figure 6.54 (symbols) along with the computations resulting from Eq. (6.73) (line). Clearly, Eq. (6.73) provides an excellent approximation to the Raman-induced downwelling irradiance.

³¹Note that the author often uses both the term “*totally* diffuse” as well as the term “isotropic” when referring to radiance that is independent of direction.

Example 3: Depth Dependence of L_u and E_d

For an example of the impact of Raman scattering within the water, we again consider a water body that is free of scattering and absorbing materials, i.e., consisting only of *pure* (sea) water, as above. The *exact* solution to the radiative transfer equation at 550 nm (with $\theta_0 = 37^\circ$) yields vertical profiles of K_d , K_u and R shown in Figure 6.55. Also shown is K_d at 464 nm, the excitation wavelength. Note that with increasing depth both K_d and K_u at 550 both tend toward K_d at 464 nm, with K_u approaching more rapidly than K_d . At shallow depths, K_d is relatively unaffected by Raman scattering. The figure also shows that R is very strongly influenced by the presence of Raman scattering, with a value approaching the order of unity at great depth.

We now apply our model for $E_d^{(0)}(z, \lambda)$ and $E_u^{(0)}(z, \lambda)$ to see how well it reproduces the exact computations of R in Figure 6.55. The solid lines on the figures are predictions of the model above. To do these predictions, we need $E_d(0, 550)$ and $E_u(0, 550)$ for *both* the elastic and Raman components. For the elastic components, we note that in a medium that scatters with a phase function that is nearly isotropic (not strongly forward scattering), as Rayleigh scattering does, the QSSA is not appropriate for the computation of $R(0)$ in $E_u(0) = R(0)E_d(0)$, so the exact value of $R(0)$ from the Monte Carlo simulation was used instead. For the elastic $E_d(0)$, we used $E_d(0) = t_f u_0 F_0 \exp(-\tau_r/2u_0)$ (Eq. (6.34) with $x = 1$), where t_f is the Fresnel transmittance of the air-water interface for incidence angle θ_0 and τ_r is the optical depth of the atmosphere. These gave $E_d^{\text{Elastic}}(0, 550) = 138.19 \text{ mW/cm}^2\mu\text{m}$, and $E_u^{\text{Elastic}}(0, 550) = 0.8278 \text{ mW/cm}^2\mu\text{m}$. The elastic components both decay exponentially in depth with decay coefficient $K_d = (a + b)/u_{0w}$, which for a medium that scatters with a nearly isotropic phase function is more appropriate than the QSSA value of $(a + b_b)/u_{0w}$. We assume that we are provided with the downward irradiance at 464 nm, e.g., measurements of $E_d(z, \lambda_E)$, so that $K_d(464)$ is known. The Raman components are computed using the formulas above but with $K_0(\lambda_E)$ replaced with $K_d(464)$. Also, $E_0/F_0 = 0.94$ at the excitation wavelength was taken from the exact computation (Eq. (6.74) gave 0.975). These resulted in

$$\begin{aligned} E_u^{\text{Total}}(z, 550) &= 0.8278 \exp(-0.0656 z) + 0.2647 \exp(-0.0146 z) \\ E_d^{\text{Elastic}}(z, 550) &= 138.19 \exp(-0.0656 z), \end{aligned} \tag{6.76}$$

in $\text{mW/cm}^2\mu\text{m}$, and, because of the integral in $E_d^{(0)}$ (Eq. (6.73)) must be carried out numerically, no simple analytic formula is available for $E_d^{\text{Raman}}(z, 550)$, only a table of values. From these, it is a simple matter to compute the $K(z)$'s and $R(z)$, as presented as the solid lines in Figure 6.55. The agreement here between the exact and approximate values of these quantities is remarkably close; however, much of the good agreement is due to the fact that in a particle-free water body the K 's for the elastic component are almost independent of depth — which is *not* true in general (Figure 6.9). However, the results do

demonstrate the efficacy of Eqs. (6.72) and (6.73) in computing the Raman-induced light fields.³²

Example 4: Transition from SSA to QSSA

To effect the transition to the QSSA approximation to the Raman contributions, the simplest approximation is to replace $c(\lambda)$ by $a(\lambda) + b_b(\lambda)$ in Eqs. (6.69)-(6.73); however, unfortunately there is no simple QSSA formula for $E_0(z)$ or $K_0(z)$. Figure 6.13, which compares the vertical profiles of K_0 , K_d , K_u , and K_L shows that K_0 varies much more rapidly with depth than any of the other K 's except K_u . Thus, how do we estimate K_0 or the scalar irradiance? K_u would be a good approximation, but there is no accurate QSSA estimate of K_u as there is for K_d (recall that in Eqs. (6.21) the estimate for K_u is not very good). The most obvious thing we can do is replace K_0 with K_d , or some averaged K_d ; however, if we know $\omega_0(\lambda_E)$ we can do a little better. Using all of the simulations discussed in Figure 6.31, a course relationship between K_d and K_0 is evident. Figure 6.58 shows K_0/K_d , evaluated at $\tau = 0.5$ used in Figure 6.31, as a function of ω_0 for the 360 simulations. The optical depth $\tau = 0.5$ rather than $\tau = 0$ is used as a compromise for improving the $K_0 \approx K_d$ assumption, as K_d at $\tau = 0.5$ is very close to $K_d(0)$ for which there is a good QSSA relationship. The line on the figure is the least-squares fit:

$$\frac{K_0}{K_d} = 1 + 0.2538\omega_0 - 2.3216\omega_0^2 + 5.2589\omega_0^3 - 3.9837\omega_0^4. \quad (6.77)$$

This fit has an R^2 of 0.9822, and for all the points shown, the average (absolute) error was 1.81% with a maximum error of 13.59% (at the point where the $K_0 \approx K_d$ assumption is in error by almost a factor of 2), and can be used to improve the estimate of K_0 given the IOPs at λ_E . Given Eqs. (6.69)-(6.73), the transition to the QSSA ($c(\lambda) \rightarrow a(\lambda) + b_b(\lambda)$), and Eq. (6.77), we have a complete set of equations to provide a reliable estimate of the Raman contribution to $L_u(0)$ and $E_u(0)$.

Example 5: Efficacy of QSSA in Computation of the Raman Contribution to AOPs

Just how effective is the transition from the SSA to the QSSA? Consider a more realistic example, e.g., a water body with particles producing a strongly-forward scattering phase function, below an atmosphere with aerosols. To estimate $L_u(0)$ in this case the first line in Eq. (6.69) must be used because $K_0(\lambda_E)$ depends strongly on depth, especially when the solar zenith angle is large. We employed a bio-optical model similar to that in the next section, but tuned to match experimental measurements in clear waters off Hawaii, to provide exact calculations of the light field at λ_E . A phytoplankton pigment concentration of 0.3 mg/m³ was used in the simulation along with a solar zenith angle

³²In situations where $K_0(z, \lambda_E)$ is not constant, the Raman components can still be computed by inserting the actual depth variations of $E_0(z, \lambda_E)$ into Eq. (6.71) and a similar equation for $E_d^{(0)}(z, \lambda)$.

60° . Entering the required quantities at the excitation wavelength into Eq. (6.69), the integration was carried out to estimate $L_u(0)$. Table 6.11 provides the results for 450 and 550 nm (excitation at 391 and 464 nm, respectively) computed using Eq. (6.69) directly (the single scattering approximation) and by replacing c by $a + b_b$ in Eq. (6.69) (the quasi-single scattering approximation, which approximately accounts for the multiple scattering of the *Raman-generated* radiance. Note that, in the quasi-single scattering approximation, $a(\lambda) + b_b(\lambda) = u_{0w}K_d(\lambda)$, and $K_d(\lambda)$ could be derived from $E_d(z, \lambda)$, i.e., measurements of the downwelling irradiance at the emission wavelength which near the surface are only weakly influenced by Raman scattering. The fraction of the Raman contribution to the total radiance at the surface turned out to be 7.7% and 14.5%, respectively, at 450 and 550 nm. The error using the QSSA compared to the SSA is reduced by a large factor

Table 6.11: Contribution to $L_u(0, \lambda)$ “(MC)” in $\text{mW}/\text{cm}^2 \mu\text{m Sr}$ and at 450 and 550 nm for a modeled water body with a pigment concentration of $0.3 \text{ mg}/\text{m}^3$, compared with values of $L_u(0, \lambda)$ computed using Eq. (6.69) with $\langle u^2 \rangle_E$ replaced by μ_d^2 . The solar zenith angle was 60° . “Model” refers to using Eq. (6.69) directly with c , or replacing $c(\lambda)$ by $a(\lambda) + b_b(\lambda)$, to transition to the QSSA.

$\lambda \rightarrow$	450 nm		550 nm	
Model	L_u (MC)	L_u Eq. (6.69)	L_u (MC)	L_u Eq. (6.69)
$c(\lambda)$	0.07164	0.03590	0.03461	0.02365
$a(\lambda) + b_b(\lambda)$	0.07164	0.06517	0.03461	0.03407

to an error of 9% at 450 nm and 1.6% at 550 nm. The larger error at 450 nm is to be expected because the single scattering albedo is significantly larger there than at 550 nm, so the quasi-single scattering approximation is less effective. Note that in this example the scalar irradiance $E_0(z, \lambda_E)$ was used directly in Eq. (6.69) rather than the $K_d \rightarrow K_0$ approximation in Exercise 4. This enabled the assessment of the SSA \rightarrow QSSA transition effectiveness alone.

Example 6: Assessment of Raman Contribution to AOP Measurements.

Even if we do not know ω_0 and cannot use Eq. (6.77), the lack of high accuracy in the $K_0 \approx K_d$ assumption does not mean that the approximate Raman contribution is not valuable. One can for example use it to qualitatively understand the influence of the various constituent concentrations on the Raman contribution in a given situation. For example, if dissolved organic material, which absorbs in a manner that increases exponentially with *decreasing* wavelength (i.e., absorbs weakly in the green, more strongly in the blue and even more strongly still in the ultraviolet) is added to the water, the Raman contribution will decrease with increasing concentration because $K_0(\lambda_E)$ will increase with increasing concentration. Thus, the Raman contribution to the upwelling radiance and irradiance at the surface in the blue and green will be significantly lower than that for pure sea water

shown in Figure 6.56. There will be a smaller decrease in the red. Similarly, phytoplankton, which absorb primarily in the blue will decrease the Raman contribution in the green, etc., by increasing $K_0(\lambda_E)$.

Another example of the utility of the approximate formulas is to assess the magnitude of the Raman contribution in experimental measurements. Consider Station C92 in Figures 6.5 and 6.6. How large is the Raman contribution to R at 550 nm (close to the maximum in R)? At this station the Sun was near the zenith, so a good approximation to $c(\lambda) \rightarrow a(\lambda) + b_b(\lambda)$ is $K_d(\lambda)$, i.e., $u_{0w} \approx 1$. The scalar irradiance just beneath the surface at λ_E was not measured, but can be estimated from Eq. (6.74) with $E_d(0)$ and $E_u(0)$ determined from extrapolation of the actual irradiance measurements to the surface. We have no way of precisely knowing $K_0(\lambda_E)$, so the best we can do is replace it with $K_d(\lambda_E)$. This exercise suggests that Raman scattering contributes approximately only 1.5% to $R(0)$ at 550 nm, i.e., in the absence of Raman scattering, $R(0)$ would be 1.5% smaller.

Example 7: Raman Influence on Angular Distribution of L_u

Finally, from a remote sensing perspective, L_u is the most important quantity to estimate from the IOPs. It can be done with reasonable accuracy through a combination of Eqs. (6.27), (6.31), (6.34)-(6.38), (6.69), and (6.77). The presence of Raman scattering will of course also change the angular distribution of the the total upwelling radiance at $z = 0$. An estimate of the influence of Raman scattering can be found as follows. First, in Eq. (6.70), because of the strong forward scattering at λ_E , approximate $L(z', u', \phi', \lambda_E)$ by $E_0(0, \lambda_E) \delta(u' - u_{0w}) \delta(\phi' - \phi_0) \exp[-K_0(\lambda_E)z']$. Then after carrying out the integrations over the primed variables, we have

$$L_u^{(0)}(z, \mu, \phi, \lambda) = \left[\frac{0.0673 b_R(\lambda_E \rightarrow \lambda) \Delta \lambda_E}{c(\lambda) + K_0(\lambda_E) \mu} (1 + 0.55 \cos^2 \alpha_0) \right] E_0(0, \lambda_E) \exp[-K_0(\lambda_E) z],$$

where $\cos \alpha_0 = -\mu u_{0w} + \sqrt{1 - \mu^2} \sqrt{1 - u_{0w}^2} \cos(\phi - \phi_0)$. The angular distribution of $L_u^{(0)}$ at the surface is then,

$$\frac{L_u^{(0)}(0, \mu, \phi, \lambda)}{L_u^{(0)}(0, 1, \phi, \lambda)} = \left[\frac{1 + 0.55 \cos^2 \alpha_0}{1 + 0.55 \langle u^2 \rangle_E} \right] \left[\frac{c(\lambda) + K_0(\lambda_E)}{c(\lambda) + K_0(\lambda_E) \mu} \right],$$

where $\langle u^2 \rangle_E$ is as usual to be approximated by u_{0w}^2 .

Example 8: Influence of Raman Scattering on L_u Across the Visible Spectrum

Finally, we have seen in Example 3 that the Raman component to $E_u(0, 550)$ accounts for about 25% of the total (Eq. (6.76)). How does the influence of Raman scattering on E_u and, more importantly, L_u vary across the spectrum? Again, we consider the same case

as in Example 3, a medium consisting only of pure sea water bounded by an aerosol-free atmosphere, because for such a water body (and atmosphere) all of the IOPs are precisely known. Figure 6.56 provides exact computations (in the scalar approximation) of the contributions to the zenith-propagating radiance (left panel) from elastic and inelastic (Raman) scattering. The right panel shows the fraction of the total radiance contributed by Raman scattering by shorter wavelengths. Note that for $\lambda \gtrsim 500$ nm Raman scattering contributes about a quarter of the total. The precipitous drop in the Raman contribution for $\lambda \lesssim 475$ nm is mostly due to the large drop in the solar irradiance (F_0) for wavelengths shorter than ~ 400 nm. The small peak in the Raman contribution near 430 nm corresponds to a similar peak in F_0 near 375 nm, the excitation wavelength (Chapter 8, Figure 8.7). How well do these calculations match real data? Figure 6.57 compares the calculated zenith-propagating radiance with measurements in the clear waters off Lanai, Hawaii. These data are from a complete set of spectral data of which those in Figure 6.1 are a subset. The radiances were extrapolated to the surface from a depth of 2 m using $L_u(z = 0, \lambda) = L_u(z = 2, \lambda) \exp[+2 \times \langle K_u(\lambda) \rangle]$, where $\langle K_u(\lambda) \rangle$ is the average value from over the depth range 2 to 6 m for the given wavelength. The agreement between the calculations and the measurements for $\lambda \gtrsim 490$ nm is truly remarkable. The lack of agreement over the 400 to 490 nm range is due to material in the water: particles (principally phytoplankton) and dissolved organic material.³³

In our treatment of the effects of Raman scattering on the in-water light field, we have neglected polarization of the scattering. Including polarization, a solution in lowest order, i.e., Eq. (6.66), is still straightforward but requires knowing the polarization state of the excitation radiance. If we only want the polarization state of the upwelling radiance just beneath the water surface, e.g., for remote sensing, we can assume as a first approximation that the excitation is just due to the (unpolarized) solar beam. Then the solution for the Stokes vector associated with the upwelling radiance is similar to that provided in Section 2.11.3 for Rayleigh scattering.³⁴ The details of the solution are left to the reader.

6.4.13 Examples of Modeling with QSSA

We complete this Section by providing some examples using the QSSA for estimating the AOPs given models of the IOPs. First, we focus on two of the Case 1 water stations

³³For $\lambda \gtrsim 490$ nm, the agreement here is a little misleading because particle scattering of solar radiation clearly provides additional radiance. However, at this location the Chlorophyll concentration is ~ 0.10 to 0.15 mg/m³ for which the Raman fraction drops to 10-15 % for $\lambda \gtrsim 490$ nm because of phytoplankton absorption at shorter wavelengths. This “missing” (compared to a particle-free medium) radiance is replaced by the particle backscattering contribution to L_u .

³⁴Recall that Raman and Rayleigh scattering share a common phase function, differing only in the value of the depolarization factor.

described earlier in Section 6.3.1: the Hawaii station, Figures 6.3 and 6.4; and C70 on Figures 6.6 and 6.5. We used the formulas displayed in Figure 6.35 to estimate the elastic contribution to L_u and Eq. (6.47) to estimate R resulting from elastic scattering. The Raman contributions were estimated using Eq. (6.69) for L_u and Eq. (6.72) for E_u . Where needed, the scalar irradiance at the Raman excitation wavelength (λ_E) was estimated from E_d and E_u (or L_u) at λ_E using Eq. (6.74), and K_0 at the excitation was estimated from $K_d(\lambda_E)$ and $\omega_0(\lambda_E)$ using Eq. (6.77). The K_d 's were estimated from Eq. (6.46) using for $\langle C'_K \rangle_X$ the average of its value for $B_p = 0.0020$ and $B_p = 0.0181$ from Table 6.8 for $X = 1$.

The IOPs required for input to the QSSA are described in detail in Chapter 5 and for convenience are summarized in Figure 6.59. Note that the Raman scattering coefficient has been included in the absorption coefficient. This is because at Raman scattering *from* a given wavelength λ is a *loss* that is indistinguishable from absorption at λ . In the formulae in Figure 6.59, C is in mg/m^3 , and the resulting quantities are in their usual units, e.g., $a_p(\lambda)$ is in m^{-1} , etc. The quantities $a_{p1}(\lambda)$ and $b_{p1}(550)$ are the values of $a_p(\lambda)$ and $b_p(550)$ (in m^{-1}) for $C = 1 \text{ mg}/\text{m}^3$. All of the quantities in the figure are fixed with the exception of f_{ap} and f_{bp} . Recall that the formulas for a_p and b_p are statistical products, e.g., $b_p(550)$ for $C = 1 \text{ mg}/\text{m}^3$ is in the range 0.12-0.45 m^{-1} for Case 1 waters. Here, this parameter is fixed at 0.30 m^{-1} , and the statistical spread is accommodated by varying f_{bp} . Similar comments apply to $a_{p1}(\lambda)$ and f_{ap} . So, given C , the only free parameters in the IOP model are f_{ap} and f_{bp} . These parameters are varied to try to obtain a good fit to the AOP data (within an acceptable range based on the statistical properties of the aggregate IOP measurements).

The modeled AOPs for the Hawaii station are L_u and K_d . Although C was not measured during the experiment, we take $C = 0.10 \text{ mg}/\text{m}^3$ – the coarse historical average. For this data we took $f_{ap} = 0.67$ and $f_{bp} = 1.10$.³⁵ Comparison of the estimated and measured AOPs are shown in Figure 6.60. The results are excellent, but the fits for both L_u and K_d do indicate that the absorption coefficient is a little too low in the region between 400 and 450 nm: the modeled L_u is too high and K_d is too low.³⁶ Also, the absorption coefficient is either a little too low, or the scattering coefficient a little too high in the region from 500 to 650 nm. However, considering the coarseness of the IOP model, e.g., the absence of any spectral features in B_p , which are known to exist, the agreement is remarkably good.

In the case of Station C70, the modeled AOPs are R and K_d . Here C was measured³⁷

³⁵These parameters were chosen by trial and error to render the spectral variation of the model close to that of the field data. There was no attempt to optimize the fit of the model to the data.

³⁶Note that $K_d \approx (a + b_b)/\langle u_{0w} \rangle_0$ and typically $a \gg b_b$, so K_d depends mostly in a . In contrast, L_u and $R \sim b_b/(a + b_b)$ and so are strongly influenced by b_b even if $a \gg b_b$.

³⁷The measurement of C here was actually of Chlorophyll *a* plus Phaeophytin *a* determined by fluorometry. The more recent measurements leading to the formulas for a_p and b_p as a function of C were established with C representing Chlorophyll *a* or total Chlorophyll measured using High Performance Liquid

to be 2.4 mg/m³. The parameter values for this station are $f_{ap} = 1.44$ and $f_{bp} = 1.33$. The resulting comparisons between the measurements and the model are also shown in Figure 6.60. The agreement for both K_d and R is excellent for $\lambda \lesssim 570$ nm. For $\lambda \gtrsim 600$ nm it appears that the modeled R is about a factor of 2 too high, however, these experimental data were analyzed before the importance the self-shading of in-water optical instruments (especially in the red) was well understood. Although self-shading could not account for all of the difference seen here, it could account for 30-40% or more of the difference.

As another example of the use of these equations, we examine the ratio $L_u(440)/L_u(550)$ as a function of the chlorophyll concentration. In forming the ratio, it was assumed that $\theta_0 = 37^\circ$. The ratio is provided in Figure 6.61 for the nominal IOPs ($f_{ap} = f_{bp} = 1$) as well as for those used for C70 and Hawaii in Figure 6.60. Also included on the figure is the original empirical algorithm for retrieval of pigment concentration from water color. The latter was based on experimental measurements of L_u and C .³⁸ The results show that the IOP model is clearly comparable with experimental measurements, but shows a strong dependence of the predicted ratio on f_{ap} . As would be expected from Figure 6.60, no single values for f_{ap} and f_{bp} appear to work well over the entire range of C in Figure 6.61.

Although a complete study of the applicability of the QSSA to the examination of experimental AOPs is beyond the scope of this work, it is hoped that the results presented in this section clearly demonstrate the value of the QSSA for the exploration of the influence of various water constituents on the AOPs.

6.5 Summary

In summary, in this chapter we have defined the AOPs of natural waters and provided examples, from direct experimental measurements and from radiative transfer simulations utilizing the IOPs, which are directly related to the concentration of the water's constituents (estimation of which is the ultimate goal of ocean color remote sensing). We then developed a simple model of radiative transfer (QSSA) and provided relationships from which the most important AOPs (from a remote sensing perspective), $L_u(0, \theta_{vw}, \phi_{vw})$ and $K_d(0)$, can be estimated given the IOPs. The contribution of inelastic processes (Raman scattering and fluorescence) can also be estimated in a simple manner, but these require the IOPs at the exciting wavelength(s) as well as the wavelength of interest. Of course, detailed radiative transfer computation of these quantities will always be superior in accuracy to

Chromatography (HPLC).

³⁸The reader is again reminded that in the earlier experiments C was actually of Chlorophyll *a* plus Phaeophytin *a* determined by fluorometry, while the IOPs used here correspond to C representing Chlorophyll *a* or total Chlorophyll measured using High Performance Liquid Chromatography (HPLC).

the relationships developed here (and validated and sometimes improved by comparison to detailed calculations); however, as the relationships between the IOPs and the constituent concentrations are rarely known with high accuracy, the simple relationships will often suffice as demonstrated in Section 6.4.13. We show in an appendix that these QSSA-based relationships make it possible to invert the AOPs to obtain some of the IOPs. Additional appendices provide alternative approaches to simple solutions to the radiative transfer equation, as well as examples of bio-optical algorithms based on measurements of the AOPs.

Finally, we have not discussed the influence of the bottom on L_u and K_d which is of interest in remote sensing of shallow coastal waters or inland lakes. We leave this area, which likely requires special techniques, to others, cogniscent of the fact that our analysis is valid only when the bottom has a negligible influence on L_u .

6.6 Appendix 1: Estimation of IOPs from AOPs

The existence of formulae relating the AOPs to the IOPs suggests the possibility of using AOP measurements to estimate some of the IOPs. Assume we are given measurements of the $E_d(z)$ and $E_u(z)$ at a given wavelength in a homogeneous water body and have somehow removed the inelastic contributions to the light field. The relevant equations are Eqs. (6.46)-(6.49):

$$\begin{aligned}
 \langle u_{0w} \rangle_0 \frac{\langle K_d \rangle_X}{c} &= 1 - \langle C'_K \rangle_X \omega_0 F, \\
 R(0) &= \sum_{n=1}^3 C'_{Rn} [QSSA'_R(u_0)]^n, \\
 QSSA'_R(\langle u_{0w} \rangle_0) &= \frac{\omega_0 B}{1 - \omega_0 F} I'_R(\langle u_{0w} \rangle_0), \\
 I'_R(\langle u_{0w} \rangle_0) &= \frac{I'_{Rp}(\langle u_{0w} \rangle_0) b_{bp} + I'_{Rw}(\langle u_{0w} \rangle_0) b_{bw}}{b_{bp} + b_{bw}}, \\
 &= \frac{I'_{Rp}(\langle u_{0w} \rangle_0) B_p b_p + I'_{Rw}(\langle u_{0w} \rangle_0) B_w b_w}{B_p b_p + B_w b_w}, \\
 F &= 1 - B \\
 B &= \frac{B_p b_p + B_w b_w}{b_p + b_w}
 \end{aligned} \tag{6.78}$$

From E_u and E_d we form $R(0)$ and $\langle K_d \rangle_X$. Then we have two known quantities. The water quantities b_{bw} and $I'_{Rw}(\langle u_{0w} \rangle_0)$ are determined given the wavelength and an estimate of

$D_0 = 1/\langle u_{0w} \rangle_0$ (e.g., see Figure 6.35). The unknown IOPs are the particle phase function (which provides I'_{Rp} and B_p), ω_0 , and c .³⁹

Clearly, there are many more unknowns than knowns. This means we need some additional assumptions reduce the number of unknowns. The obvious quantity to focus on for this purpose is the phase function. If we assume, as we have in much of this chapter, that the particle phase function is a modified Petzold (or Petzmas) with a given B_p , then $I'_{Rp}(\langle u_{0w} \rangle_0)$ is determined (Figure 6.44) reducing the unknowns to ω_0 and c .⁴⁰ This suggests that we should be able to obtain ω_0 and c , and thus all of the IOPs from $R(0)$ and $\langle K_d \rangle_X$. However, there is a pitfall. Note the nearness of $\langle C'_K \rangle_X$ to unity in Tables 6.6 and 6.7, i.e., with an error of only a few %, $\langle C'_K \rangle_X = 1$. When $\langle C'_K \rangle_X = 1$, we have

$$\langle u_{0w} \rangle_0 \frac{\langle K_d \rangle_X}{c} = 1 - \omega_0 F \quad \implies \quad a + b_b = \langle u_{0w} \rangle_0 \langle K_d \rangle_X$$

and

$$QSSA'_R(\langle u_{0w} \rangle_0) = \frac{\omega_0 B}{1 - \omega_0 F} I'_R(\langle u_{0w} \rangle_0) \quad \implies \quad b_b = \frac{\langle u_{0w} \rangle_0 \langle K_d \rangle_X QSSA'_R(\langle u_{0w} \rangle_0)}{I'_R(\langle u_{0w} \rangle_0)}.$$

Thus, the only IOPs that are derivable from the AOPs are a and b_b . This is not surprising; with a strongly forward scattering phase function, small angle scattering makes a significant contribution to b (and therefore c), but in radiative transfer small angle scattering is essentially equivalent to no scattering at all, so it should not affect the AOPs in a significant manner.⁴¹ Since $B = b_b/b$, one might think that it would be possible to determine b from the retrieved b_b , but remember that B is the integral of the phase function over the backward solid angles, and the phase function has been *assumed* not determined.

A similar retrieval of IOPs can be effected using the often-measured $L_u(z = 0, \mu = 1, \phi)$ along with $E_d(0)$. Recall that

$$\begin{aligned} \frac{L_u(0, \mu = 1, \phi; u_0, \phi_0)}{E_d(0)} &= QSSA'_L(u_0) = \frac{\omega_0}{4\pi(1 - \omega_0 F)} \frac{P_r(u_0 \rightarrow -1, \phi_0 \rightarrow \phi)}{u_0 + 1} \\ &= \frac{\omega_0 B}{(1 - \omega_0 F)} \left(\frac{1}{u_0 + 1} \right) \frac{P_r(u_0 \rightarrow -1, \phi_0 \rightarrow \phi)}{4\pi B} \\ &= \frac{b_b}{a + b_b} \left(\frac{1}{u_0 + 1} \right) \frac{P_r(u_0 \rightarrow -1, \phi_0 \rightarrow \phi)}{4\pi B}. \end{aligned}$$

Thus, replacing E_u by L_u in the measurements simply replaces $I_{R'}$ by P_r/B and a geometrical factor. The conclusions are then the same as above with the replacement of $QSSA'_R$

³⁹This should not be surprising, these are just the IOPs that appear in the radiative transfer equation, which determines the AOP's, given the incident radiance on the water surface.

⁴⁰Given ω_0 and c , $b_p = \omega_0 c - b_w$, and if we start from the bottom equation in Eqs. (6.78) and work our way upward, we find that all the parameters required to compute $\langle K_d \rangle_X$ and $R(0)$ are available.

⁴¹This, in fact, is the basis of the QSSA.

with $QSSA'_L$. Of course, as before, one still needs to assume a particle phase function to provide P_r/B . Again, having assumed a P_r , only a and b_b can be obtained with any confidence.

Although one can use these equations to estimate a and b_b , it is possible that significant error can occur because of large deviations of the AOPs from Eqs. (6.78) for some IOPs, e.g., Table 6.7 and Figures 6.33 and 6.45. Therefore we seek a better AOP \rightarrow IOP inversion technique that avoids these errors. In the next two sections we provide such a method.

6.6.1 Iterative Inversion Algorithm — Homogeneous Water Bodies

The basic idea of an iterative AOP \rightarrow IOP inversion algorithm is as follows. From the illumination on the sea surface and measurements of depth profiles of either $E_u(z)$ and $E_d(z)$ or $L_u(z)$ and $E_d(z)$, we must find IOPs that, when inserted into the radiative transfer equation, reproduce the measured AOP profiles within experimental error. In this brief sketch of the method, we limit the discussion to the simplest case: a homogeneous water body with no inelastic processes, or with the effects of inelastic processes removed, e.g., through the methods of Section 6.4.12.

Consider first the case where $E_u(z)$ and $E_d(z)$ are measured. Recall from Section 6.2 that

$$a(z) = \langle \mu(z) \rangle K_V(z), \quad \text{where} \quad K_V(z) = -\frac{d}{dz} \ell n [E_d(z) - E_u(z)].$$

We form $K_V(z)$ and $R(z)$ from the irradiance data, and would like to use K_V to find a ; however, we do not yet know $\langle \mu(z) \rangle$, so initially we approximate it by μ_{0w} . Thus, the algorithm starts with an initial guess for the absorption coefficient, $a^{(0)}(z) = \mu_{0w} K_V(z)$ which is averaged over depth according to

$$a^{(0)} = \frac{\int_0^{z_m} a^{(0)}(z) f(z) dz}{\int_0^{z_m} f(z) dz}, \quad (6.79)$$

where z_m is the maximum depth of the measurements and $f(z)$ is a weighting function that could for example be used to give less weight to the less accurate values of $K_V(z)$ near the surface (due to wave-induced light field fluctuations). In our Monte Carlo simulations, the results are found to be insensitive to the form of f , and we simply take $f = 1$. Now, recall Eq. (6.48) that the reflectance just beneath the water surface is given by

$$QSSA'_R(u_{0w}) = \frac{b_b}{a + b_b} I'_R(u_{0w}) \approx \frac{b_b}{a} I'_R(u_{0w})$$

and from Figure 6.44 $I'_R \sim 0.33$ for the Petzold Phase-T phase function. The lowest order relationship, $QSSA'_R \approx R$, provides an initial estimate of b_b :

$$b_b^{(0)}(z) = 3R(z)a^{(0)}(z). \quad (6.80)$$

This is then averaged over depth in a manner identical to $a^{(0)}(z)$. As in all inverse algorithms, an assumption must be made for the scattering phase function because it is required to solve the radiative transfer equation. This assumption (e.g., the Petzold Phase-T) provides B , which through $b^{(0)} = b_b^{(0)}/B$, yields the initial estimate of the scattering coefficient. For the retrievals here, we used the actual phase function employed in the Monte Carlo simulations, i.e., the correct phase function. At this point we have all of the IOPs required to solve the radiative transfer equation: $a^{(0)}$, $b^{(0)}$, and the scattering phase function. The solution of the transfer equation provides $E_u^{(0)}$, $E_d^{(0)}$, and $E_0^{(0)}(z)$, yielding an estimate for the $\langle \mu(z) \rangle$ profile: $\langle \mu(z) \rangle^{(0)}$. The estimated $\langle \mu(z) \rangle$ provides a new estimate of $a(z)$, i.e., $a^{(1)}(z) = \langle \mu(z) \rangle^{(0)} K_V(z)$. We get a revised estimate of $b_b(z)$ in the following manner. Calling the computed value of $R(z)$ after the initial iteration $R^{(0)}(z)$, $\Delta R(z) = R(z) - R^{(0)}(z)$ is formed, and the change in $b_b(z)$ is taken to be $\Delta b_b(z) = 3\Delta R(z)a^{(1)}(z)$. However, since the relationship $R(z) \approx 0.33b_b(z)/a(z)$ is not precise, the revised estimate for b_b is taken to be

$$b_b^{(1)}(z) = b_b^{(0)}(z) + \epsilon \Delta b_b(z)$$

where $\epsilon < 1$. This has the property that it pushes (or nudges) b_b in the right direction, but with $\epsilon \approx 0.5$, there is little possibility of overshoot. These are then averaged over depth as before, and $b_b^{(1)}$ is combined with B to provide $b^{(1)}$. The revised IOPs are then inserted into the radiative transfer equation and its solution provides the quantities required for the next step in the iteration: $E_u^{(1)}(z)$, $E_d^{(1)}(z)$, and $E_0^{(1)}(z)$. This is then repeated many times. The algorithm is stopped when the residual error after n iterations, defined as

$$\delta^{(n)} \triangleq \frac{1}{N} \sum_{i=1}^N |\ell n[E_d^{(n)}(z_i)] - \ell n[E_d(z_i)]| + \frac{1}{N} \sum_{i=1}^N |\ell n[E_u^{(n)}(z_i)] - \ell n[E_u(z_i)]|,$$

where the superscripted irradiances are those computed in the n^{th} iteration and z_i are the depths at which the irradiance data are given, reaches a minimum.

Applying the algorithm to simulated data shows that the retrieved a and b_b values are only weakly dependent on the assumed particle phase function (in the QSSA they depend on the phase function only through I'_R). Of course, when the correct phase function is used, the value of B is correct, so an estimate of b can be obtained as well; however, in general b cannot be obtained using such an algorithm.

In the case where $L_u(z)$ is measured rather than $E_u(z)$. The algorithm is still applicable: $E_u(z)$ is estimated from $L_u(z)$ using the “ Q -factor,” $Q = E_u/L_u$. At each iteration, the Q -factor (which is initially set to π) is updated based on the results of the previous iteration,

i.e., after the n^{th} iteration $E_u^{(n)}(z) = Q^{(n)}(z)L_u(z)$. This $E_u^{(n)}(z)$ is combined with measured $E_d(z)$ and used as in the earlier E_u and E_d algorithm. The performance of the $L_u(z)$ and $E_d(z)$ is only slightly degraded compared to the $E_u(z)$ and $E_d(z)$ algorithm.

6.6.2 Iterative Inversion Algorithm — Stratified Water Bodies

How do we deal with the vertical stratification of water bodies? From what we saw earlier in our discussion of stratified media, we can expect that K_V , in the manner of K_d will be a “local” property, i.e., $K_V(z)$ will depend mostly on the IOPs at the depth z (Figure 6.51). Thus, Eq. (6.79) should still provide a good approximation to $a(z)$. However, $R(z)$ is not local, i.e., it depends on the IOPs at all depths greater than z , so Eq. (6.80) can hardly be a reasonable approximation to estimate $b_b^{(0)}$, nor will Δb_b estimated this way be realistic. When L_u is the measured upwelling quantity, this can be remedied by using our discussion in Section 6.4.11 on stratified media. Recall Eq. (6.57):

$$R_L^{(1)}(\tau_a, \mu, \phi) = - \int_{\tau_a}^{\infty} R_{LH}^{(1)}(\tau, \mu, \phi) \frac{d}{d\tau} \exp[-(\tau - \tau_a)M(u_0, \mu)] d\tau.$$

If we compute $dR_L^{(1)}/d\tau_a$, we find

$$R_{LH}^{(1)}(\tau_a) = R_L^{(1)}(\tau_a) - \frac{1}{M} \frac{dR_L^{(1)}(\tau_a)}{d\tau_a},$$

where $R_{LH}^{(1)}(\tau_a)$ is the first order reflectance of a homogeneous medium with the same IOPs as the medium of interest has at depth τ_a . Introducing the QSSA for which $d\tau_a = c(z) dz \rightarrow [a(z) + b_b(z)] dz \rightarrow \langle u_{0w} \rangle_0 K_d(z) dz$, we have

$$R_{LH}(z) = R_L(z) - \frac{1}{M \langle u_{0w} \rangle_0 K_d(z)} \frac{dR_L(z)}{dz} = R_L(z) - \frac{1}{(1 + \langle u_{0w} \rangle_0) K_d(z)} \frac{dR_L(z)}{dz}. \quad (6.81)$$

$R_{LH}(z)$ is the value $R_L(z)$ would have if the medium were homogeneous for depths $> z$ with the IOPs that the actual medium has at z . We refer to R_{LH} as R_L for the “equivalent” homogeneous medium. To illustrate the efficacy of Eq. (6.81) we apply it to the three-layer example discussed in Figure 6.52. This is shown in Figure 6.62 (Right Panel). Note that the rather confusing behavior of $R_L(z)$ with depth has been simplified, with R_{LH} making it fairly obvious that the medium consists of three *homogeneous* layers. It is the function $R_{LH}(z)$ derived from the actual profile of $R_L(z)$ that should be used as the initial guess and the updating of the IOPs. That is, along with Eq. (6.79) as the initial guess for a ,

$$b_b^{(0)}(z) = 3\pi R_{LH}(z)a^{(0)}(z)$$

should be used as the initial guess for $b_b(z)$ with the initial $Q(z)$ taken to be π . Then, from the solution that yields $R_{HL}^{(0)}(z)$, $\langle\mu^{(0)}(z)\rangle$, and $Q^{(0)}(z)$, the trial IOPs for the next iteration are

$$a^{(1)}(z) = \langle\mu(z)^{(0)}\rangle K_V(z) \quad \text{and} \quad b_b^{(1)}(z) = b_b^{(0)}(z) + \epsilon \Delta b_b^{(0)}(z),$$

where now

$$\Delta b_b^{(0)}(z) = 3Q^{(0)}(z) \Delta R_{LH}^{(0)}(z) a^{(1)}(z) \quad \text{with} \quad \Delta R_{LH}^{(0)}(z) = R_{LH}(z) - R_{LH}^{(0)}(z).$$

Thus, what we are trying to do is to find IOPs such that when inserted into the radiative transfer equation reproduce AOPs of the stratified medium, by updating the trial IOPs using the “equivalent” homogeneous medium. Note that the equivalent homogeneous medium is used only for *initiating and updating* the IOPs.

Examples of the efficacy of this inversion scheme are provided in Figure 6.63, where we have applied this procedure to the three-layer profiles of R_L in Figure 6.62. The figures show that a is retrieved to high accuracy (error $< 1\%$). The absorption coefficient is mostly determined by K_d (actually K_V) and this quantity can be accurately computed.⁴² In contrast, given a , b_b is mostly determined by R_L , which being the ratio of two noisy quantities exhibits significant noise. In addition, the trial b_b also depends on dR_L/dz , which is proportional to the difference of two noisy quantities, and also exhibits even more noise than the individual quantities. Thus, the retrieved b_b is expected to be noisier than a , and the results bear this out. It is seen here that the error in b_b appears to be $\sim 10\%$ near the surface, and decreasing away from the surface. Some of this error is, of course, the result of the approximate nature of Eq. (6.81); however, we note that the overall error does not seem to depend on the stratification, so Eq. (6.81) does in fact remove most of its effects.⁴³

One should note that IOPs obtained by such inversions are actually preferable to those measured by conventional in-situ instruments, which usually examine small sample volumes (few cubic cm) and thus undersample the contribution of larger particles (which are few in number, but there nonetheless). IOPs determined from AOPs sample volume sizes that are appropriate to radiative transfer (volume size is $\sim 1/K_d^3$, which can be $\sim 10\text{--}1000 \text{ m}^3$

⁴²These profiles of R_L and E_d are the result of Monte Carlo simulations, and thus can have considerable noise. The K 's can be accurately found even in the presence of noise by using the observation that the irradiances decay more or less exponentially. Thus, rather than differentiating the quantities themselves, we differentiate the log-transformed quantities. This reduces the noise (and improves the accuracy) of the K estimates.

⁴³Note that in these inversions we have used the stratified version of the algorithm, i.e., the one employing Eq. (6.81), even for the homogeneous case ($\omega_0 = 0.7$ at all depths). Using the homogeneous version of the algorithm (i.e., dropping the dR_L/dz term in Eq. (6.81) for this particular case reduces the error by about a factor of two.

or greater). It is these IOPs that, when inserted into the radiative transfer equation, or our simple relationships, yield the appropriate AOPs.

6.7 Appendix 2: Bio-Optical Algorithms

Although not really a part of the AOP subject per se, it seems appropriate here to discuss the so-called “bio-optical algorithms:” relationships between the AOPs and the biologically important parameters like the pigment concentration and the Chlorophyll *a* concentration. In Section 6.2 we defined the remote sensing reflectance as $R_{rs}(\lambda) = L_w(\lambda)/E_d^+(\lambda)$, where $L_w(\lambda)$ is the water-leaving radiance propagating toward the zenith and $E_d^+(\lambda)$ is the downward irradiance from the Sun and sky falling on the water surface. Here, we examine important relationships between the remote sensing reflectance at various wavelengths and the pigment concentration C_P . Figure 6.5 shows that (in Case 1 waters) as C_P increases, the reflectance in the blue decreases and the reflectance in the green and yellow increases. This change in the reflectance spectrum (the “color” of the water) with C_P is what originally led to ocean color remote sensing.

Figure 6.5 suggests that C_P could be estimated from a ratio of reflectance in the blue to that in the green. The first ocean color sensor in space, the Coastal Zone Color Scanner (CZCS) had spectral bands at 443, 520, 550, and 670 nm, and the first empirical algorithms for relating reflectance to C_P used ratios formed with the first three of these bands. Figure 6.64 shows the original bio-optical algorithms used with the CZCS for Case 1 waters. In CZCS the idea was to use the $R_{rs}(443)/R_{rs}(550)$ algorithm (left panel) for lower values of C_P , and for higher values, where retrieving accurate values of $R_{rs}(443)$ became very difficult due to its small value, switch to the $R_{rs}(520)/R_{rs}(550)$ algorithm (right panel). The blue to green ratio is clearly more sensitive to changes in C_P than the blue-green to green ratio, i.e., $d(\log_{10}(C_P))/d(\log_{10}(R)) = -1.68$ for the blue to green and compared to -6.15 for the blue-green to green. The performance of the blue-green ratio algorithm suggests that C_P can be estimated from the ratio with an uncertainty of $\sim 30\%$.

SeaWiFS had more spectral bands in the visible than CZCS. In addition to CZCS, it had bands at 412, 490, and 510 nm.⁴⁴ This led to a four-band algorithm called “OC4” using 443, 490, 510, and 550. This algorithm was of the form

$$\log_{10}(C_P) = \sum_{n=0}^4 a_n \left(\log_{10}(R) \right)^n,$$

⁴⁴The green band on SeaWiFS was 555 nm compared to 550 nm in CZCS, but that is basically irrelevant to the C_P -retrieval.

where R is the largest among $R_{rs}(443)/R_{rs}(550)$, $R_{rs}(490)/R_{rs}(550)$ and $R_{rs}(510)/R_{rs}(550)$. An example of this is provided in Figure 6.65 for the same data set used in Figure 6.64. The curve on the figure is a quadratic (rather than quartic) least-squares fit to the data.

Algorithms based on differences in reflectances rather than ratios were considered early on in the CZCS program, but fell out of favor because ratio algorithms were thought to be more forgiving of errors in atmospheric correction. However, difference algorithms react to error in $R_{rs}(\lambda)$ differently from ratio algorithms. For the former, if

$$R = \frac{R_{rs}(\lambda_1)}{R_{rs}(\lambda_2)}, \quad \text{then} \quad R + \Delta R = \frac{R_{rs}(\lambda_1) + \Delta R_{rs}(\lambda_1)}{R_{rs}(\lambda_2) + \Delta R_{rs}(\lambda_2)} \approx R + \frac{\Delta R_{rs}(\lambda_1)}{R_{rs}(\lambda_1)} - \frac{\Delta R_{rs}(\lambda_2)}{R_{rs}(\lambda_2)},$$

and the error in R vanishes in first order if the *relative* error is the same at both wavelengths. In contrast, for the algorithm using differences, if $D = R_{rs}(\lambda_1) - R_{rs}(\lambda_2)$, then

$$D + \Delta D = R_{rs}(\lambda_1) + \Delta R_{rs}(\lambda_1) - R_{rs}(\lambda_2) + \Delta R_{rs}(\lambda_2) = D + (\Delta R_{rs}(\lambda_1) - \Delta R_{rs}(\lambda_2)),$$

and the error vanishes if the *absolute* error is the same at both wavelengths. An algorithm using differences called the “color index” (CI) has been developed and used in processing of SeaWiFS and MODIS imagery for oligotrophic waters. It is given by

$$CI = R_{rs}(550) - \left[R_{rs}(443) + \left(\frac{550 - 443}{670 - 443} \right) (R_{rs}(670) - R_{rs}(443)) \right].$$

On a graph of R_{rs} as a function of λ , the term in the square brackets is the value $R_{rs}(550)$ would have if the spectral variation was a straight line drawn from $R_{rs}(443)$ to $R_{rs}(670)$. The difference between that value and the actual value is CI (so CI is negative for low C_P and positive for high C_P). For this color index, if the error in remote sensing reflectance is linear in λ , i.e., $\Delta R_{rs}(\lambda) \sim e_1 + e_2\lambda$, then

$$\Delta CI = e_1 + 550e_2 - \left[e_1 + 443e_2 - \left(\frac{550 - 443}{670 - 443} \right) (e_1 + 670e_2 - e_1 - 443e_2) \right] = 0, \quad (6.82)$$

and the error in CI vanishes when the error in $R_{rs}(\lambda)$ is a *linear* function of λ . For the same data as in Figure 6.64, the relationship between the color index (Eq. (6.82)) and C_P is shown in the right panel of Figure 6.65.⁴⁵ The straight line on the figure is a linear regression using only data with $CI < 0$: $\log_{10}(C_P) = -0.58 + 70.48 CI$.

All of these algorithms (with the exception of CI operating over a limited range) display essentially the same uncertainty in retrieving C_P ($\sim 30\%$).⁴⁶ This is in part due to the natural variability of the IOPs with C_P (Chapter 5). Similar algorithms with similar accuracies have been developed from a much larger data base using C in place of C_P . These are used in all post-CZCS sensors.

⁴⁵The quantity actually used in Figure 6.65 was not R_{rs} , it was πR_{rs} , which is called the normalized water-leaving reflectance and which is used extensively in later chapters.

⁴⁶This assumes *no* error in the associated $R_{rs}(\lambda)$'s.

6.8 Appendix 3: An Alternative to the QSSA: The Zaneveld Solution

In this appendix we derive an alternative approximate solution for L_u/E_d at the surface of an absorbing-scattering medium based on the work of J.R.V. Zaneveld. Consider a 1-d medium, i.e., one in which the IOPs depend only on depth, illuminated by collimated irradiance F_0 . Within the medium, and in the absence of internal sources, the radiance $L(z, \hat{\xi}, \lambda)$ is governed by

$$\hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial z} L(z, \hat{\xi}, \lambda) = -c(z, \lambda) L(z, \hat{\xi}, \lambda) + \int_{\text{All } \hat{\xi}'} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}, \lambda) L(z, \hat{\xi}', \lambda) d\Omega(\hat{\xi}').$$

(We now drop the λ 's from the argument list, but remember that this is for monochromatic radiation.) In general, we note that $\beta(z; \hat{\xi}' \rightarrow \hat{\xi})$ is slowly varying with $\hat{\xi}'$ for $\hat{\xi}' \bullet \hat{\xi} < 0$, while $L(z, \hat{\xi}')$ is slowly varying for $\hat{\xi}' \bullet \hat{e}_z < 0$. Thus, it is convenient to break the integral term into two terms, one for integrations over the lower hemisphere ($\hat{\xi}' \bullet \hat{e}_z > 0$) and one for integration over the upper hemisphere ($\hat{\xi}' \bullet \hat{e}_z < 0$), indicated by Ω'_d and Ω'_u , respectively:

$$\begin{aligned} \int_{\Omega'_u + \Omega'_d} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) L(z, \hat{\xi}') d\Omega(\hat{\xi}') &= \int_{\Omega'_u} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) L(z, \hat{\xi}') d\Omega(\hat{\xi}') \\ &+ \int_{\Omega'_d} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) L(z, \hat{\xi}') d\Omega(\hat{\xi}'). \end{aligned}$$

As we will be interested in radiance traveling in the upward direction, we examine cases for which $\hat{\xi}$ is upward, i.e., take $\hat{\xi} \bullet \hat{e}_z < 0$. Consider the Ω'_d integral. Since $\beta(z; \hat{\xi}' \rightarrow \hat{\xi})$ is slowly varying over the range of integration, we define an angle $\hat{\zeta}$ (unknown) such that

$$\int_{\Omega'_d} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) L(z, \hat{\xi}') d\Omega(\hat{\xi}') = \beta(z; \hat{\zeta} \rightarrow \hat{\xi}) \int_{\Omega'_d} L(z, \hat{\xi}') d\Omega(\hat{\xi}') = \beta(z; \hat{\zeta} \rightarrow \hat{\xi}) E_{0d}(z), \quad (6.83)$$

where $E_{0d}(z)$ is the downwelling *scalar* irradiance at z . Likewise, we note that as $L(z, \hat{\xi}')$ is slowly varying to the upward integration, we choose $\hat{\eta}$ so that

$$\int_{\Omega'_u} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) L(z, \hat{\xi}') d\Omega(\hat{\xi}') = L(z, \hat{\eta}) \int_{\Omega'_u} \left[\frac{L(z, \hat{\xi}')}{L(z, \hat{\eta})} \right] \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}'). \quad (6.84)$$

With these changes,

$$\begin{aligned} \hat{\xi} \bullet \hat{e}_z \frac{\partial}{\partial z} L(z, \hat{\xi}, \lambda) &= -c(z, \lambda) L(z, \hat{\xi}, \lambda) \\ &+ \beta(z; \hat{\zeta} \rightarrow \hat{\xi}) E_{0d}(z) + L(z, \hat{\eta}) \int_{\Omega'_u} \left[\frac{L(z, \hat{\xi}')}{L(z, \hat{\eta})} \right] \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}'). \end{aligned} \quad (6.85)$$

Now, apply this to the surface ($z = 0$) of a medium with no refracting interface ($m = 1$) illuminated by a beam of irradiance propagating in the direction $\hat{\xi}_0$:

$$\begin{aligned} \hat{\xi} \bullet \hat{e}_z \left(\frac{\partial}{\partial z} L(z, \hat{\xi}) \right)_{z=0} &= -c(0, \lambda) L(0, \hat{\xi}) \\ &+ \beta(0, \hat{\xi}_0 \rightarrow \hat{\xi}) E_{0d}(z) + L(0, \hat{\xi}) \int_{\Omega'_u} \left[\frac{L(0, \hat{\xi}')}{L(0, \hat{\xi})} \right] \beta(0, \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}'), \end{aligned} \quad (6.86)$$

where we have replaced $\hat{\eta}$ by $\hat{\xi}$, as that is the direction we want the radiance. Defining the attenuation coefficient for radiance in the direction $\hat{\xi}$, $K_L(z, \hat{\xi})$, through

$$K_L(z, \hat{\xi}) = -\frac{1}{L(z, \hat{\xi})} \frac{\partial}{\partial z} L(z, \hat{\xi}),$$

we have

$$\begin{aligned} -\hat{\xi} \bullet \hat{e}_z K_L(0, \hat{\xi}) L(0, \hat{\xi}) &= -c(0, \lambda) L(0, \hat{\xi}) \\ &+ \beta(0, \hat{\xi}_0 \rightarrow \hat{\xi}) E_{0d}(0) + L(0, \hat{\xi}) \int_{\Omega'_u} \left[\frac{L(0, \hat{\xi}')}{L(0, \hat{\xi})} \right] \beta(0, \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}'), \end{aligned}$$

If we define the shape factor, $f_L(z, \hat{\xi})$, through

$$f_L(z, \hat{\xi}) b_f(z) \triangleq \int_{\Omega'_u} \left[\frac{L(z, \hat{\xi}')}{L(z, \hat{\xi})} \right] \beta(z, \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}'), \quad (6.87)$$

where $b_f(z)$ is the forward scattering coefficient (sometimes called $F(z)b(z)$ in the text), then

$$L(0, \hat{\xi}) = \frac{\beta(0, \hat{\xi}_0 \rightarrow \hat{\xi}) E_{0d}(0)}{-\hat{\xi} \bullet \hat{e}_z K_L(0, \hat{\xi}) + c(0) - f_L(0, \hat{\xi}) b_f(0)}, \quad (6.88)$$

or, considering just the radiance propagating toward the zenith (which we have indicated by L_u in the text),

$$L_u(0) = \frac{\beta(0, \hat{\xi}_0 \rightarrow -\hat{e}_z) E_{0d}(0)}{K_L(0, -\hat{e}_z) + c(0) - f_L(0, -\hat{e}_z) b_f(0)}, \quad (6.89)$$

Recall that the average cosine of the downwelling light field, $\langle u_d(z) \rangle$, is defined through $\langle u_d(z) \rangle = E_d(z)/E_{0d}(z)$, so

$$\frac{L_u(0)}{E_d(0)} = \frac{\beta(0, \hat{\xi}_0 \rightarrow -\hat{e}_z)}{\langle u_d(0) \rangle [K_L(0, -\hat{e}_z) + c(0) - f_L(0, -\hat{e}_z) b_f(0)]}, \quad (6.90)$$

In the case we are examining, the quantity $\langle u_d(0) \rangle$ is just $u_0 = \cos \theta_0$, so finally,

$$\frac{L_u(0)}{E_d(0)} = \frac{\beta(0, \hat{\xi}_0 \rightarrow -\hat{e}_z)}{u_0[K_L(0, -\hat{e}_z) + c(0) - f_L(0, -\hat{e}_z)b_f(0)]}, \quad (6.91)$$

Equation (6.91) is an *exact* solution of the radiative transfer problem, but unfortunately it involves two unknown AOPs,⁴⁷ $K_L(0, -\hat{e}_z)$ and $f_L(0, -\hat{e}_z)$. Following Zaneveld, it may be better to refer to this equation as a *restatement of the radiative transfer equation*. To proceed in making it useful and to connect it to what we have done earlier, we need to approximate the AOPs.

The shape factor f_L defined in Eq. (6.87) has not been encountered in the text, so we need to familiarize ourselves with its behavior as a function of the IOPs. As a simple example of the variation of f_L with the IOPs, we shall characterize β using the Henyey-Greenstein (H-G) phase function and calculate f_L as a function of the phase function asymmetry parameter g and the single scattering albedo ω_0 . Note that as g increases from 0 to unity the H-G phase functions progresses from isotropic to very strongly peaked in the forward direction. As a further simplification we assume the Sun is at the zenith, so $\hat{\xi}_0 = \hat{e}_z$. The results of this exercise are shown in Figure 6.67 (right panel), which provides $f_L(0, -\hat{e}_z)$, when $\hat{\xi}_0 = +\hat{e}_z$, as a function of ω_0 and the asymmetry parameter g . Included for comparison on the figure is the result for pure Rayleigh scattering (Ray.). In the particular case that we are examining, f_L is given by

$$f_L(z, -\hat{e}_z) = \frac{1}{b_f(z)} \int_{\Omega_u} \left[\frac{L(z, \hat{\xi}')}{L(z, -\hat{e}_z)} \right] \beta(z, \hat{\xi}' \rightarrow -\hat{e}_z) d\Omega(\hat{\xi}').$$

If $L(z, \hat{\xi})$ is uniform, then the integral is just $b_f(z)$ and $f_L = 1$. As ω_0 increases, there is more and more multiple scattering, which renders $L(z, \hat{\xi})$ more and more uniform and so $f_L \rightarrow 1$ as $\omega_0 \rightarrow 1$.⁴⁸ This is clearly seen in all of the cases in the figure. If β is strongly peaked in the forward direction, then the largest contribution to the integral is from $\hat{\xi} = -\hat{e}_z$ and the result approaches unity as β becomes more and more forward peaked. For phase functions Petzold and Petzmas, which have an integrable singularity at zero scattering angle, f_L is closer to unity for all ω_0 than seen in Figure 6.67 for the Henyey-Greenstein phase function. In the case of Rayleigh scattering, which has no forward peak, f_L is closer to unity than for HG, but for a different reason: the radiance distribution in the Rayleigh case is closer to uniform than the HG cases for all ω_0 . For small ω_0 a simple formula can be developed for f_L at the surface in terms of the phase function using the SSA or QSSA:

$$f_L(0, -\hat{e}_z) = \frac{4\pi}{P(\pi)} \int_{\Theta=90^\circ}^{\Theta=180^\circ} \frac{P(\Theta)P(\pi-\Theta)}{1-\cos\Theta} \sin\Theta d\Theta.$$

⁴⁷The quantity $K_L(0, -\hat{e}_z)$ was referred to as $K_L(0)$ in the text.

⁴⁸Actually, the upwelling radiance distribution in general will never be uniform, so $f_L = 1$ is not actually reached at $\omega_0 = 1$.

This can be used to evaluate f_L as $\omega_0 \rightarrow 0$. The Petzold (and Petzmas) phase function is more difficult to deal with. As seen in Figure 6.67 (left panel) it scatters much more in the forward direction than either of the HG phase functions in the figure. In fact, between $0.1^\circ \leq \Theta \leq 1^\circ$, $P(\Theta) \sim \Theta^{-1.38}$, so the manner in which the singularity is treated can affect the resulting f_L at the level of $\sim 1\%$ or so. Fitting $P(\Theta)$ in the range $0.1^\circ \rightarrow 1^\circ$ to a power law, which was assumed valid to $\Theta = 0$, resulted in $f_L = 0.994$, using the above formula for $\omega_0 \rightarrow 0$.

Equation (6.91) is exact, but has some similarities to the QSSA formulas. How are they related? As we have just seen, when the phase function is strongly peaked in the forward direction, as are those encountered in marine optics, the Ω'_u integral, Eq. (6.84), to a good approximation is

$$\int_{\Omega'_u} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) L(z, \hat{\xi}') d\Omega(\hat{\xi}') \approx L(z, \hat{\xi}) \int_{\Omega'_u} \beta(z; \hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}').$$

The integral on the right hand side of this is just $b(1 - P_{ru}(u, \phi))$, where

$$P_{ru}(u, \phi) \triangleq \frac{1}{4\pi} \int_0^{2\pi} d\phi' \int_0^1 P_r(u' \rightarrow u, \phi' \rightarrow \phi) du',$$

$u = -\hat{e}_z \bullet \hat{\xi}$ (and ϕ is irrelevant), and P_r is the scattering phase function *remaining* after the Dirac delta function has replaced the forward peak in the QSSA. The function P_{ru} is provided in Figure 6.15 for the Petzold phase function (as well as the Petzmas phase function). Thus, with this approximation

$$f_L(0, -\hat{e}_z) b_f(0) = b(0)(1 - P_{ru}(1)) = c(0)\omega_0(0)(1 - P_{ru}(1)).$$

In addition, we have seen that for the limited number of examples considered, when $\hat{\xi} = -\hat{e}_z$, $K_L(0, -\hat{e}_z) = K_L(0) \approx K_d(0)$, but an excellent QSSA approximation to $K_d(0)$ is seen in Figure 6.17, where $K_d(0) = u_0 c(0)[1 - \omega_0(0)(1 - P_{ru}(u_0))]$. With these replacements,

$$\frac{L_u(0)}{E_d(0)} = \frac{\beta(0, \hat{\xi}_0 \rightarrow -\hat{e}_z)}{c(0)[1 - \omega_0(0)(1 - P_{ru}(u_0))] + u_0 c(0)[1 - \omega_0(0)(1 - P_{ru}(1))]}.$$

Noting that $P_{ru}(1) = B$ and $P_{ru}(u_0) \approx B$ for $\theta_0 \lesssim 35^\circ - 40^\circ$ (Figure 6.15), we finally obtain

$$\frac{L_u(0)}{E_d(0)} = \frac{1}{4\pi} \frac{\omega_0(0)P(0, \hat{\xi}_0 \rightarrow -\hat{e}_z)}{[1 - \omega_0(0)F(0)](1 + u_0)}, \quad (6.92)$$

where $\beta/c = \omega_0 P/4\pi$ and $F = 1 - B$. Equation (6.92) is identical to the modified QSSA approximation to L_u/E_d , i.e., $QSSAL'(u_0)$ (Eq. (6.24)). Thus, one might argue that Eq.

(6.91) is an *exact* generalization of the QSSA in which the AOPs are explicitly retained rather than approximated.⁴⁹

Equation (6.88) also shows that the departure of the BRDF from that predicted by the single scattering approximation (or by the QSSA) with the IOPs, is governed by the variation of $K_L(0, \hat{\xi})$ and $f_L(0, \hat{\xi})$ with $\hat{\xi}$.

What about application to a real waterbody with a refracting interface and an atmosphere above. The form of Eq. (6.90) remains the same, but $\hat{\xi}_0$ is replaced by the unknown angle ζ from Eq. (6.83), i.e.,

$$\frac{L_u(0)}{E_d(0)} = \frac{\beta(0, \zeta \rightarrow -\hat{e}_z)}{\langle u_d(0) \rangle [K_L(0) + c(0) - f_L(0, -\hat{e}_z) b_f(0)]}.$$

In the absence of the atmosphere, the interface provides refraction of the solar beam, changing $\hat{\xi}_0$ to $\hat{\xi}_{0w}$, and reflection of upwelling radiance from below. Thus, the downwelling radiance incident on the waterbody just beneath the interface consists of a beam propagating in the direction $\hat{\xi}_{0w}$ ⁵⁰ and diffuse radiance reflected from the interface. Since the direct solar part is much stronger than the diffuse component, the angle $\zeta \approx \hat{\xi}_{0w}$ and $\langle u_d(0) \rangle \approx u_{0w}$, or in case of a rough surface $\langle u_d(0) \rangle \approx D_0^{-1}$ (see Section 6.4.4.2). The addition the atmosphere strengthens the diffuse part relative to the refracted direct solar beam, but the approximations are still valid, albeit a bit weaker. Thus, in a real waterbody, an excellent approximation is

$$\frac{L_u(0)}{E_d(0)} \approx \frac{D_0 \beta(0, \hat{\xi}_{0w} \rightarrow -\hat{e}_z)}{K_L(0) + c(0) - f_L(0, -\hat{e}_z) b_f(0)}, \quad (6.93)$$

where of course the AOPs are evaluated in the presence of the air-water interface and the atmosphere.

One fascinating feature of Eq. (6.91) and (6.93) is the fact that they are applicable to a medium with vertically stratified IOPs as well as a homogeneous medium. Since all of the IOPs in the equation are evaluated at $z = 0$, any effect of the vertical stratification on L_u/E_d must be *manifest* in the values of the AOPs. These equations can be considered the generalizations of the QSSA to stratified media.

Earlier in Section 6.4.11 we derived an approximate theory of L_u/E_d in a stratified medium and provided examples of profiles of K_L and L_u/E_d for a three-layer medium in

⁴⁹Another form of Eq. (6.91) that facilitates its comparison to the QSSA is

$$\frac{L_u(0)}{E_d(0)} = \frac{\omega_0 P(\hat{\xi}_0 \rightarrow -\hat{e}_z)}{u_0 [K_L/c + 1 - f_L(0, -\hat{e}_z) \omega_0 F]},$$

where all quantities are evaluated just beneath the surface.

⁵⁰Or if the surface is rough, propagating in a small range of directions close to $\hat{\xi}_{0w}$.

Figures 6.51 and 6.52, respectively. One should particularly take note of the very strong effect of the stratification on $K_L(\tau)$. For the three profiles examined in the figures, the values for $f_L(0, -\hat{e}_z)$ were all close to unity. Thus, in this case, virtually all of the influence of stratification is manifest in $K_L(0)$,⁵¹ and it *quantitatively* explains the variation of L_u/E_d at the surface with the varying stratification. The exact Eq. (6.88) is applicable to all depths in the medium, and the reader can see that it *qualitatively* predicts the behavior of $L_u(\tau)/E_d(\tau)$ in Figures 6.51 and 6.52. One exception is within the layer where $\omega_0 = 0$. There, $K_L = a$, $c = a$, and $\beta = 0$, so $b_f = 0$ yielding $0/0$, and the formula yields an indeterminate result. However, for a real waterbody, $\omega_0 = 0$ is not possible.

6.9 Appendix 4: An Alternative to the QSSA: The Modified δ -Isotropic Solution

In this appendix we develop an alternative approach to the QSSA that is similar in spirit but different in some important details. Recall that in the QSSA, the actual phase function in the problem was approximated and replaced by a Dirac delta function and a remainder P_r :

$$P(\hat{\xi}' \rightarrow \hat{\xi}) = A_\delta \delta^{(2)}(\hat{\xi}' - \hat{\xi}) + A_r P_r(\hat{\xi}' \rightarrow \hat{\xi}).$$

In the QSSA, we chose P_r to be the actual phase function for scattering angles greater than some value that was determined by the geometry of the particular problem. Here we replace P_r by a constant value over the entire range of scattering angles, so the scattering phase function is replaced by a Dirac delta function *plus* an isotropically scattering component with backscattering probability B . Evaluating A_δ and A_r yields

$$A_\delta = 4\pi(1 - 2B) \quad \text{and} \quad A_r = 2B.$$

We call this procedure the “ δ -Isotropic” approximation.

Inserting this phase function into the radiative transfer equation for a homogeneous medium provides the following equation for the radiance:

$$u \frac{dL(z, u)}{dz} = [-c + b(1 - 2B)]L(z, u) + bB \int_{-1}^1 L(z, u') du'.$$

Rewriting in terms of optical depth τ and single scattering albedo ω_0 gives

$$u \frac{dL(\tau', u)}{d\tau'} = -L(z, u) + \frac{\omega'_0}{4\pi} \int_{-1}^1 L(z, u') du', \quad (6.94)$$

⁵¹There is a very small $\langle u_d(0) \rangle$ change caused by the stratification. It varied by at total of about 1% between the two extreme profiles.

where

$$\tau = cz, \quad \omega_0 = \frac{b}{c}, \quad \tau' = \tau(1 - \omega_0(1 - 2B)), \quad \text{and} \quad \omega'_0 = \frac{2\omega_0 B}{1 - \omega_0(1 - 2B)}.$$

We apply this equation to the problem of radiance exiting the surface ($z = 0$) of such a medium when it is illuminated by the solar beam of irradiance F_0 propagating in a direction specified by $\hat{\xi}_0 = (u_0, \phi_0)$. S. Chandrasekhar has provided the *exact* solution to this problem in the following form:

$$\frac{L(0, \mu)}{E_d(0)} = \frac{\omega'_0}{4\pi} \frac{H(u_0, \omega'_0)H(\mu, \omega'_0)}{u_0 + \mu}, \quad (6.95)$$

where the H -functions are solutions to the nonlinear integral equation

$$H(u, \omega_0) = 1 + \frac{\omega_0}{2} u H(u, \omega_0) \int_0^1 \frac{H(u', \omega_0)}{u' + u} du', \quad (6.96)$$

and the propagation direction is $\hat{\xi} = (-\mu, \phi)$. The beauty of this solution is that it is exact, i.e., it includes all orders of multiple scattering. The H -functions can be derived numerically by iteration; however, there are several approximate formulas available that provide them with high accuracy. The one we use in the computations described below is

$$\frac{1}{H(u, \omega_0)} = \left[1 - (1 - \sqrt{1 - \omega_0})u \times \left\{ r_0 + \left(1 - \frac{r_0}{2} - r_0 u \right) \ln \left(\frac{1 + u}{u} \right) \right\} \right], \quad (6.97)$$

where

$$r_0 = \frac{2}{1 + \sqrt{1 - \omega_0}} - 1.$$

The error over the full range of both u and ω_0 is less than 1%.⁵²

Equation (6.95) cannot reproduce the variations in the BRDF that are due to structure in the scattering phase function, as isotropic scattering has no such structure. We can remedy this by noting that, as scattering is similar to a diffusion process, the phase function-induced structure in the BRDF comes from the first few scatterings. With this in mind, we subtract the single scattering contribution to the BRDF we have computed, and replace it with the single scattering solution derived with the correct phase function. This procedure results in a revised BRDF;

$$\frac{L(0, \mu, \phi)}{E_d(0)} = \frac{\omega'_0}{4\pi} \frac{H(u_0, \omega'_0)H(\mu, \omega'_0)}{u_0 + \mu} + \frac{\omega_0}{4\pi} \left(\frac{P(u_0, \phi_0 \rightarrow -\mu, \phi) - 2B}{u_0 + \mu} \right), \quad (6.98)$$

⁵²A simpler approximation, but with errors as high as 4% is

$$H(u, \omega_0) = \frac{1 + 2u}{1 + 2u\sqrt{1 - \omega_0}}.$$

where P is the actual phase function for the medium. In actuality, the δ -Isotropic approximation is being used to *assess the effect of multiple scattering*. We refer to this procedure as the “modified δ -Isotropic” approximation (MDIA). Note that in the limit that $\omega_0 \rightarrow 0$, for which $H(u, \omega_0) \rightarrow 1$,

$$\frac{L(0, \mu, \phi)}{E_d(0)} \rightarrow \frac{\omega_0}{4\pi} \left(\frac{P(u_0, \phi_0 \rightarrow -\mu, \phi)}{u_0 + \mu} \right),$$

the correct single-scattering result.

To see how well the MDIA approximates the L_u/E_d at nadir, we compare it with the QSSA in Figures 6.66 and 6.68, for water parameters that are the same as those in Section 6.4.3, Figure 6.26 (the right panels in Figures 6.66 and 6.68 are identical to Figure 6.26). We note that that Eq. (6.98) is slightly better than the QSSA for $B_p = 0.0181$, but somewhat poorer for $B_p = 0.0020$, where the QSSA becomes quite accurate because the phase function more closely resembles a Dirac delta function.

How well does Eq. (6.98) perform in computing the shape of the BRDF? In Figure 6.69 we compare the exact and computed (normalized) radiance distributions as functions of the viewing angle, $\arccos(\mu)$. The computations in the figure are for both the Petzold and the Petzmas phase functions (both have $B_p = 0.0181$). The Sun is at the zenith. Note that in contrast to the QSSA, which always yields a normalized BRDF that is too low at all viewing angles, the modified δ -Isotropic approximation performs better in reproducing the BRDF. For the Petzmas phase function it can reduce the QSSA-error in the predicted normalized BRDF in this case from about 10% to less than about 2%. The performance for the Petzold phase function, which has more structure in the backward scattering angles, is poorer, but the QSSA-error reduction is still significant, e.g., from as much as -20% for $\omega_0 = 0.9$ and $\theta_v = 30^\circ$ to about $+4\%$.

In light of the success of the modified δ -Isotropic approximation in the computation of the remote-sensing reflectance and the BRDF, one is led to expect that it should also be reasonably accurate for computing $R(0)$. The solution to Eq. (6.95) for the diffuse reflectance, $E_u(0)/E_d(0)$, can be shown to be

$$R(0) = 1 - H(u, \omega'_0) \sqrt{(1 - \omega'_0)},$$

so, from the integration of Eq. (6.98),

$$R(0) = \int_0^{2\pi} d\phi \int_{-1}^1 d\mu \frac{\mu L(0, \mu, \phi)}{E_d(0)} = 1 - H(u, \omega'_0) \sqrt{(1 - \omega'_0)} + R_{\text{Phase}} - R_{\text{Iso.}}, \quad (6.99)$$

where

$$R_{\text{Phase}} = \omega_0 I'_R(u_0) \quad \text{and} \quad R_{\text{Iso.}} = \frac{\omega_0}{2} \left[1 - u_0 \ln \left(\frac{1 + u_0}{u_0} \right) \right],$$

are the single scattering reflectances for the true phase function and for an isotropic phase function, respectively. The I'_R is provided in Figure 6.18. To test the quality of the modified δ -Isotropic approximation in the prediction of $R(0)$, the same cases are employed as those that were examined in Figure 6.25, i.e., combinations of particles and water for two particle phase functions: one with $B_p = 0.0181$ (Petzold) and one with $B_p = 0.0020$ (modified Petzold). The simulations used three solar zenith angles ($\theta_0 = 10^\circ, 37^\circ$ and 60°), nine single scattering albedos ($\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85$, and 0.9) and values of the total backscattering probability B ranging from B_p to 0.12 . Clearly the smaller B_p yields significantly better performance than the larger. For $B_p = 0.0181$ the performance degrades as θ_0 increases. The largest errors are usually associated with the largest values of ω_0 and θ_0 . At small θ_0 the agreement is excellent. For $\theta_0 = 10^\circ$, the average-absolute error is 2.4% for $B_p = 0.0020$ and 4.2% for $B_p = 0.0181$. In contrast, for $\theta_0 = 60^\circ$ the corresponding errors are 9.8% and 16.3%. Overall the error is 5.7% and 9.8%, respectively for $B_p = 0.0020$ and $B_p = 0.0181$, compared to 1.6% and 6.3% for $QSSA'_R$. Interestingly, unlike the situation with the BRDF, the phase function correction to the modified δ -Isotropic — the R_{Phase} and $R_{\text{Iso.}}$ terms in Eq. (6.99) — only make as small improvement in the performance for $R(0)$: overall the error for $B_p = 0.0181$ without the correction is 13.2% compared to 9.8% with the correction

An important feature of the δ -Isotropic approximation is that it can also be used to estimate the AOPs beneath the surface from the IOPs. Assume that we have solved Eq. (6.94) for $L(\tau', u)$. From this one can find all of the AOPs. For example,

$$\kappa'_d(\tau') = -\frac{d \ln E_d(\tau')}{d\tau'},$$

where $\tau' = c'z$, and $c' = c(1 - \omega_0(1 - 2B)) \triangleq c\gamma$. Therefore,

$$\kappa'_d(\tau') = -\frac{1}{\gamma c} \frac{d \ln E_d(\tau')}{dz} = \frac{K_d(\tau')}{\gamma c}, \quad \text{so} \quad K_d(z) = \gamma c \kappa'_d(\gamma c z).$$

Thus, if we want $K_d(z_0)$ in a medium characterized by ω_0 , B , and c , we find ω'_0 , solve Eq. (6.94) for $L(\tau', u)$, and compute $\kappa'_d(\tau')$. The required $K_d(z_0)$ is then $\gamma c \kappa'_d$, with κ'_d evaluated at $\tau' = \gamma c z_0$. Similar equations are easily verified for $K_u(z)$, $K_L(z)$, etc. The irradiances can be directly found in the same way, e.g., $E_d(z_0) = E_d(\tau')$, with $E_d(\tau')$ evaluated at $\tau' = \gamma c z_0$.

To see how well the δ -Isotropic approximation predicts the AOPs, we examine two of the cases in Figures 6.66, 6.68 and 6.70. Both cases have $B = 0.10$ and $\omega_0 = 0.8$. The first one has $B_p = 0.0020$, for which the modified δ -Isotropic approximation yields quite accurate values for $L_u(0)$ and $R(0)$, while the second has $B_p = 0.0181$ for which the δ -Isotropic approximation is in error by as much as 10%. Several of the AOPs are provided for these two cases at the surface and at a water depth of 10 m in Table 6.12. Again, there is no

atmosphere and the refractive index of water is taken to be unity. The results are clearly excellent for $B_p = 0.0020$. As expected, for $B_p = 0.0181$ the results are poorer, with errors in the various K 's of as much as 10%. Clearly, the δ -Isotropic approximation can provide reasonable, and sometimes excellent, estimates of AOPs within the water column.

Table 6.12: Comparison between AOPs computed from the exact (Monte Carlo) solution and the δ -Isotropic approximation to the radiative transfer problem with $\theta_0 = 60^\circ$, $\omega_0 = 0.80$, and $B = 0.10$. There is no atmosphere and the refractive index of water is taken to be unity.

	$B_p = 0.0020$ ($c = 0.0480 \text{ m}^{-1}$)		$B_p = 0.0181$ ($c = 0.0555 \text{ m}^{-1}$)	
Quantity	Exact (MC)	δ -Isotropic	Exact (MC)	δ -Isotropic
$R(0)$	0.1365	0.1355	0.1519	0.1355
$L_u(10 \text{ m})$	0.0124	0.0124	0.0119	0.0124
$E_d(10 \text{ m})$	0.3835	0.3843	0.3608	0.3684
$E_u(10 \text{ m})$	0.0524	0.0521	0.0553	0.0499
$K_L(10 \text{ m})$	0.0273	0.0267	0.0327	0.0308
$K_d(10 \text{ m})$	0.0269	0.0266	0.0335	0.0308
$K_u(10 \text{ m})$	0.0268	0.0268	0.0343	0.0312
$K_0(10 \text{ m})$	0.0262	0.0260	0.0325	0.0304
$\langle \mu_d(10 \text{ m}) \rangle$	0.480	0.483	0.465	0.482
$\langle \mu_u(10 \text{ m}) \rangle$	0.413	0.423	0.388	0.423

What effect does the atmosphere and the reflecting water surface have on the effectiveness of the δ -Isotropic approximation? We provide at least a glimpse of the effectiveness of the approximation in a more realistic setting, by looking at the same two cases as above, but with the atmosphere (characteristic of the actual atmosphere at 400 nm with molecular and aerosol scattering) in place and a reflecting interface between the water and the atmosphere. We compare the full, exact computation with for $\theta_0 = 60^\circ$, with the δ -Isotropic approximation (in the absence of the atmosphere), but for $\theta_0 = 40.52^\circ$, the refracted solar zenith angle in the water for $\theta_0 = 60^\circ$. This is in the same spirit as replacing u_0 in the QSSA without the refracting interface by u_{0w} with a refracting interface. The results are provided in Table 6.13, where because the atmosphere and surface will attenuate the absolute irradiances in the water, we compare only the ratio quantities. Again the agreement between the exact and approximate computations is quite good. The somewhat large difference in the values of K_0 (even for $B_p = 0.0020$) is attributed to internally reflected radiance in the case with the refracting interface. This causes E_0 to decay faster with depth than it does when the reflecting interface is missing.

The modified δ -Isotropic approach discussed here is an alternative to the QSSA for computing the remote-sensing reflectance and the diffuse reflectance. For the tests we examined it is somewhat less accurate than the QSSA in most respects, but more accurate

Table 6.13: Comparison between AOPs computed from the exact (Monte Carlo) solution and the δ -Isotropic approximation to the radiative transfer problem with $\omega_0 = 0.80$, and $B = 0.10$. For the exact computation $\theta_0 = 60^\circ$, the atmosphere is characteristic of that at 400 nm, and the refractive index of water is 1.333. For the δ -Isotropic there is no atmosphere, the refractive index of water is taken to be unity, and $\theta_0 = 40.52^\circ$, the refracted solar zenith angle for $\theta_0 = 60^\circ$. The quantity $L_u^*(10 \text{ m})$ is $L_u(10 \text{ m})/E_d(10 \text{ m})$.

	$B_p = 0.0020$ ($c = 0.0480 \text{ m}^{-1}$)		$B_p = 0.0181$ ($c = 0.0555 \text{ m}^{-1}$)	
Quantity	Exact (MC)	δ -Isotropic	Exact (MC)	δ -Isotropic
$R(0)$	0.1147	0.1119	0.1211	0.1119
$L_u^*(10 \text{ m})$	0.0320	0.0296	0.0321	0.0296
$K_L(10 \text{ m})$	0.0173	0.0174	0.0216	0.0197
$K_d(10 \text{ m})$	0.0179	0.0178	0.0218	0.0207
$K_u(10 \text{ m})$	0.0174	0.0171	0.0208	0.0198
$K_0(10 \text{ m})$	0.0167	0.0149	0.0195	0.0179
$\langle \mu_d(10 \text{ m}) \rangle$	0.709	0.714	0.690	0.714
$\langle \mu_u(10 \text{ m}) \rangle$	0.475	0.442	0.439	0.442

in one: the computation of the shape of the normalized BRDF.⁵³ In addition, it can be applied to the approximation of many AOPs *within* the water column that are not available in the QSSA. Unlike QSSA, at present it can only be applied to a homogeneous medium, but it should be investigated further. However, for the pedagogical purposes the author believes the QSSA is more appropriate for most applications because of its transparency.

6.10 Bibliographic Notes

Unless otherwise indicated, the radiative transfer simulations presented in this chapter are new and were carried out by the author specifically to illustrate ideas and concepts in the text.

6.2 AOP Definitions

The material in this section is basic to ocean optics. It was largely developed by R.

⁵³This one improvement is not to be underestimated. In Chapter 13 we discuss the validation of ocean color sensors, for which comparison of surface and sensor water-leaving radiances are required. For such comparisons the surface and sensor measurements of the water-leaving radiance are rarely made viewing in the same direction. Since the water-leaving radiance depends on the viewing direction it is important to have a method of transforming the measurements from one viewing direction to another – just what the modified δ -Isotropic is better at.

Preisendorfer [Preisendorfer, 1961, 1976a] and is described in greater detail in Mobley [1994].

6.3.1 Examples of AOPs Derived from Experimental Measurements

The data from Hawaii (Figures 6.1 to 6.4) are from Gordon et al. [2009]. The data from Figures 6.5 to 6.7 are from Morel and Prieur [1975] (see also Morel and Prieur [1977]). In the latter figures, those labeled “D” are from the Caribbean and the Sargasso Sea, while those labeled “C” are off the coast of Mauritania.

6.3.2 Examples of AOPs Derived from Computed Light Fields

The examples of the AOPs displayed in Figures 6.8 through 6.13 are the results of Monte Carlo simulations carried out by the author. In these examples the radiative transfer code did *not* account for the effects of polarization or inelastic scattering on the light field. The code is described in Mobley et al. [1993] along with comparisons with other codes.

6.4 Relationships between AOPs and IOPs

Most of this section is based on work by the author (published and unpublished). Although numerical codes are available for full radiative transfer computations of all AOPs given the IOPs — most notably “HydroLight” [Mobley, 1989, 1994] — the author believes there is value in simple, easy to understand (and derive) analytical relationships between the various relevant quantities. Much of the material in this section is an attempt to explore the full usefulness of the QSSA for providing such relationships between the AOPs to the IOPs.

The technique of replacing a portion of the forward peak of the scattering phase function with a Dirac delta function seems to have originated with Potter [1969]. The initial QSSA applied to ocean optics was described by Gordon [1973] to approximate $R(0)$, and has been refined over the years, e.g., Gordon et al. [1975]. The application to L_u was first made by Gordon et al. [1988], even though the basic ingredient was already in the 1973 paper. In 1973, unlike irradiance, radiance measurements were few at the time and did not become routine until the mid 1980’s.

In this section, for the least-squares fits of the “exact” L_u to $QSSA'_L(u_{0w})$, the radiance L_u was calculated with a radiative transfer code that *included* the full effects of polarization.

6.4.3 Decomposition of Scattering into Water and Particle Components

We note that

$$\begin{aligned} QSSA'_L(\mu, \phi; u_0, \phi_0) &= \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \frac{P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} \\ &= \frac{b_b}{a + b_b} \frac{P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)/B}{u_0 + \mu}, \end{aligned}$$

and

$$QSSA'_R(u_0) = \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \int_0^{2\pi} d\phi \int_0^1 \frac{\mu P_r(u_0 \rightarrow -\mu, \phi_0 \rightarrow \phi)}{u_0 + \mu} d\mu = \frac{b_b}{a + b_b} \frac{I_R}{B}.$$

In earlier works, e.g., [Gordon et al. \[1988\]](#) and [Gordon et al. \[1975\]](#), L_u/E_d and R were expanded in *powers* of the parameter $G = b_b/(a + b_b)$; however, here linear relationships are found. This is due to the fact that in the earlier studies the phase function consisted of the Petzold Phase-T alone, so those results corresponded to a *single* value of B . In contrast, in this Section, B varied over the range from 0.003 to 0.12 and the linear relations are better in this more realistic case (but see 6.4.9.1).

6.4.9.1 L_u/E_d

The hybrid model, in which different formalisms are used for ω_0 smaller or larger than 0.7, described in this section has not been previously published.

6.4.9.4 An Important Observation

The observation that b and c cannot be estimated from R is developed in detail in [Gordon \[1993\]](#).

6.4.10 The Penetration Depth

The relationship between the penetration depth and the K_d was first worked out by [Gordon and McCluney \[1975\]](#).

6.4.11 QSSA Estimate of the Influence of Vertical Structure in the IOPs

The influence of vertical stratification of the IOPs on the AOPs — particularly on $R(0)$ — has been studied by [Gordon \[1978a\]](#), [Gordon and Clark \[1980b\]](#), [Zaneveld \[1982\]](#), [Sathyendranath and Platt \[1989\]](#) and [Gordon \[1992\]](#). The derivation here is the author's; however, it was inspired by [Zaneveld et al. \[2005\]](#). It is superior to the earlier treatments.

A demonstration that K_d is a “local” property of a stratified medium is made in [Gordon \[1980\]](#).

6.4.12 The Influence of Inelastic Processes on the AOPs

The influence of fluorescence on the AOPs was first worked out in [Gordon \[1979\]](#) and that of Raman scattering by [Sugihara et al. \[1984\]](#), [Stavn and Weidemann \[1988\]](#) and [Marshall and Smith \[1990\]](#), and more recently by [Gordon \[1999\]](#) based on newer measurements of the Raman scattering cross section by [Bartlett et al. \[1998\]](#). The independence of the volume fluorescence function β_F on scattering angle was directly verified by [Gordon et al. \[1993\]](#) for several species of phytoplankton. High spectral resolution measurements of the effect of inelastic processes on the depth of solar Fraunhofer absorption lines in natural waters (not discussed here) can be found in [Ge et al. \[1995\]](#) and provide a direct separation between the contribution of elastic and inelastic processes to the AOPs. The model presented in 6.4.12.3 is the author's own development [[gordon 2014.](#)], although similar ideas can be found in [Bartlett \[1996\]](#) and [Sathyendranath and Platt \[1998\]](#).

6.4.13 Examples of Modeling with the QSSA

The possibility of a large instrument self-shading error in early spectral radiometric measurements, especially in the red (as mentioned in this section) was pointed out and quantified by [Gordon and Ding \[1992\]](#).

6.6 Appendix 1: Estimation of IOPs from AOPs

Various methods for AOP \rightarrow IOP inversion are reviewed in considerable detail in [Gordon \[2002\]](#). The method described in this appendix for homogeneous waters is from [Gordon and Boynton \[1997\]](#). It was extended to water bodies with depth-varying IOPs in [Gordon and Boynton \[1998\]](#) and waters where Raman scattering is important in [Boynton and Gordon \[2000\]](#), as well as waters for which the phase function has a significant Rayleigh component (very clear water) [[Boynton and Gordon, 2002](#)]. For preparation of Figure 6.63, in the [Gordon and Boynton \[1998\]](#) algorithm we replaced the original [Gordon and Clark \[1980b\]](#) treatment of vertical structure by Eq. (6.81), which is superior. Examples of the application of these methods to experimental data can be found in [Gordon et al. \[2009\]](#) and [Gordon et al. \[2009\]](#). The fact that the scattering coefficient cannot be retrieved from the AOPs alone because near-forward scattering contributes strongly to b but only weakly to the AOPs is shown directly through simulations by [Gordon \[1993\]](#).

6.7 Appendix 2: Bio-Optical Algorithms

Ratio algorithms of the type shown in Figure 6.64 were first given by [Gordon and Clark \[1980a\]](#) and in the form here by [Gordon et al. \[1983\]](#). The OC4-type algorithms are described in [O'Reilly et al. \[1998\]](#) and [O'Reilly et al. \[2000\]](#). Difference algorithms were first introduced by [Viollier et al. \[1980\]](#) and the novel *CI* algorithm was developed by [Hu](#)

et al. [2012].

6.8 Appendix 3: An Alternative to the QSSA: The Zaneveld Solution

The solution of the radiative transfer equation culminating in Eq. (6.88) closely follows that provided by Zaneveld [1982] and Zaneveld [1995].

6.9 Appendix 4: An Alternative to the QSSA: The Modified δ -Isotropic Solution

The solution to Eq. (6.95) is provided in Chandrasekhar [1950]. The approximation to the H -function in Eq. (6.98) is from Hapke [1993] and the simpler approximation in the footnote from Hapke [1981]. This particular phase function was first (to our knowledge) used by DiToro [1978] as a simplification to the QSSA. DiToro derived Eq. (6.95), but then solved it in the QSSA approximation rather than adapting Chandrasekhar's exact solution for R as we did here. Following Chandrasekhar [1950], he also solved for the asymptotic radiance distribution and K_∞ (Chapter 2), as did Gordon and Xu [1996], in applying the approximation to inelastic processes.

6.11 Figures

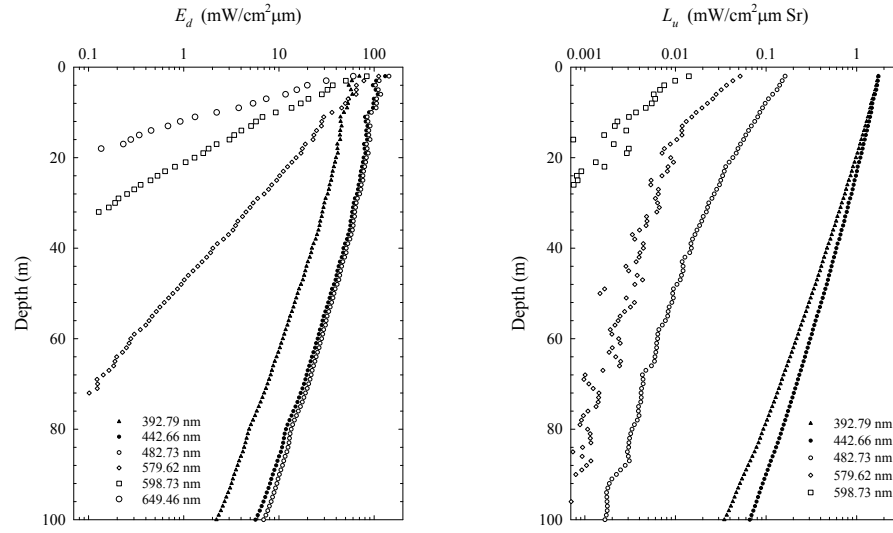


Figure 6.1: Examples of the spectral E_d (left panel) and L_u (right panel) obtained in the clear waters off Lanai, Hawaii. The sky was clear and the solar zenith angle was 36° .

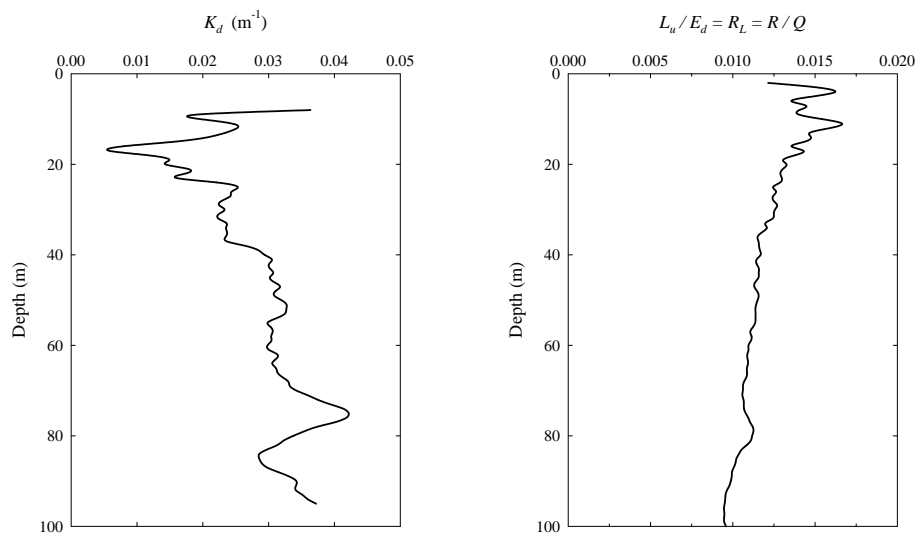


Figure 6.2: Profiles of K_d (left panel) and R_L (right panel) at 482.73 nm derived from the data in Figure 6.1.

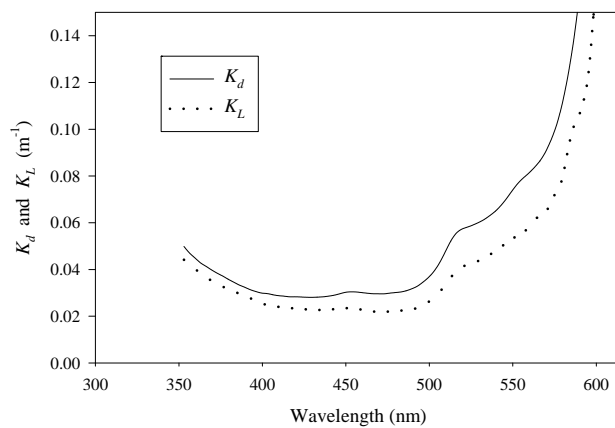
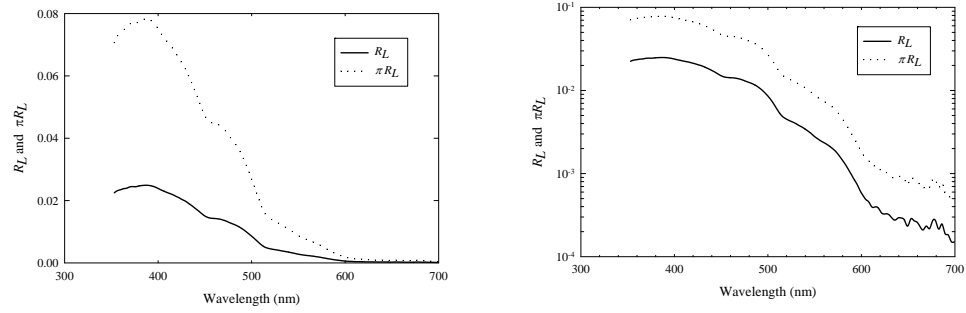
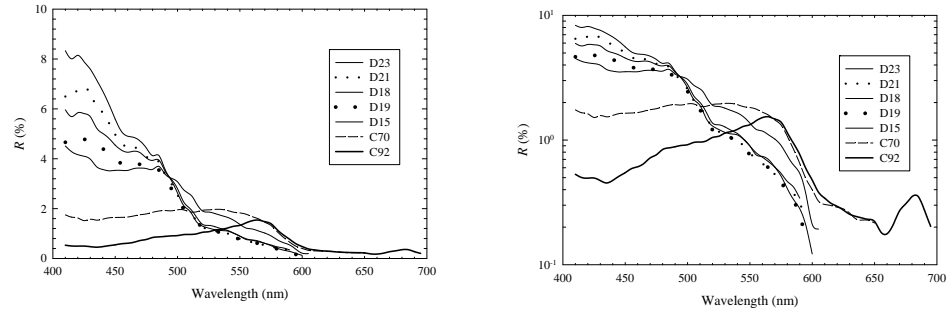
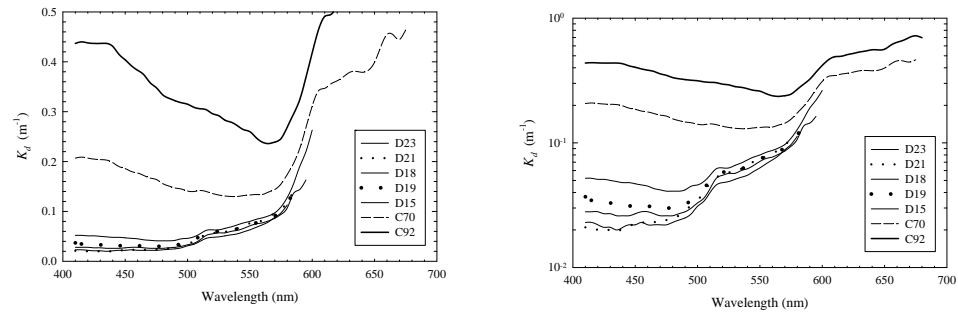


Figure 6.3: Spectra of K_d and K_L for the Data from Hawaii in Figure 6.1.

Figure 6.4: Spectra of R_L and πR_L for the Data from Hawaii in Figure 6.1.Figure 6.5: Spectra of R for Case 1 waters with a range of pigment concentration.Figure 6.6: Spectra of K_d for Case 1 waters with a range of pigment concentration.

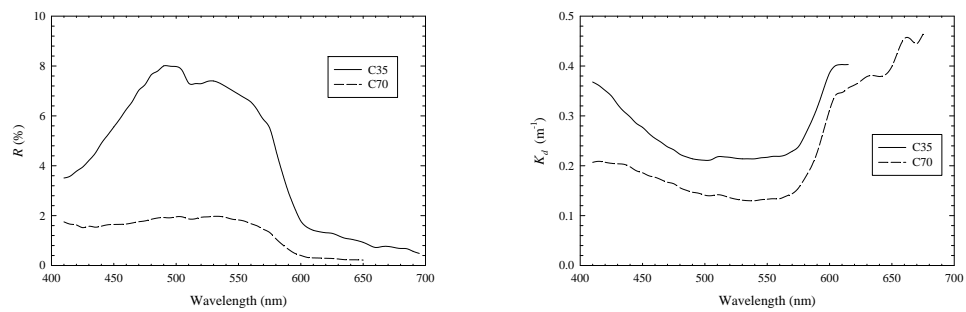


Figure 6.7: Examples of R and K_d providing a comparison of spectra between Case 1 (C70) and sediment-dominated Case 2 (C35) waters with comparable pigment concentrations.

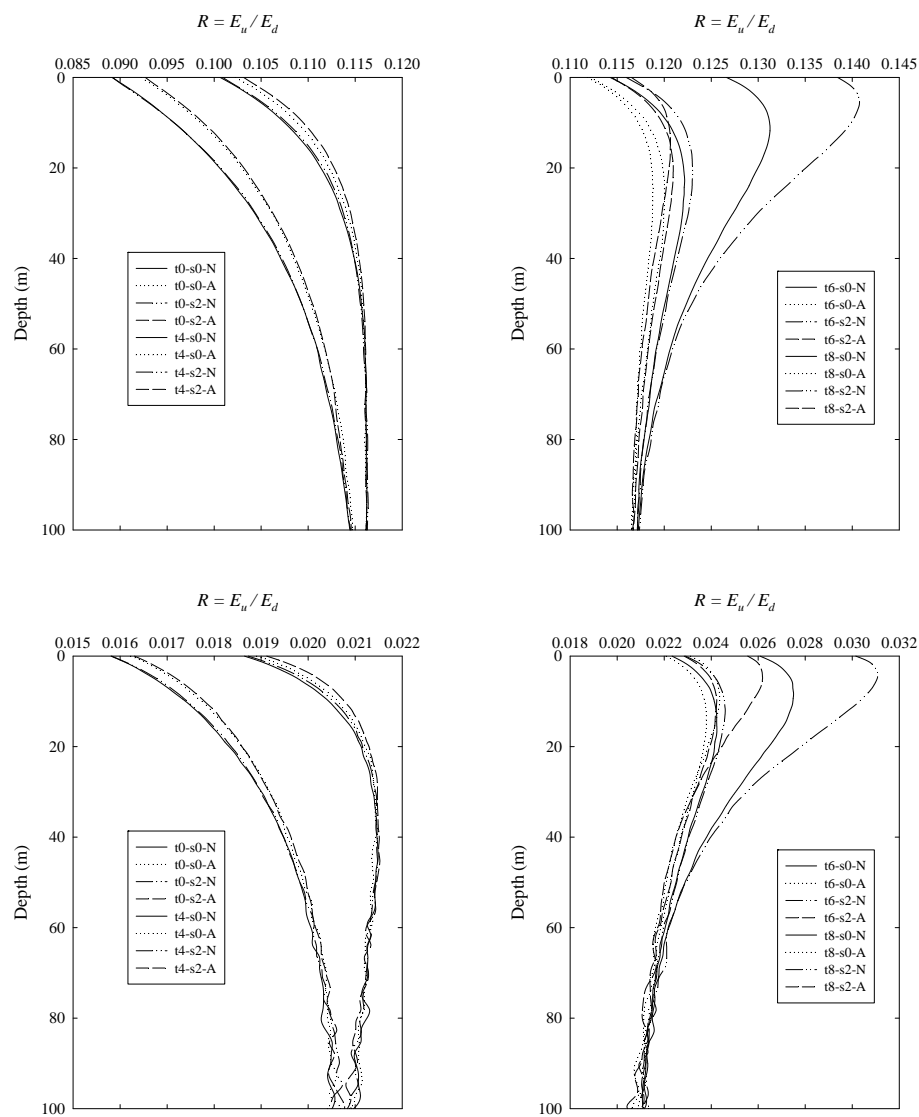


Figure 6.8: Depth profiles of the irradiance reflectance ($R = E_u/E_d$) at 440 nm (top) and 550 nm (bottom). The key to the legend describing the various curves is as follows: “t0-s0-N” means the solar zenith angle is $\theta_0 = 0^\circ$ (t0) the surface slope standard deviation is $\sigma = 0$ (s0) and there is no atmosphere (N), e.g., the case labeled “t8-s2-A” refers to $\theta_0 = 80^\circ$, $\sigma = 0.2$ with an atmosphere present over the water.

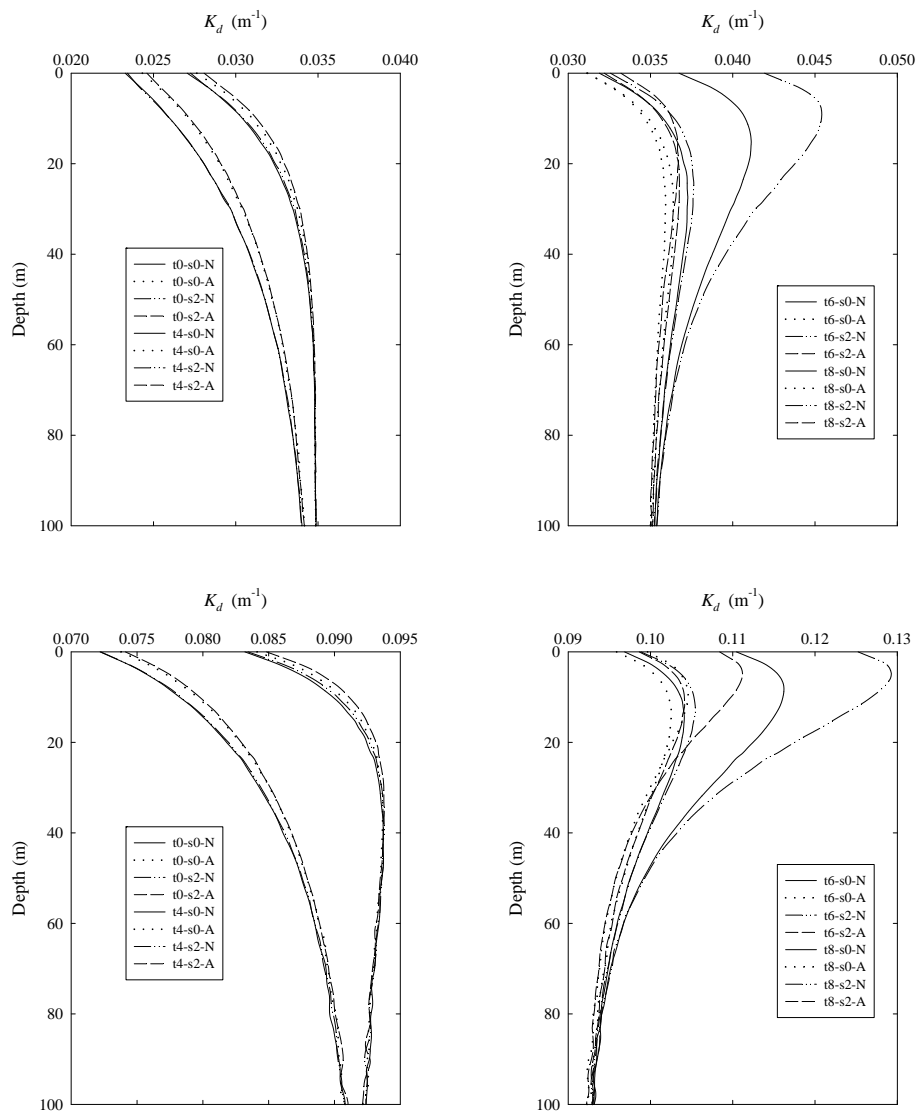


Figure 6.9: Depth profiles of the irradiance attenuation coefficient K_d at 440 nm (top) and 550 nm (bottom). The key to the legend is the same as in Figure 6.8.

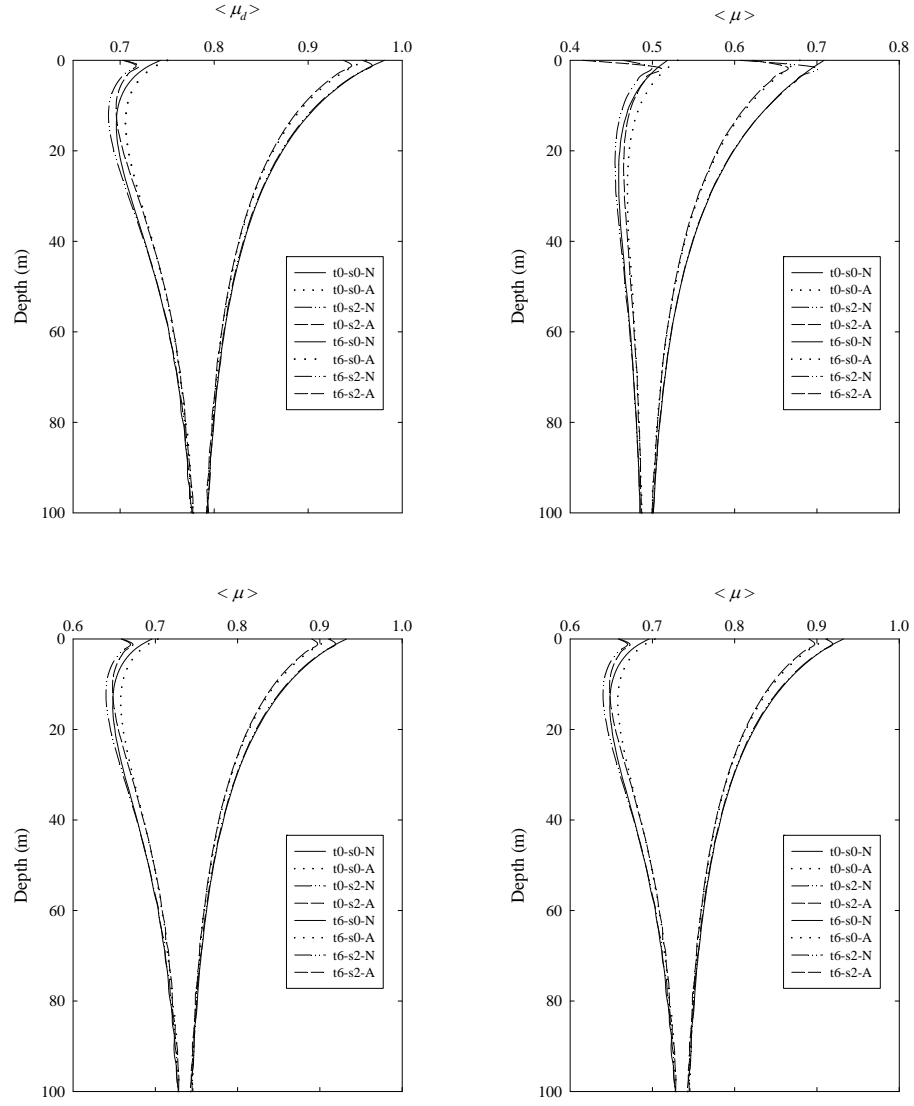


Figure 6.10: Depth profiles of average cosine of the downwelling light field $\langle \mu_d \rangle = E_d/E_{0d}$ (top: 440 nm left, 550 nm right) and the average cosine of the entire light field $\langle \mu \rangle = (E_d - E_u)/E_0$ (bottom: 440 nm left, 550 nm right). The key to the legend is the same as in Figure 6.8.

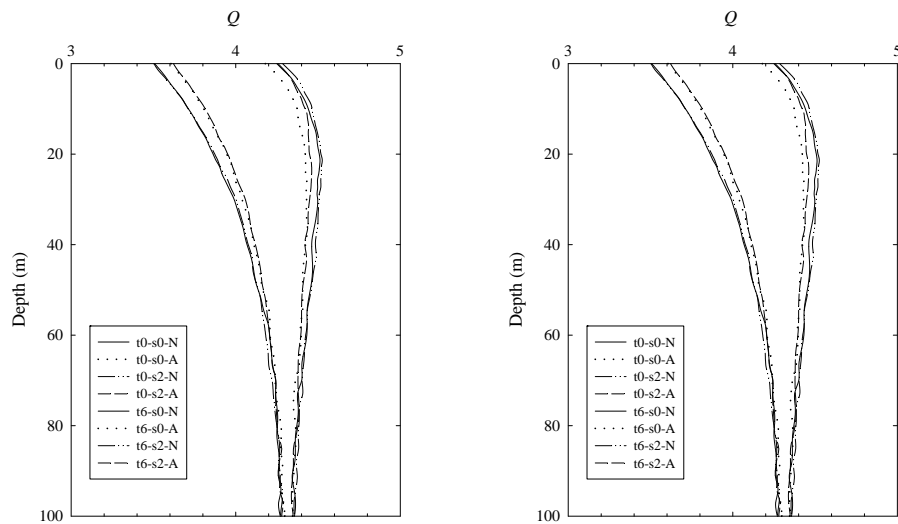


Figure 6.11: Depth profiles of $Q = E_u/L_u$ at 440 nm (left) and 550 nm (right). The key to the legend is the same as in Figure 6.8.

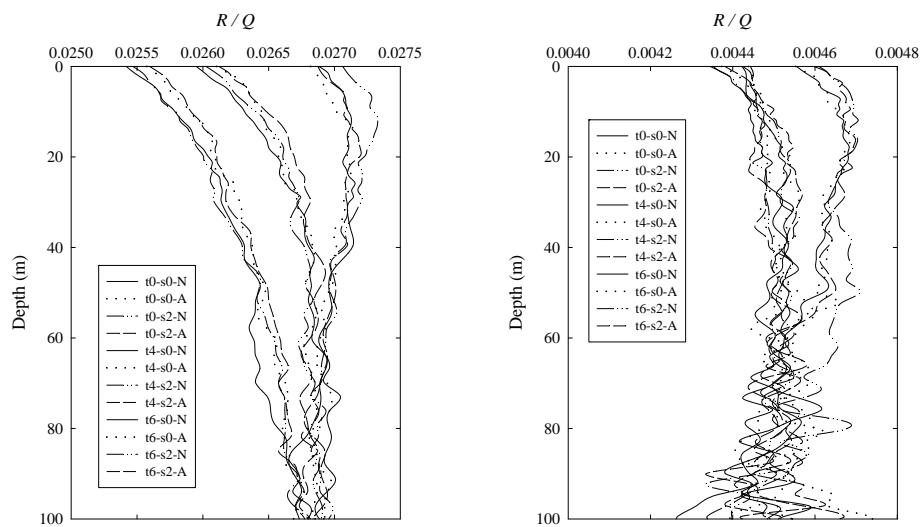


Figure 6.12: Depth profiles of $R/Q = L_u/E_d$ at 440 nm (left) and 550 nm (right). The key to the legend is the same as in Figure 6.8.

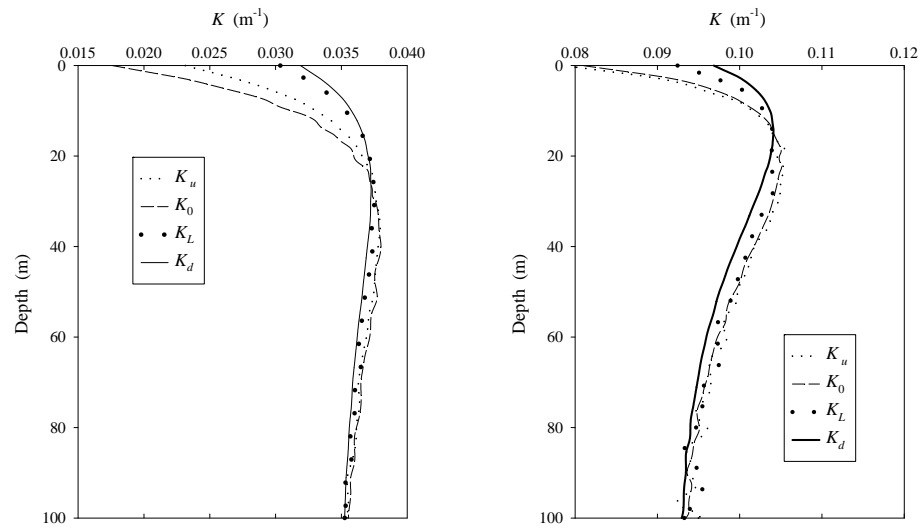


Figure 6.13: Depth profiles of the various K functions (left, 440 nm; right, 550 nm). This is for $\theta_0 = 60^\circ$ and $\sigma = 0$, no atmosphere.

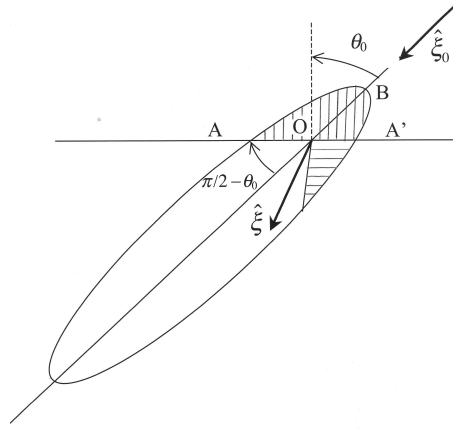


Figure 6.14: Schematic for help understanding $P_{ru}(u_0, \phi_0)$ and $P_{rd}(u_0, \phi_0)$. Sunlight is propagating along $\hat{\xi}_0$ making an angle θ_0 with the vertical (the dashed line) and then scattered in the medium in the direction $\hat{\xi}$. The elliptical-looking curve represents the scattering phase function on a polar plot (phase function in a given direction is proportional to the distance from the point O in that direction, and the point B is direct backscattering). The surface of the medium is along the line AA'. $P_{ru}(u_0, \phi_0)$ is found by accounting for all directions in which the vertical component of $\hat{\xi}$ is upward (the vertically striped volume). The portion without striping is replaced by a Dirac Delta function at $\hat{\xi} = \hat{\xi}_0$. Thus, $P_{rd}(u_0, \phi_0)$ is the volume represented by the horizontal striping. Taken together, $P_{ru}(u_0, \phi_0) + P_{rd}(u_0, \phi_0)$ is the integral of the phase function over solid angles from $\theta = \pi/2 - \theta_0$ to π and $\phi - \phi_0 = 0$ to 2π .

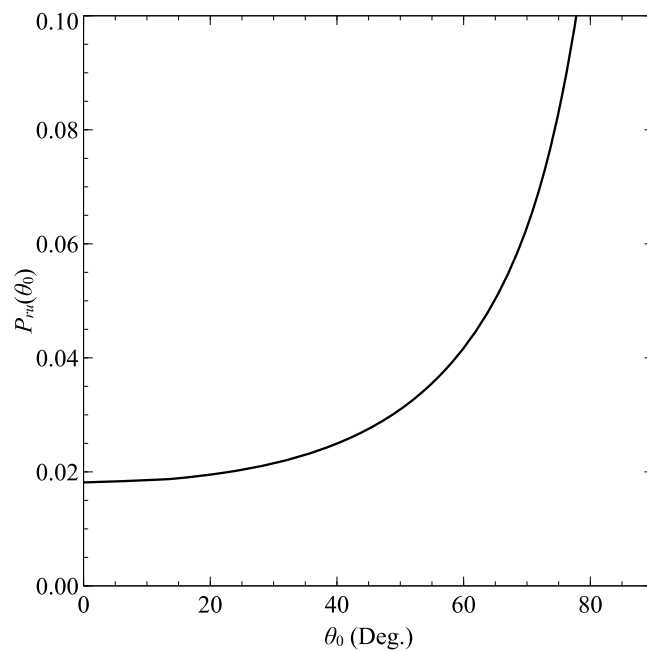


Figure 6.15: $P_{ru}(u_0, \phi_0)$ as a function of the solar zenith angle.

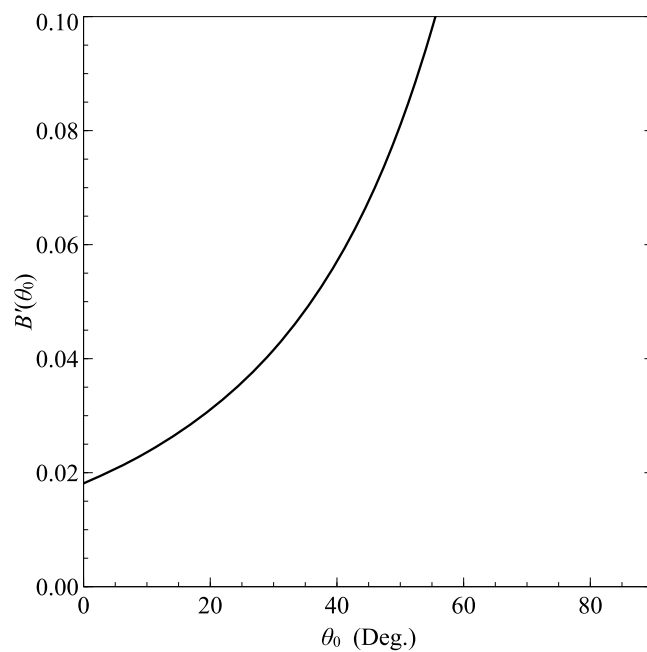


Figure 6.16: B' as a function of the solar zenith angle.

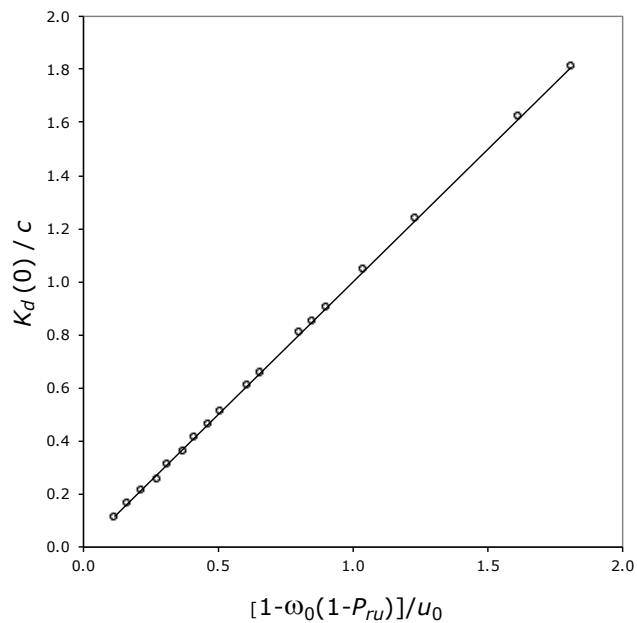


Figure 6.17: Exact $K_d(0)/c$ as a function of the QSSA prediction, $[1 - \omega_0(1 - P_{ru})]/u_0$, for $\theta_0 = 0$ and 60° . The line is a linear regression showing the near equality of these two quantities. Its slope is 0.9999 and $R^2 = 0.9998$.

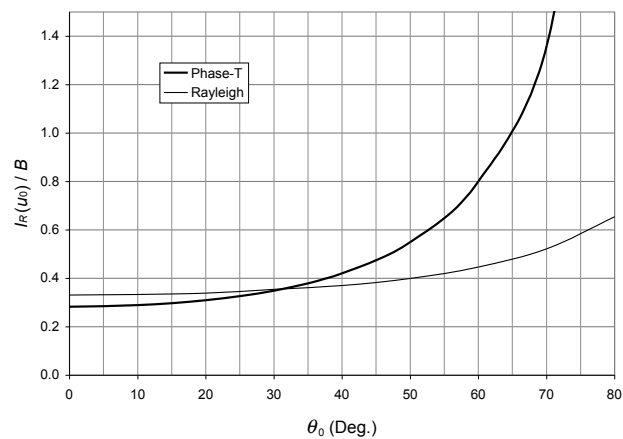


Figure 6.18: Value of the integral $I_R(u_0)/B = I'_R(u_0)$ appearing in the QSSA for $R(0)$ as a function of the solar zenith angle for Phase-T and for Rayleigh scattering.

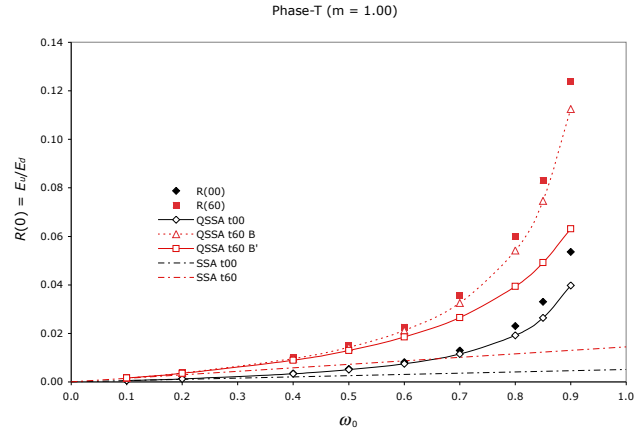


Figure 6.19: Comparison between “exact” computations of $R(0)$ (solid symbols), QSSA predictions (solid or dotted lines with open symbols) and SSA predictions (dash-dot lines). The notation “t00” refers to $\theta_0 = 0^\circ$ and “t60” to 60° .

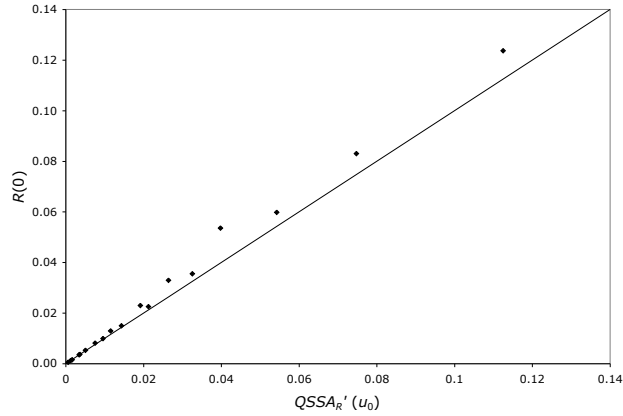


Figure 6.20: Comparison between “exact” computations of $R(0)$ (symbols) and the function $QSSA'_R(u_0)$. Both use the Petzold Phase-T phase function.

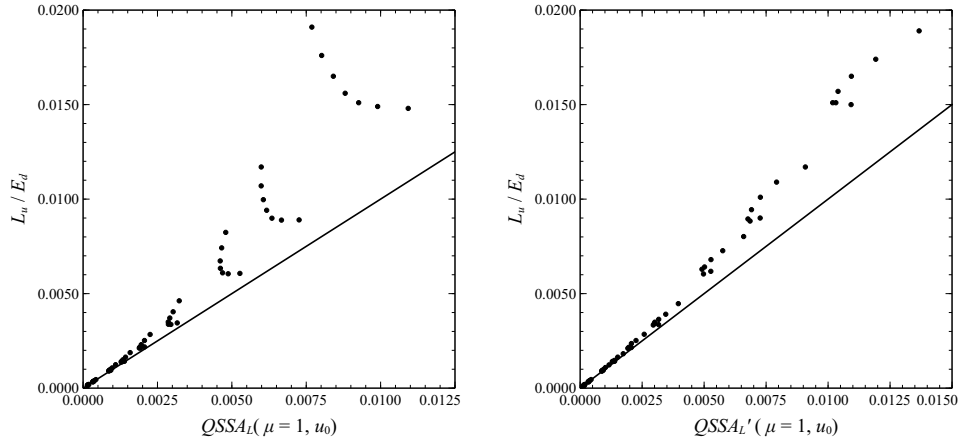


Figure 6.21: Left Panel: Exact computations of the zenith propagating radiance exiting the medium divided by the downwelling irradiance, for several values of ω_0 and solar zenith angles, compared with $QSSA_L(\mu = 1, u_0)$. The Petzmas phase function was used in the computations. Right Panel: Same as the left panel, but $B'(u_0)$ in $QSSA_L(\mu = 1, u_0)$ is replaced by B , the backscattering probability, forming $QSSA'_L(\mu = 1, u_0)$. The solid line represents one-to-one agreement.

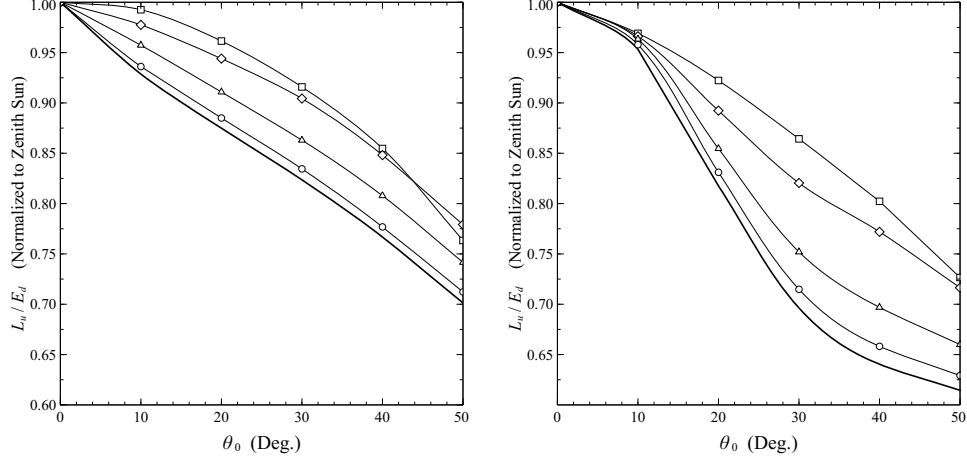


Figure 6.22: Exact computations for zenith-propagating radiance with the Petzmas phase function (left panel) and the Petzold phase function (right panel) as a function of the solar zenith angle θ_0 . The computations are normalized to the value of L_u/E_d for $\theta_0 = 0$. Curves from bottom to top correspond to $QSSA'_L$ (thicker line) and $\omega_0 = 0.2, 0.5, 0.8, \text{ and } 0.9$.

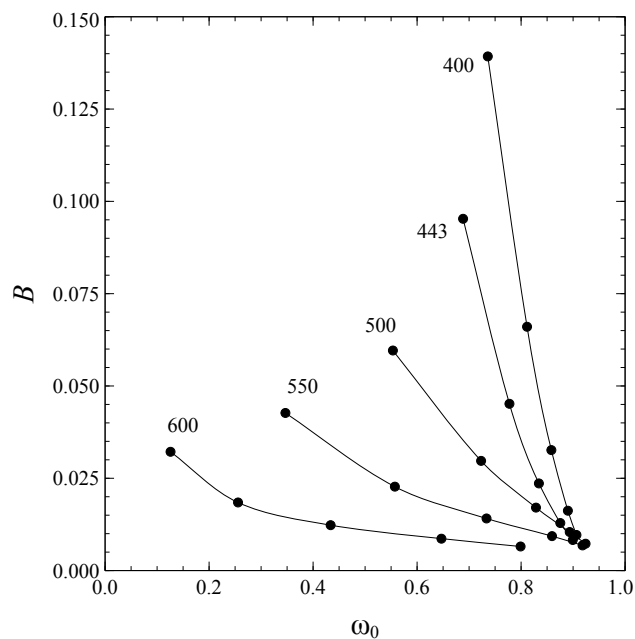


Figure 6.23: The correlation between values of B and ω_0 for the IOP-Chlorophyll model described in Chapter 5. Each curve is for a specific wavelength of the light, written next to the curve (in nm). The points from left to right on each curve are for Chlorophyll a concentrations of 0.03, 0.10, 0.30, 1.0, and 3.0 mg.m^3 .

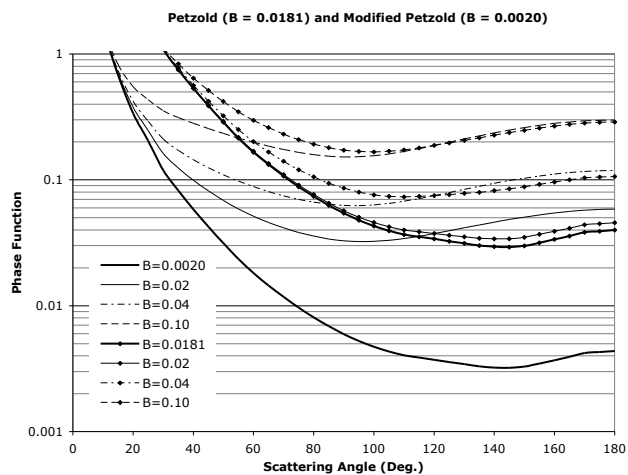


Figure 6.24: Phase functions for particles and for particles *plus* water combined to yield the indicated value of B .

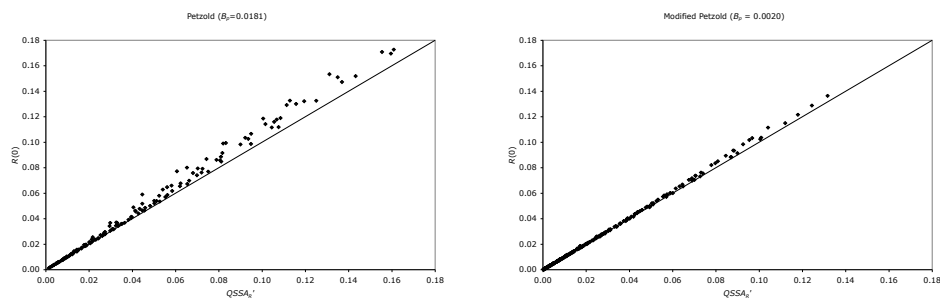


Figure 6.25: Comparison of exact computations of $R(0)$ and the $QSSA'_R$ predictions for $B_p = 0.0181$ (left) and 0.002 (right). The diagonal line represents perfect agreement between the two.

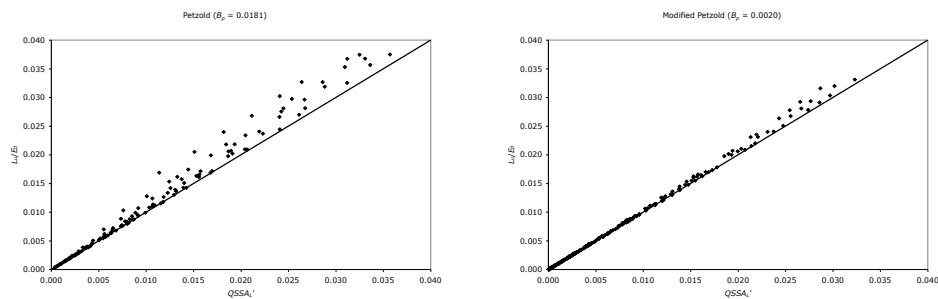


Figure 6.26: Comparison of exact computations of L_u/E_d and the $QSSA'_L$ predictions for $B_p = 0.0181$ (left) and 0.002 (right). The diagonal line represents perfect agreement between the two.

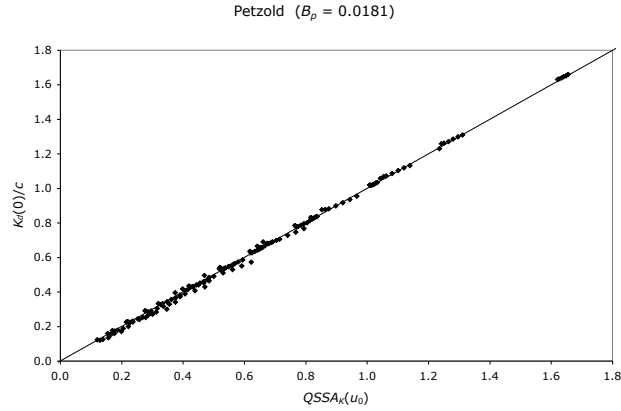


Figure 6.27: Comparison of exact computations of $K_d(0)$ and the $QSSA_K$ predictions for $B_p = 0.0181$. The diagonal line represents perfect agreement between the two. The parameters are identical to the left panel in Figure 6.26.

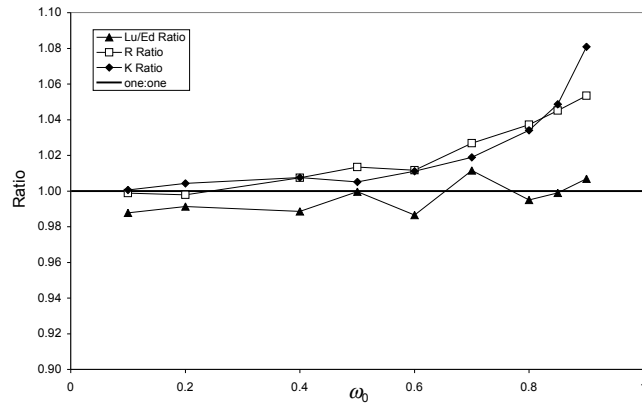


Figure 6.28: Influence of the refracting surface on L_u/E_d , R , and K_d , all evaluated just beneath the interface. The plot provides the ratio of these quantities for $m = 1.333$ and $m = 1.000$. The solar zenith angle *beneath the surface* is the same for both indices (40°). Phase-T is used in the simulations.

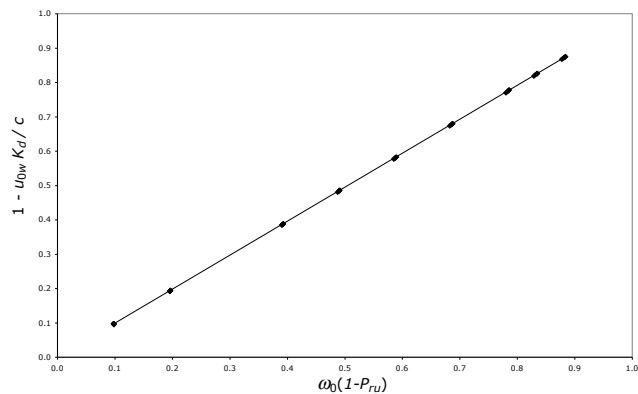


Figure 6.29: Linear fit of K_d to Eq. (6.28) for the Petzmas phase function with ω_0 ranging from 0.1 to 0.9 and θ_0 from 0 to 58° , and $m = 1.333$.

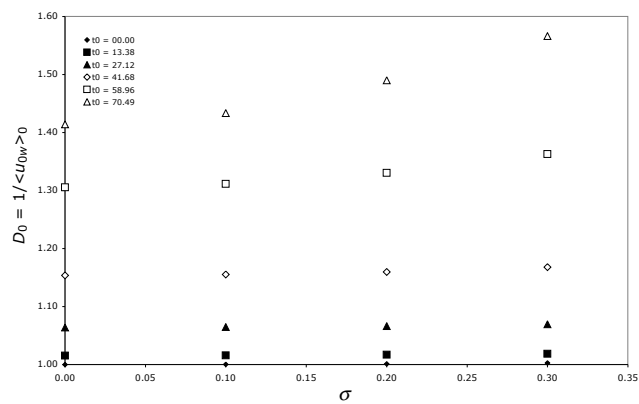


Figure 6.30: Simulated values of $D_0 = 1/\langle u_{0w} \rangle_0$ as a function of the surface roughness parameter. (“t0” stands for θ_0 in degrees). There is no atmosphere.

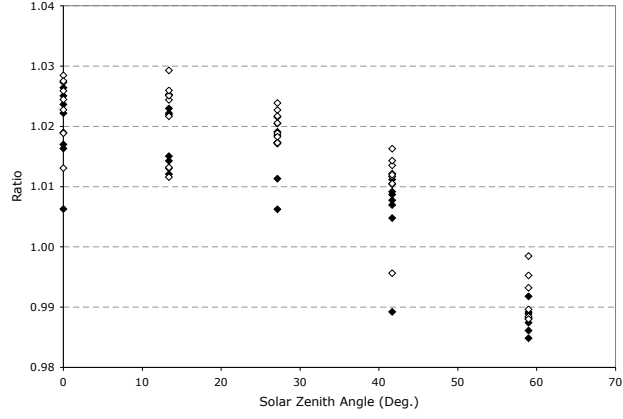


Figure 6.31: The ratio $(L_u/E_d)_{\text{Atm}} \div (L_u/E_d)_{\text{NoAtm}}$ at 400 nm for the Petzmas phase function with ω_0 ranging from 0.1 to 0.9, θ_0 from 0 to 58° ($\theta_{0w} = 0$ to 40°) and $\sigma = 0$ (filled symbols) and 0.2 (open symbols) as described in the text.

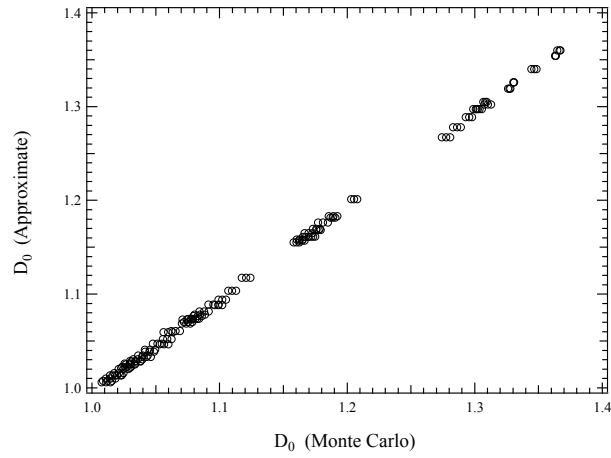


Figure 6.32: The value of D_0 computed using Eqs. (6.32) through (6.38), and the “exact” value determined through Monte Carlo simulations.

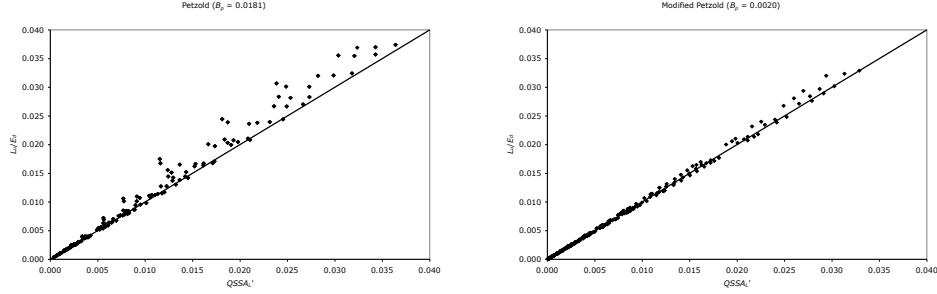


Figure 6.33: Comparison of exact computations of L_u/E_d and the $QSSA'_L$ predictions for $B_p = 0.0181$ (left) and 0.002 (right). The diagonal line represents perfect agreement between the two. This utilizes $\langle u_{0w} \rangle$ (or $1/D_0$) computed using Eqs. (6.34) through (6.38).

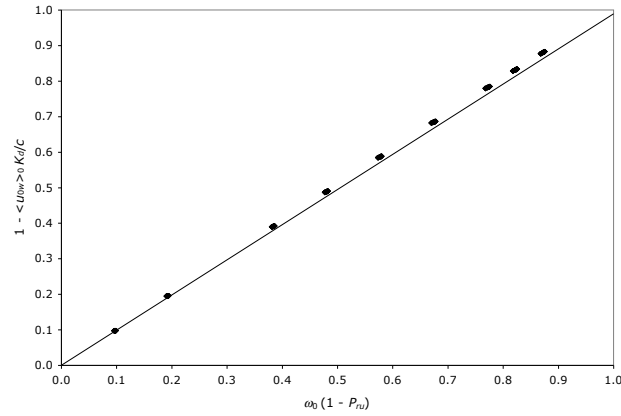


Figure 6.34: The influence of the atmosphere and water surface on K_d for the Petzmas phase function with ω_0 ranging from 0.1 to 0.9, θ_0 from 0 to 58° and $\sigma = 0$ and 0.2. The atmospheric parameters are provided in the text. The line is the fit of K_d to Eq. (6.28) in the absence of the atmosphere. Note that u_{0w} in Figure 6.29 has been replaced by $\langle u_{0w} \rangle_0$ here.

To compute L_u :

$$L_u = E_d QSSA'_L(\langle u_{0w} \rangle_0)$$

$$QSSA'_L(\langle u_{0w} \rangle_0) = \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \frac{P_r(\langle u_{0w} \rangle_0 \rightarrow -1)}{1 + \langle u_{0w} \rangle_0}$$

To compute E_d :

$$E_d(\lambda) = E_d^+(\lambda) t_f (1 - 0.48 R)^{-1}, \quad R \approx \pi QSSA'_L(\langle u_{0w} \rangle_0)$$

$$E_d^+(\lambda) = u_0 F_0(\lambda) \exp[-\tau_r(\lambda)/(x(\lambda)u_0)]$$

$$x(\lambda) \approx 2.3 - 0.001(\lambda - 400)$$

To compute $\langle u_{0w} \rangle_0$:

$$\langle u_{0w} \rangle_0 \stackrel{\text{def}}{=} 1/D_0$$

$$D_0 = f D_0^S + (1 - f) D_0^\oplus$$

$$D_0^\oplus = C_{D0} + C_{D1}\sigma + C_{D2}\sigma^2,$$

$$C_{D0} = 1/u_{0w},$$

$$C_{D1} \approx 0.0155 (C_{D0} - 1),$$

$$C_{D2} \approx 0.03 - 0.1831 (C_{D0} - 1) + 7.268 (C_{D0} - 1)^2.$$

$$D_0^S \approx 1.1876 + 0.1505\sigma + 0.675\sigma^2$$

$$f = \frac{gT^S}{gT^S + (1 - g)T^\oplus}$$

$$g(\lambda) = 1 - \exp \left[-\frac{\tau_r(\lambda)}{u_0} \left(1 - \frac{1}{x(\lambda)} \right) \right].$$

$$T^S \approx 0.933 + 0.04\sigma$$

$$T^\oplus = t_f(u_0)$$

Figure 6.35: Summary of equations to compute L_u using the modified QSSA in the presence of the water surface and the atmosphere. Note that this uses $C_L = 1$ in Eq. (6.27).

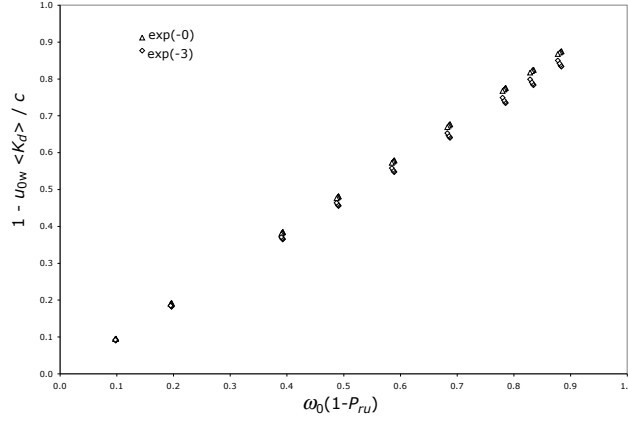


Figure 6.36: The relationship between the average K_d between the surface and a depth for which $E_d(z)/E_d(0) = \exp(-3)$ (diamonds). Included for comparison is $K_d(0)$ (triangles). The computations are for all combinations of $\omega_0 = 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85$, and 0.9 , and $\theta_{0w} = 0, 10^\circ, 20^\circ, 30^\circ$, and 40° .

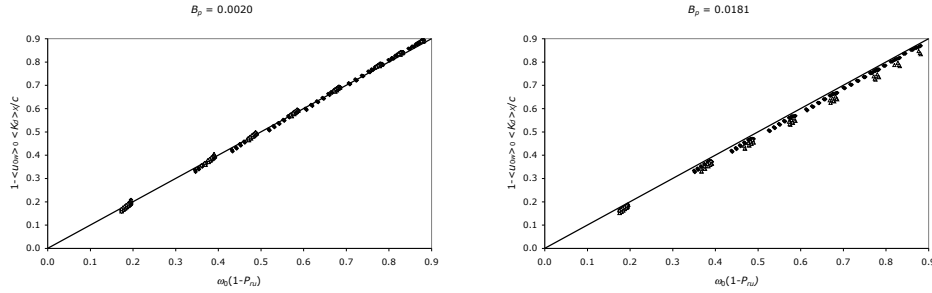


Figure 6.37: Comparison of exact computations of $1 - \langle u_{0w} \rangle_0 \langle K_d \rangle_X / c$ with QSSA predictions for $B_p = 0.0020$ and $B_p = 0.0181$. The total backscattering probability B varies from 0.0020 to 0.12 . Solid line represents exact agreement between the Monte Carlo simulation and the QSSA predictions. Filled symbols are for $X = 0$ and open triangles are for $X = 3$.

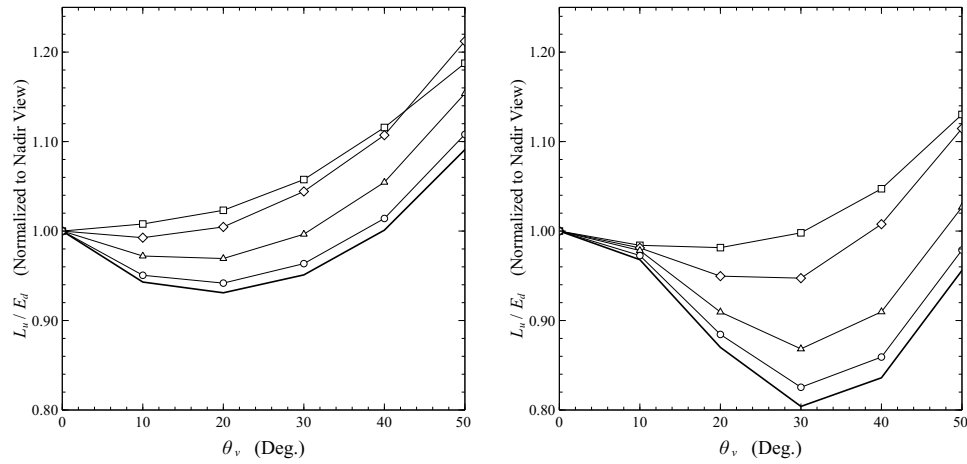


Figure 6.38: L_u/E_d for viewing radiance propagating in the direction θ_v (measured relative to the *upward* normal to the surface) computed with the Petzmas phase function (left panel) and the Petzold phase function (right panel) as a function of the viewing angle θ_v . The Sun is at the zenith. The computations are normalized to the value of L_u/E_d for $\theta_v = 0$ and were computed from Figure 6.22 using reciprocity. Curves from bottom to top correspond to the QSSA and $\omega_0 = 0.2, 0.5, 0.8$, and 0.9 .

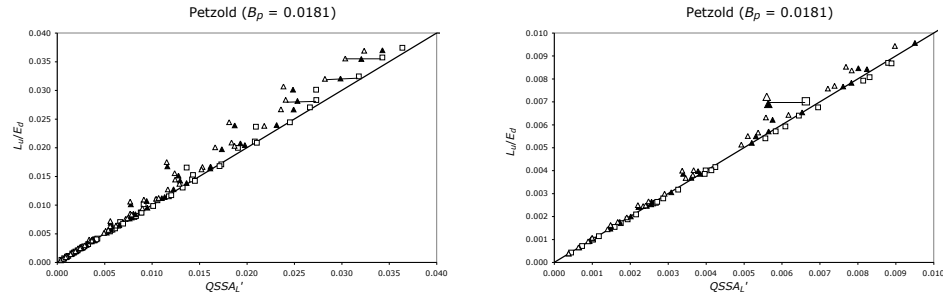


Figure 6.39: Comparison of exact computations of L_u/E_d and the $QSSA'_L$ predictions for $B_p = 0.0181$. The right panel is the expanded lower left part of the left panel. The diagonal line represents perfect agreement between the two. This utilizes $\langle u_{0w} \rangle$ (or $1/D_0$) computed using Eqs. (6.34) through (6.38). The open squares are for $\theta_0 = 10^\circ$, the solid triangles $\theta_0 = 37^\circ$, and the open triangles $\theta_0 = 60^\circ$. Horizontal lines connect points of equivalent IOPs.

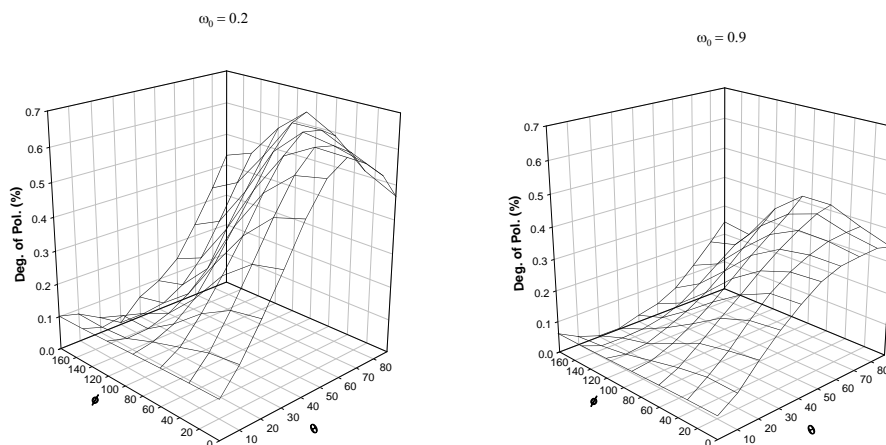


Figure 6.40: Angular distribution of the degree of polarization of $L_u(0, \theta_{vw}, \phi_{vw})$ for $\omega_0 = 0.2$ and 0.9 with the Petzold Phase-T phase function and the Voss and Fry normalized (to P_{11}) Mueller phase matrix. The solar zenith angle is 40° . The atmosphere is aerosol free and the wavelength is 500 nm .

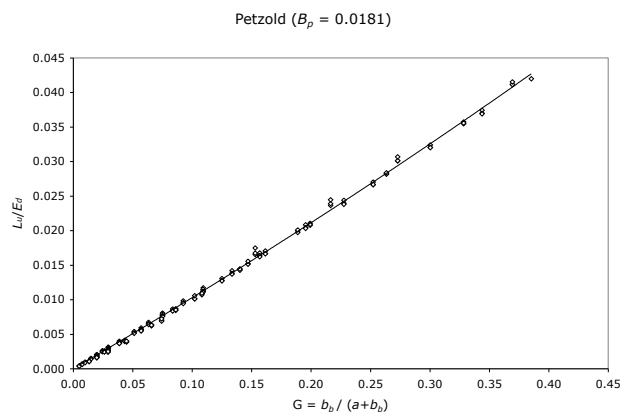


Figure 6.41: L_u/E_d as a function of G for the same data as in Figure 6.39. The line represents a least-squares fit: $L_u/E_d = 0.1001 G + 0.0278 G^2$.

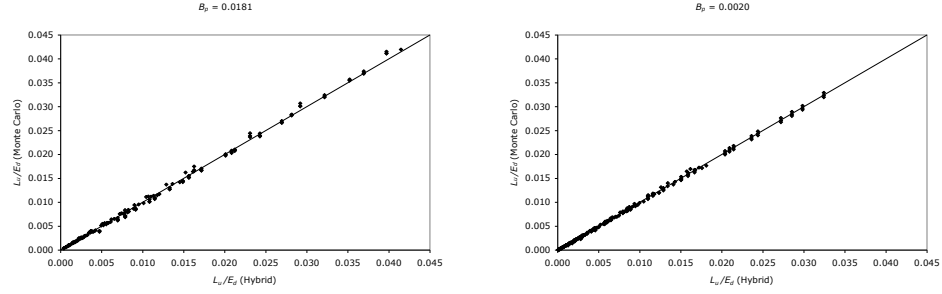


Figure 6.42: Performance of the "Hybrid" model (Eq. (6.45)) for the same cases as used in Figure 6.33. For the figures, $g_1 = 0.1054$ and $g_2 = 0.0057$.

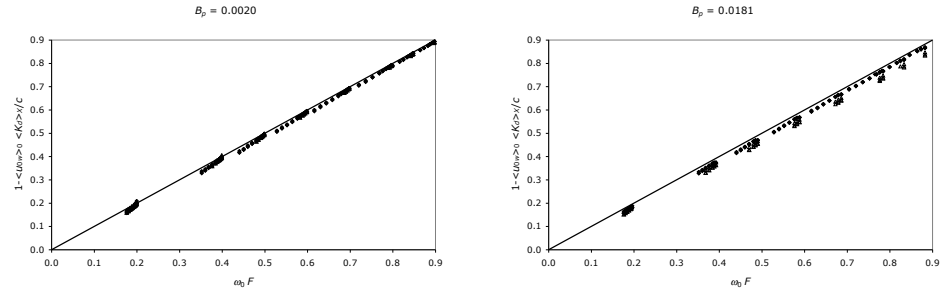


Figure 6.43: Comparison of exact computations of $1 - \langle u_{0w} \rangle_0 \langle K_d \rangle_X / c$ with $\omega_0 F$ predictions for $B_p = 0.0020$ and $B_p = 0.0181$. The total backscattering probability B varies from 0.0020 to 0.12. Solid line represents exact agreement between the Monte Carlo simulation and the prediction of Eq. (6.46) with $\langle C'_K \rangle_X = 1$. Filled symbols are for $X = 0$ and open triangles are for $X = 3$.

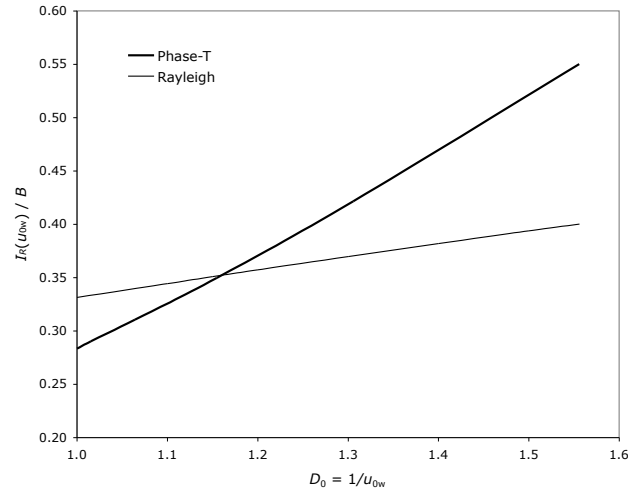


Figure 6.44: $I'_R = I_R/B$ for the Phase-T and Rayleigh scattering phase functions.

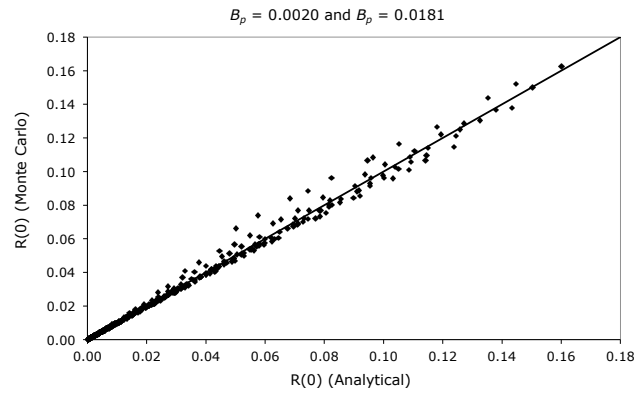


Figure 6.45: Comparison between Monte Carlo-simulated $R(0)$ and that computed using Eq. (6.47). The Monte Carlo results include the atmosphere and water surface.

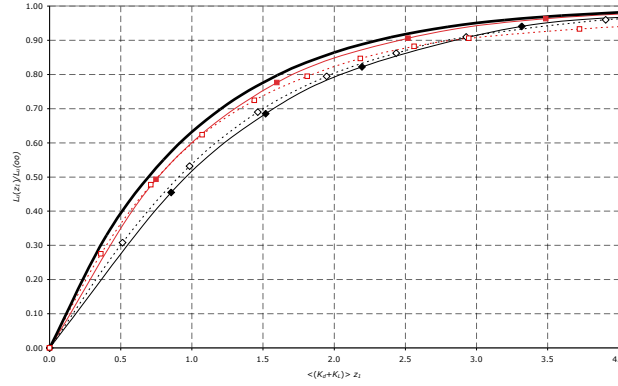


Figure 6.46: $L_u(z_1)/L_u(\infty)$ as a function of $\langle K_d + K_L \rangle z_1$. The parameters of the simulations are those in Figure 6.9. The red points connected by thin lines are for $\theta_0 = 0$ with the solid line corresponding to 440 nm and the dashed to 550 nm. Similarly, the black points and thin lines are for $\theta_0 = 60^\circ$. The thick black line is the prediction of Eq. (6.54).

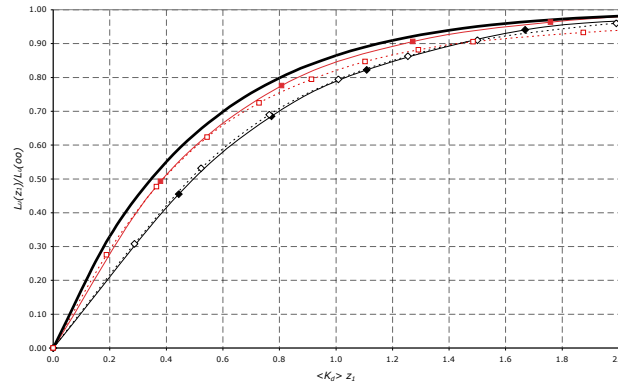


Figure 6.47: $L_u(z_1)/L_u(\infty)$ as a function of $\langle K_d \rangle z_1$. The points and thin lines are representative of the same quantities as in Figure 6.46. The thick black line is the prediction of Eq. (6.55).

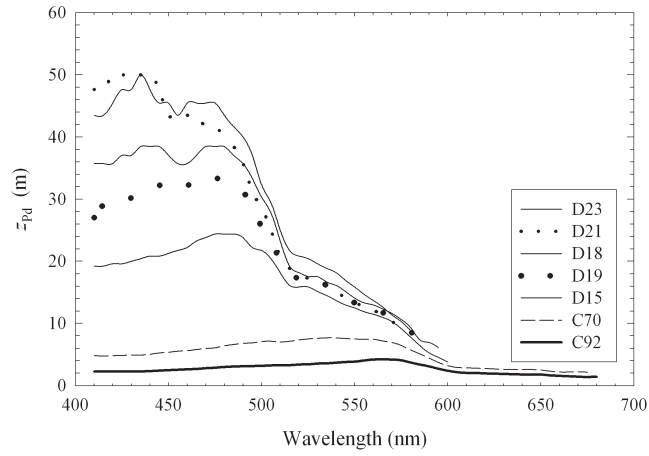


Figure 6.48: The penetration depth for the various stations presented in Figures 6.5 and 6.6. Here z_{Pd} is taken to be $1/K_d$, and K_d is taken from Figure 6.6.

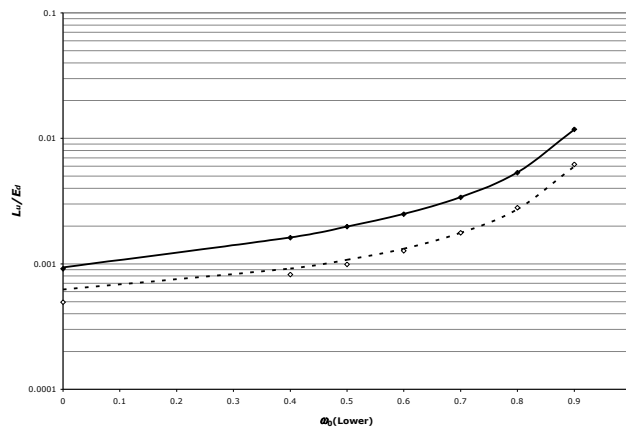


Figure 6.49: Comparison between exact computations (symbols) and Eq. (6.60) (lines) for the two-layer case described in the text. Solid lines and symbols are for $\theta_0 = 0$ and dashed lines and open symbols are for $\theta_0 = 60^\circ$.

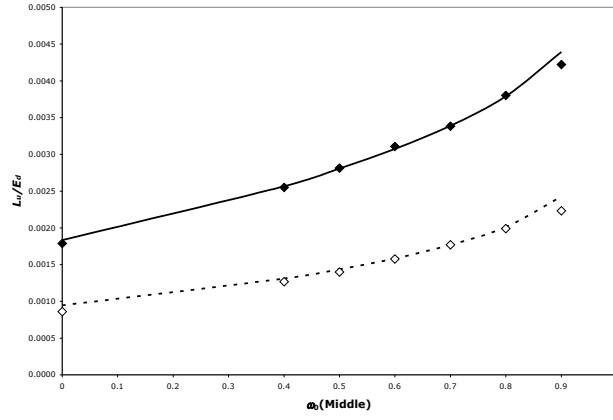


Figure 6.50: Comparison between exact computations (symbols) and Eq. (6.60) (lines) for the three-layer case described in the text. Solid lines and symbols are for $\theta_0 = 0$ and dashed lines and open symbols are for $\theta_0 = 60^\circ$.

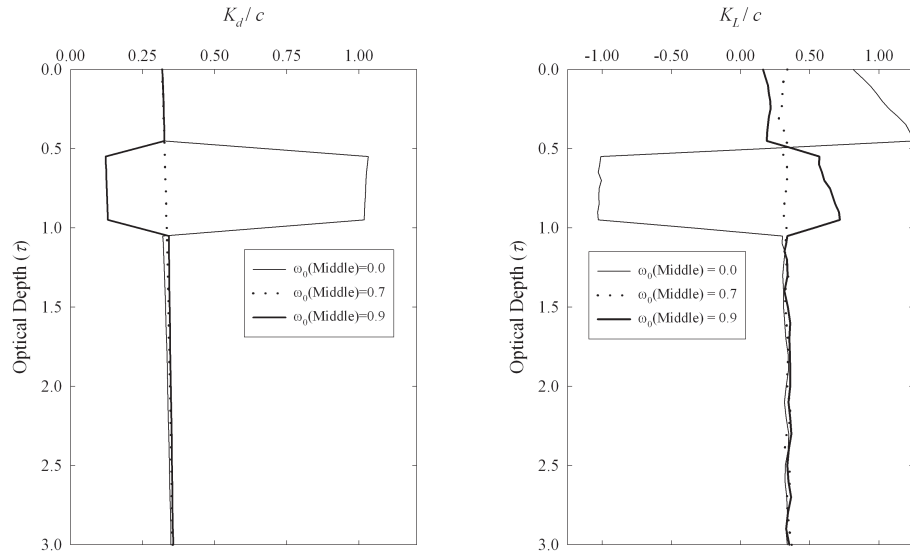


Figure 6.51: Examples of K_d (left panel) and K_L (right panel) for a three-layer medium with $\theta_0 = 0$. The value of ω_0 for the top and bottom layers is 0.7. The value for the middle layer is indicated in the legend.

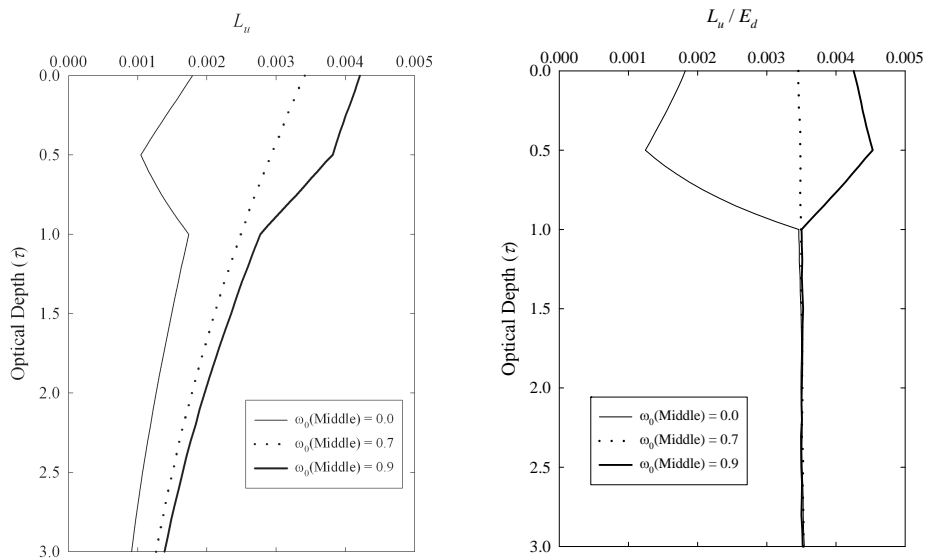


Figure 6.52: Zenith-propagating radiance, L_u , (normalized to unit irradiance incident on the water surface) for a three-layer medium with $\theta_0 = 0$ (Left Panel). Radiance reflectance L_u/E_d for the same cases (Right Panel). The value of ω_0 for the top and bottom layers is 0.7. The value for the middle layer is indicated in the legend.

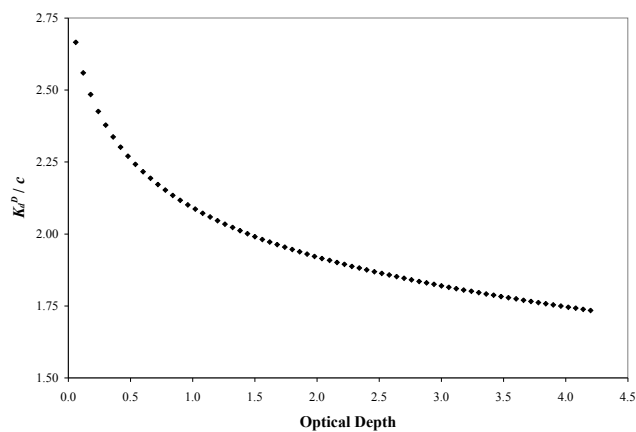


Figure 6.53: Attenuation of downward irradiance that originates from the reflection of totally diffuse upward irradiance by the air-water interface.

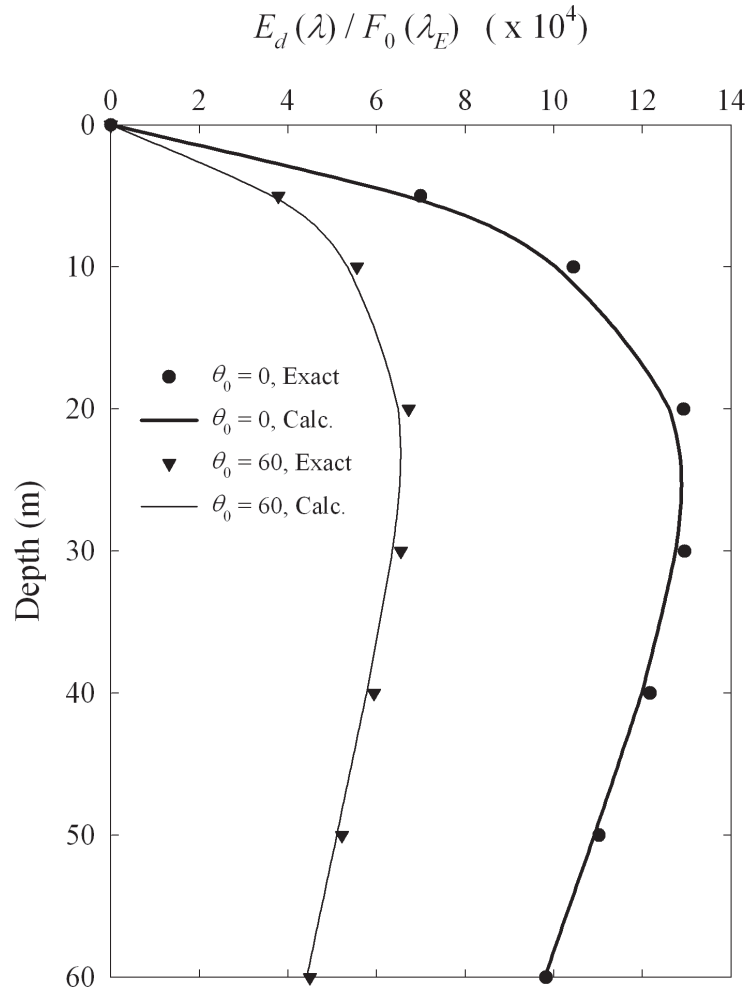


Figure 6.54: The Raman-induced $E_d(z)$ at 550 nm from exact computations (symbols) along with the results from Eq. (6.73) (lines).

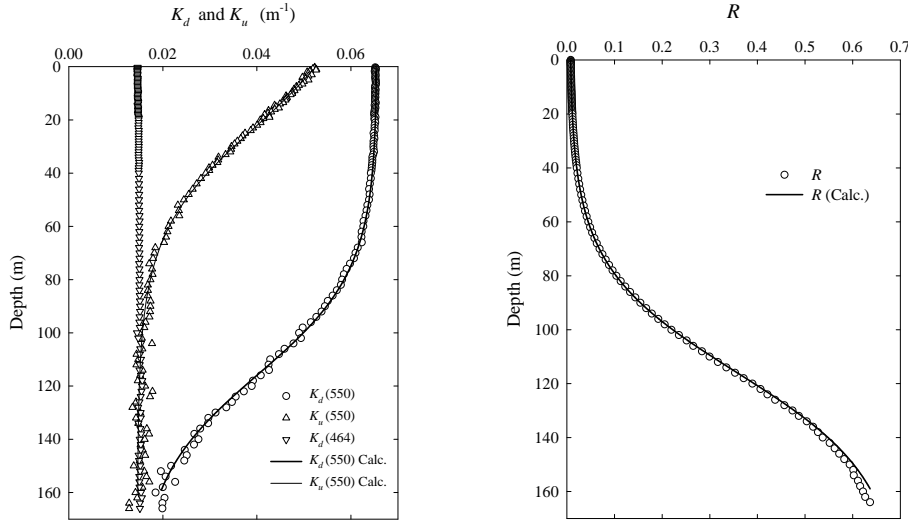


Figure 6.55: Zenith-propagating radiance, L_u , (normalized to unit irradiance incident on the water surface) for a three-layer medium with $\theta_0 = 0$ (Left Panel). Radiance reflectance L_u/E_d for the same cases (Right Panel). The value of ω_0 for the top and bottom layers is 0.7. The value for the middle layer is indicated in the legend. K-R-550 Left Panel: exact computations of the K_d and K_u at 550 nm along with K_d at 464 nm for a water body consisting of pure water only (symbols). Right Panel: exact computations of reflectance at 550 for the same water body (symbols). The sky was aerosol free and the solar zenith angle was 36° . The lines correspond to predictions of the lowest-order model described in the text.

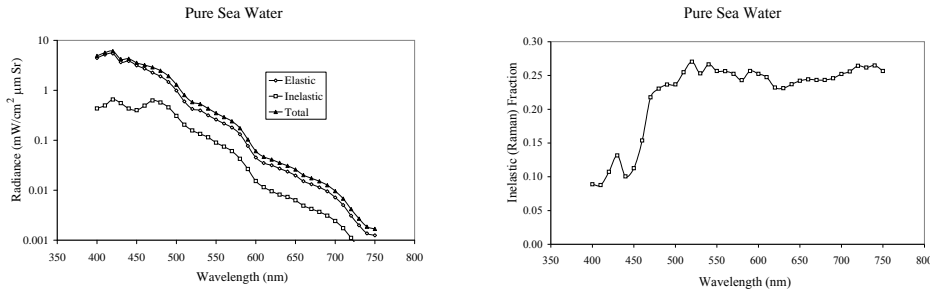


Figure 6.56: Left Panel: The elastic and inelastic (Raman) contributions to the zenith-propagating radiance, just beneath the surface of a water body consisting only of pure seawater. The atmosphere was aerosol free and the solar zenith angle was 37° . Right panel: The fraction of the total radiance in the left panel that is contributed by Raman scattering.

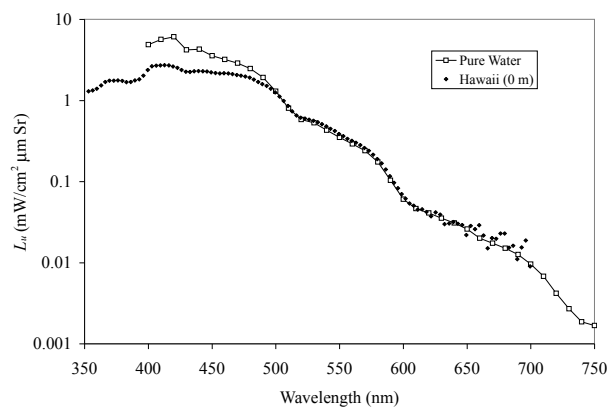


Figure 6.57: Upwelling radiance just beneath the surface at Hawaii derived from the full data set, part of which was used to produce Figure 6.1, (filled symbols) along with the theoretical upwelling radiance (open symbols) for a water body consisting only of pure (sea) water below an aerosol-free atmosphere. Both elastic and Raman scattering are accounted for in the theoretical calculation.

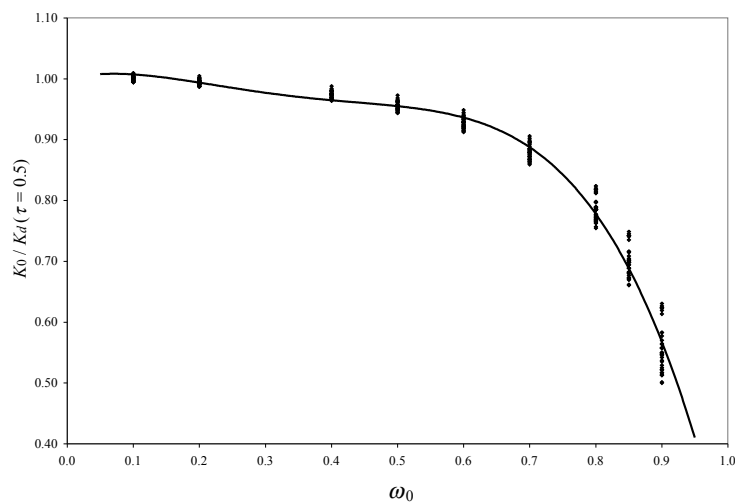


Figure 6.58: K_0/K_d at $\tau = 0.5$ for several values of ω_0 , refracted solar zenith angles (0° , 10° , 20° , 30° , and 40°), and $\sigma = 0$ and 0.2 in the absence of the atmosphere. In addition, the same water and surface properties are combined with a model atmosphere at $\lambda = 400$, 500 and 600 nm. The model atmosphere's aerosol optical thicknesses $\tau_a(\lambda)$ were 0.126 , 0.107 , and 0.094 and Rayleigh optical thicknesses $\tau_r(\lambda)$ were 0.364 , 0.145 , and 0.069 , respectively at, $\lambda = 400$, 500 , and 600 nm.

To Compute the Absorption Coefficient:

$$a(\lambda) = a_w(\lambda) + a_p(\lambda) + a_y(\lambda) + b_{\text{Raman}}(\lambda_E \text{ Band} \rightarrow \lambda)$$

$$a_p(\lambda) = f_{ap} a_{p1}(\lambda) C^{E_p(\lambda)}$$

$$a_y(\lambda) = a_y(440) \exp[-S_y(\lambda - 440)]; \quad S_y = 0.014 \text{ nm}^{-1}$$

$$a_y(440) = 0.2 \times [a_w(440) + a_p(440)]$$

To Compute the Scattering Coefficient:

$$b(\lambda) = b_w(\lambda) + b_p(\lambda)$$

$$b_p(\lambda) = f_{bp} b_{p1}(550) \left(\frac{\lambda}{550} \right)^v C^{0.62}; \quad b_{p1}(550) = 0.30 \text{ m}^{-1}$$

$$v = 0.5(\log_{10} C - 0.3) \text{ for } C \leq 2 \text{ mg/m}^3 \text{ and } v = 0 \text{ otherwise}$$

To Compute the Backscattering Probability:

$$B(\lambda) = \frac{B_w b_w(\lambda) + B_p b_p(\lambda)}{b_w(\lambda) + b_p(\lambda)}; \quad B_w = 0.5$$

$$B_p(\lambda) = \alpha_s B_{ps} + (1 - \alpha_s) B_{p\ell}; \quad B_{ps} = 0.0140, \quad B_{p\ell} = 0.0019$$

$$\alpha_s = \frac{0.0020 + 0.01 \left(\frac{1}{2} - \frac{1}{4} \log_{10}(C) \right) - B_{p\ell}}{B_{ps} - B_{p\ell}}$$

To Compute the Scattering Phase Function:

$$P(\Theta, \lambda) = \frac{P_w(\Theta) b_w(\lambda) + P_p(\Theta) b_p(\lambda)}{b_w(\lambda) + b_p(\lambda)}$$

$$P_p(\Theta) = \alpha_s P_{ps}(\Theta) + (1 - \alpha_s) P_{p\ell}(\Theta)$$

$$P_{ps}(\Theta) \text{ \& } P_{p\ell}(\Theta) \text{ are Modified Petzold}$$

To Compute the Raman Scattering Parameters:

$$b_{\text{Raman}}(\lambda_E \text{ Band} \rightarrow \lambda) = \left(\frac{589}{\lambda} \right)^{4.8} \times 2.61 \times 10^{-4} \text{ m}^{-1}$$

$$\lambda_E = \frac{\lambda}{1 + 3353 \times 10^{-7} \lambda}$$

Figure 6.59: Summary of equations used to compute the IOPs required for the examples described in the text.

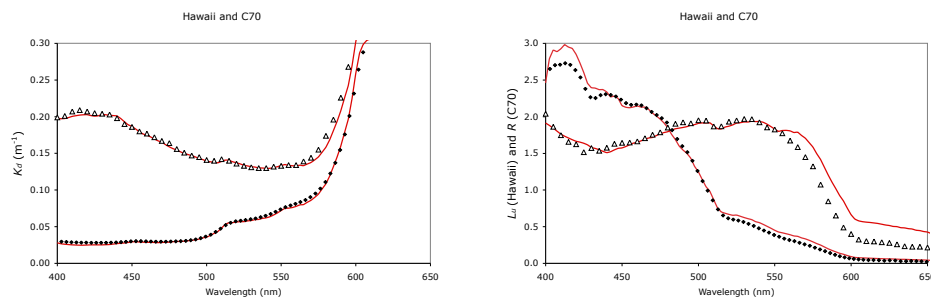


Figure 6.60: Comparison between measurements and modeling (as described in the text) for Hawaii and C70 stations shown in Figures 6.3-6.6. Open symbols and filled symbols are the measurements for C70 and Hawaii, respectively. Solid lines are the model results. On the right panel, L_u is in $\text{mW}/\text{cm}^2\mu\text{m}$ and R is in %.

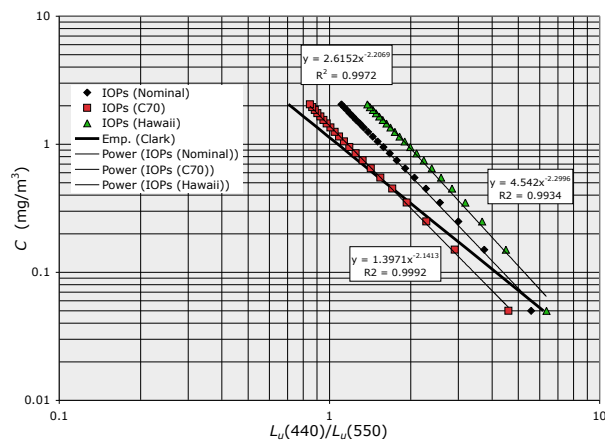


Figure 6.61: The chlorophyll concentration as a function of the ratio of L_u in the blue to that in the green, as in the initial algorithms to estimate the pigment concentration from the water color. The modeled curves using the model's nominal ($f_{ap} = f_{bp} = 1$) IOPs as well as the C70 and Hawaii IOPs are compared to the original empirical algorithm (Emp.) based on experimental data. The lines and equations represent regressions of the model results. The regression line for the experimental measurements is $y = 1.13x^{-1.71}$.

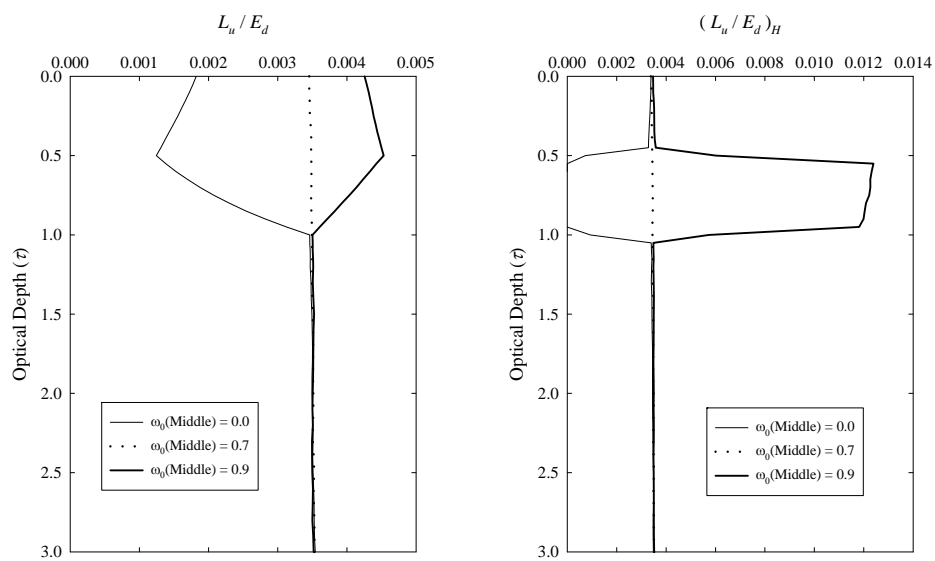


Figure 6.62: Zenith-propagating radiance reflectance, $L_u/E_d = R_L$, for a three-layer medium with $\theta_0 = 0$ (Left Panel). Radiance reflectance $(L_u/E_d)_H = R_{LH}$ after application of Eq. (6.81) to the R_L profiles (Right Panel). The value of ω_0 for the top and bottom layers is 0.7. The value for the middle layer is indicated in the legend.

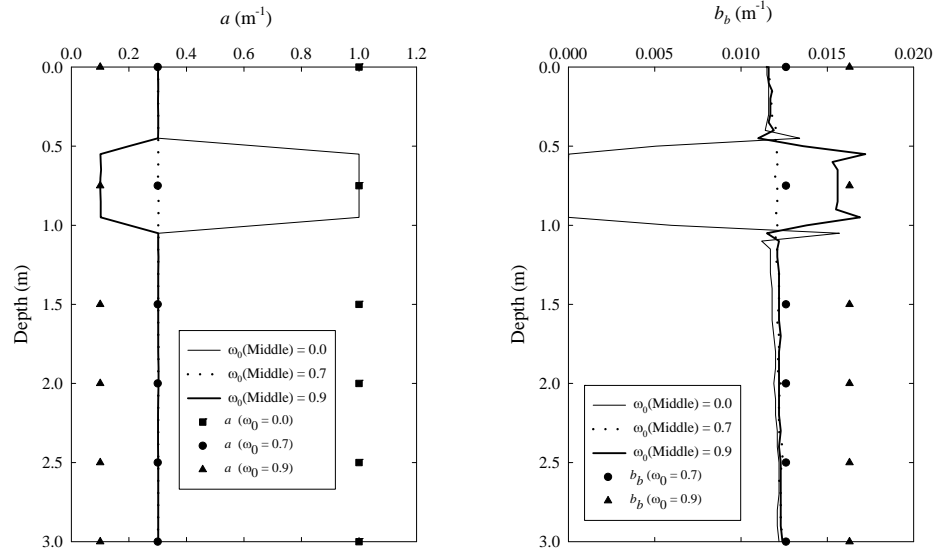


Figure 6.63: Retrieved profiles of a and b_b from a three-layer medium for the cases provided in Figure 6.62. The lines are the retrievals for the stratified media, and the symbols are the values of a and b_b for homogeneous media with the IOPs indicated in the caption. The retrievals should match the symbols in the appropriate layer (as they do in the left panel for a). In preparing the figure we have taken $c = 1$ m⁻¹.

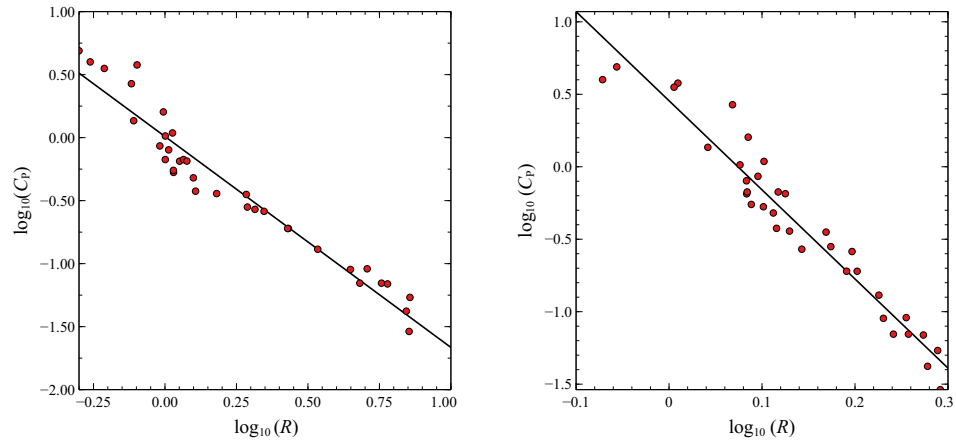


Figure 6.64: Bio-optical algorithms for retrieving the pigment concentration from CZCS-retrieved remote sensing reflectances in Case 1 waters. Left panel: $R = R_{rs}(443)/R_{rs}(550)$. Right panel: $R = R_{rs}(520)/R_{rs}(550)$. C_P is in mg/m³. The lines on both panels are the result of linear regression.

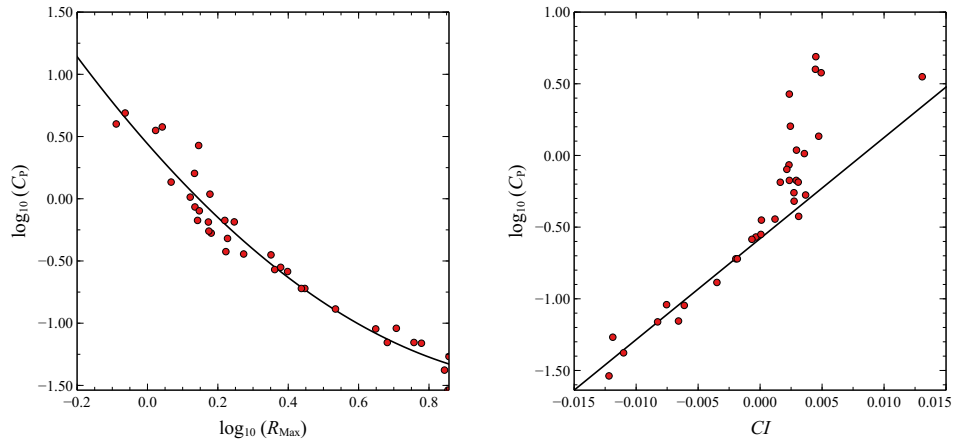


Figure 6.65: Newer bio-optical algorithms for retrieving the pigment concentration from SeaWiFS-like remote sensing reflectances in Case 1 waters. Left panel: R_{Max} is the maximum among $R_{rs}(443)/R_{rs}(550)$, $R_{rs}(490)/R_{rs}(550)$ and $R_{rs}(510)/R_{rs}(550)$. Right panel: relationship between C_P in mg/m^3 and the color index CI defined in the text. The line on the left panel is the least-squares fit to a quadratic, and the line on the right panel is a linear regression ($\log_{10}(C_P)$ vs. CI) of the data for $CI < 0$.

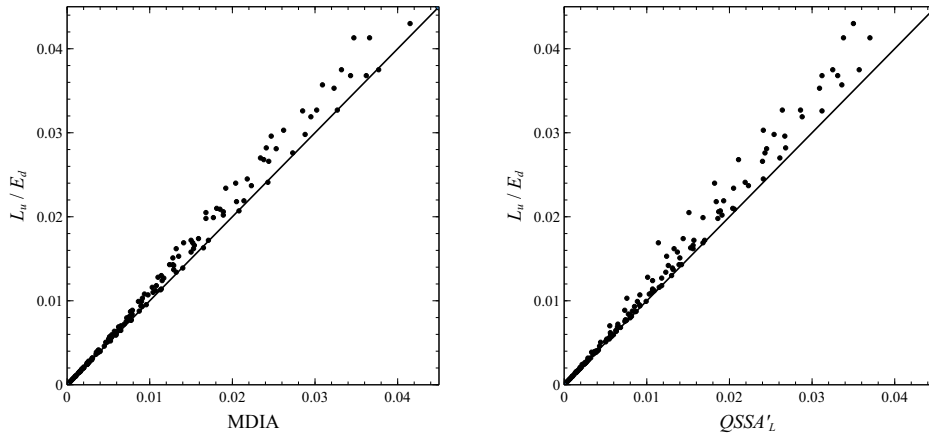


Figure 6.66: Comparison of exact computations of L_u/E_d with the modified δ -Isotropic approximation (MDIA) predictions (left panel) and the $QSSA'_L$ predictions (right panel) for $B_p = 0.0181$. There is no atmosphere and the refractive index of water is taken to be unity.

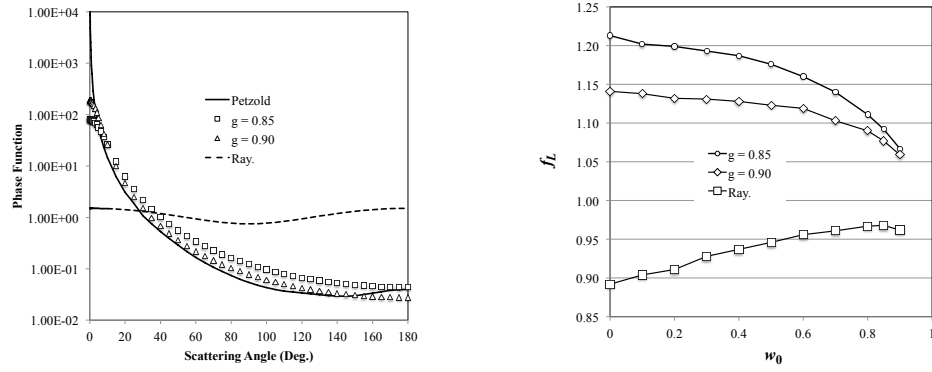


Figure 6.67: Zaneveld's shape factor f_L as a function of ω_0 (right panel) for a water body that scatters with a Henyey-Greenstein phase function for various values of the asymmetry parameter g (left panel). The Sun is at the zenith and there is no atmosphere.

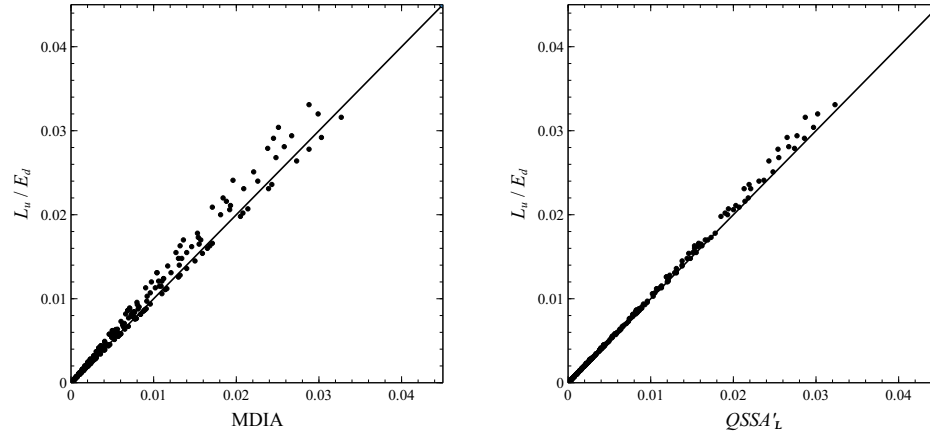


Figure 6.68: Comparison of exact computations of L_u/E_d with the modified δ -Isotropic approximation (MDIA) predictions (left panel) and the $QSSA'_L$ predictions (right panel) for $B_p = 0.002$. There is no atmosphere and the refractive index of water is taken to be unity.

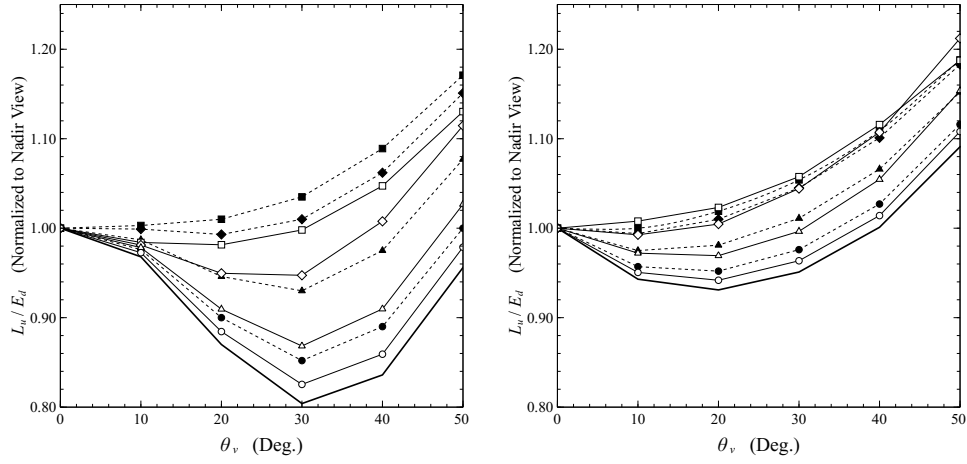


Figure 6.69: The normalized BRDF, $L_u(0, \mu)/E_d$, for the Petzold (left panel) and Petzmas (right panel) phase functions ($B_p = 0.0181$) and the Sun at the zenith. The view angle is given by $\theta_v = \arccos(\mu)$. The solid line at the bottom of the cluster is the QSSA result. The open symbols are for exact computations (circle is for $\omega_0 = 0.2$, triangle is for $\omega_0 = 0.5$, diamond is for $\omega_0 = 0.8$, and square is for $\omega_0 = 0.9$). Closed symbols are for Eq. (6.98), with the symbol shape having the same meaning as for the exact computations.

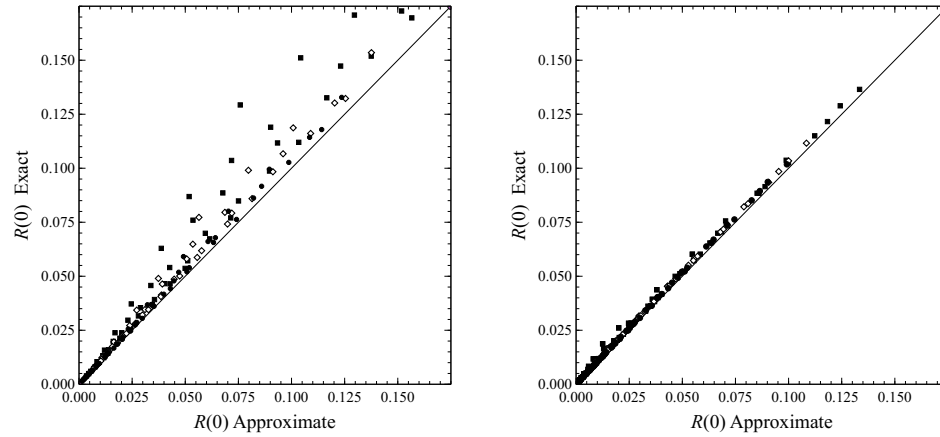


Figure 6.70: The comparison of the reflectance of a homogeneous medium computed by Monte Carlo methods (“Exact”) and the modified δ -Isotropic approximation (“Approximate”). The phase function is a mixture of the modified Petzold for the scattering by particles and Rayleigh for the scattering by water. The solid circles are for $\theta_0 = 10^\circ$, the open diamonds for $\theta_0 = 37^\circ$ and the solid squares for $\theta_0 = 60^\circ$. The values of the backscattering probability for the particles B_p is 0.0181 (left panel) and 0.002 (right panel). The refractive index of the water is taken to be unity. Similar plots comparing the exact and $QSSA'_R$ values of $R(0)$ are provided in Figure 6.25.

Chapter 7

Optical Effects of the Water Surface

7.1 Introduction

This chapter we consider aspects of the water surface that are important in remote sensing. Noting that the basic idea of ocean color remote sensing is to measure the radiance backscattered out of the water and propagated to satellite altitudes, and to relate that radiance to certain constituents in the water, we can enumerate some of the ways the water surface can be expected to affect this process. First, because of the change in refractive index at the interface, the surface will reflect a portion of the incident Sun light and sky light back into the atmosphere. Some of this reflected light may reach the sensor. Next, the surface will refract both of these components into the water, providing the photons available for backscattering toward the surface. Finally, the subsurface radiance that is backscattered will again interact with the surface, with some being reflected back into the water and the rest being transmitted through into the atmosphere. If there are breaking waves on the surface causing foam patches called “whitecaps,” these patches will reflect Sun light and sky light toward the sensor.

We begin by examining the optics of a flat water surface (most of which has already been described in Chapter 2). Next, we extend the analysis to a first-order analytical model of a wind-ruffled surface which accounts for most of the observed surface effects as long as the solar zenith angle and the viewing angles are not in excess of about 60° . Along the way we describe how this model was incorporated into the Monte Carlo radiative transfer simulation code that was used in Chapter 6 to study surface effects numerically. Finally,

we describe the salient properties of whitecaps.

7.2 Optical Effects of a Flat (Smooth) Water Surface.

The optics of a flat surface separating two media with different refractive indices has been described in considerable detail in Chapter 2. Here, we summarize the important points. Let radiance $L_i(\hat{\xi}_i)$ be incident on the interface from medium “ i ” (with refractive index m_i). Medium “ t ” has refractive index m_t (Figure 7.1). The radiance that is reflected back into medium “ i ” by the interface, $L_i(\hat{\xi}_r)$, is given by

$$L_i(\hat{\xi}_r) = r_f L_i(\hat{\xi}_i),$$

where $\hat{\xi}_r = \hat{\xi}_i - 2\hat{e}_z(\hat{e}_z \bullet \hat{\xi}_i)$ (or $\chi_r = \chi_i$) and r_f is the Fresnel reflectance of the interface.¹ The radiance that is transmitted through the interface $L_t(\hat{\xi}_t)$ is

$$L_t(\hat{\xi}_t) = \left(\frac{m_t}{m_i}\right)^2 t_f L_i(\hat{\xi}_i),$$

where $t_f = 1 - r_f$ is the Fresnel transmittance of the interface. The vector $\hat{\xi}_t$ is given by

$$m_2 \hat{\xi}_t = m_1 \hat{\xi}_i + (m_2 \cos \chi_t - m_1 \cos \chi_i) \hat{e}_z.$$

Note that the vectors $\hat{\xi}_i$, $\hat{\xi}_r$, $\hat{\xi}_t$ and \hat{e}_z are coplanar, and $m_i \sin \chi_i = m_t \sin \chi_t$.

We can use these relationships to answer the questions posed in the Introduction. First, what is the reflection and refraction of the direct Sun light? The radiance of the solar beam (Sun in a black sky) is $L(\hat{\xi}_i) = F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_i)$, so the reflected radiance is $L(\hat{\xi}_r) = r_f F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_r)$ and the transmitted radiance is

$$L(\hat{\xi}_t) = t_f \left(\frac{m_t}{m_i}\right)^2 F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_t),$$

where, for the air-water interface, $m_t = m_w$ and $m_i = 1$.² Similarly, the reflection and

¹The reader is reminded that when the incident radiance is unpolarized

$$r_f = \frac{1}{2} \left[\frac{\sin^2(\chi_i - \chi_t)}{\sin^2(\chi_i + \chi_t)} + \frac{\tan^2(\chi_i - \chi_t)}{\tan^2(\chi_i + \chi_t)} \right],$$

with $m_i \sin \chi_i = m_t \sin \chi_t$.

²The irradiance of the transmitted beam is easy to find. Noting that (Chapter 2) $m_t^2 \cos \chi_t d\Omega(\hat{\xi}_t) = m_i^2 \cos \chi_i d\Omega(\hat{\xi}_i)$, we find that $E_d(\hat{\xi}_t) = t_f E_d(\hat{\xi}_i)$, where the $E_d(\hat{\xi})$'s are, respectively, the irradiance incident on the water surface ($\hat{\xi} = \hat{\xi}_i$) and the irradiance transmitted through the water surface ($\hat{\xi} = \hat{\xi}_t$). Similarly, the reflected irradiance is $E_u(\hat{\xi}_r) = r_f E_d(\hat{\xi}_i)$ since $d\Omega(\hat{\xi}_r) = d\Omega(\hat{\xi}_i)$ and $\chi_r = \chi_i$.

refraction of sky radiance $L_{\text{Sky}}(\hat{\xi}_i)$ are, respectively, $L(\hat{\xi}_r) = r_f L_{\text{Sky}}(\hat{\xi}_i)$ and

$$L(\hat{\xi}_t) = t_f m_w^2 L_{\text{Sky}}(\hat{\xi}_i).$$

Finally, for radiance underneath the water surface propagating toward the surface, $L_u(\hat{\xi}_i)$, we can still use Figure 7.1, but now $m_i > m_t$ so $\chi_t > \chi_i$. Thus, the transmitted radiance is

$$L_t(\hat{\xi}_t) = \frac{t_f}{m_w^2} L_u(\hat{\xi}_i).$$

In this case we can also have $\chi_t = 90^\circ$, which occurs when $m_w \sin \chi_i = 1$. This value of χ_i is the critical angle χ_c . For $\chi_i > \chi_c$ there is no transmitted wave: all of the radiance in these angles is reflected with $\chi_r = \chi_i$, i.e., $r_f(\chi_i > \chi_c) = 1$.

7.3 Optical Effects of a Wind-ruffled Water Surface.

We now turn to the optics of a water surface ruffled by the wind. As in the case of a flat surface we will describe the computation of the various quantities of interest in remote sensing: reflection of Sun and sky light by the surface and transmission of upwelling subsurface radiance $L_u(\hat{\xi}_i)$ through the interface. We will also describe how the presence of a rough interface has been included in the radiative transfer process for use in other chapters.

7.3.1 Introduction

Consider standing near the water surface facing in the general direction of the Sun. If the water surface is flat, you will see a reflected image of the Sun in a small range of viewing angles determined by the solar angular diameter. However, if the water surface is ruffled by the wind you will see many individual images of the Sun (that flash on and off) in directions well outside the range of directions for which the image is observed with a flat surface. An individual flash in a given direction occurs when a portion of the moving surface has the correct orientation to reflect the solar beam toward the observer. The resulting visual pattern of flashes is usually called the “glitter pattern” or “Sun glint.” Seen from a great distance, this glitter pattern is a bright patch of individual flashes (usually) centered on the position the Sun’s image would occupy in the case of a flat surface. Standing close to the water surface, e.g., on the deck of a ship, looking toward the horizon during the rising or setting Sun, one would see something that looks schematically like the drawing in Figure 7.2, with the triangle in the drawing representing the average reflected image of the Sun, i.e., a photograph of the water surface and horizon taken with a long exposure

time. An actual photograph of such a pattern (but not a time exposure) is provided in Figure 7.3. This pattern has been referred to as “the road to happiness” in the Russian literature. In this section we examine the optics of a rough reflecting and refracting water surface. The as mentioned above, the goal will be to describe the effect of such a surface on the radiance it reflects and refracts, and along the way, explain some of the features in Figures 7.2 and 7.3.

7.3.2 Reflection of Light by a Wind-Ruffled Surface

There is a rigorous methods of treating reflection and refraction from a rough surface. This involves using the measured energy spectrum of surface waves (for various wind conditions), from which one can develop a realization of the water surface by synthesizing its shape, assuming that the phase of each component wave in the spectrum is a random angle. Then, tracing individual photons with the given (incident) radiance distribution through their (multiple) interactions with this surface determines the angular distribution of the reflected and refracted components. One then synthesizes a new realization the surface and repeats the process. Averaging over a large number of such realizations one can arrive at the radiances reflected from, and transmitted through, a randomly rough surface given an initial incident radiance distribution $L_i(\hat{\xi}_i)$.³ Such a procedure, although accurate, is computationally intensive, totally numerical and does not lend itself well to pedagogy, nor does it follow the historical development of the subject. Here we follow the traditional development that, with some limitations, is an excellent approximation, and nearly as accurate as the numerical procedure over a wide range of the important parameters.

Consider the water surface to be composed of a collection of flat facets whose surface normals follow an, as yet unknown, probability distribution. The reflective and refractive properties of a patch of surface containing a large number of such facets will allow estimation of this probability distribution. However, in order to determine this distribution, understanding the geometry of the reflection and refraction process is paramount. We will start with a description of reflection from an individual facet (Figure 7.4). Let the normal to the facet be denoted by \hat{n}_f , with the convention that it is directed *into* the water. The direction of the incident photon is denoted by $\hat{\xi}_i$ and that of the reflected photon by $\hat{\xi}_r$. The law of reflection (Chapter 1) states that

$$\hat{\xi}_r = \hat{\xi}_i - 2\hat{n}_f(\hat{n}_f \bullet \hat{\xi}_i). \quad (7.1)$$

Notice that this equation requires $\hat{\xi}_r$, $\hat{\xi}_i$ and \hat{n}_f to all lie in a plane. Let the nominal level sea surface be located at $z = z_1$. The instantaneous height of sea surface is $z - z_1 = \zeta(x, y, t)$

³What is “random” about the “randomly rough surface?” It’s the phases of the individual waves.

or $z_1 = z - \zeta(x, y, t) \triangleq f(x, y, t)$. The normal to this surface \hat{n}_f is given by

$$\hat{n}_f = \frac{\nabla f}{|\nabla f|} = \frac{\hat{e}_z - \hat{e}_x \frac{\partial \zeta}{\partial x} - \hat{e}_y \frac{\partial \zeta}{\partial y}}{\sqrt{1 + \left(\frac{\partial \zeta}{\partial x}\right)^2 + \left(\frac{\partial \zeta}{\partial y}\right)^2}}.$$

Letting

$$\vec{s} \triangleq \hat{e}_x \frac{\partial \zeta}{\partial x} + \hat{e}_y \frac{\partial \zeta}{\partial y} \triangleq \hat{e}_x s_x + \hat{e}_y s_y,$$

and $|\vec{s}| \triangleq s$, this can be written

$$\hat{n}_f = \frac{\hat{e}_z - \vec{s}}{\sqrt{1 + s^2}}. \quad (7.2)$$

Note the \hat{n}_f is a function of x , y , and t . The components of \vec{s} are the slopes of the lines of intersection of the facet with planes parallel to the x - z and y - z planes at x , y , and t . Now, take the scalar product of Eq. (7.1) with \hat{e}_z , noting that $\hat{n}_f \bullet \hat{e}_z = 1/\sqrt{1 + s^2}$, we find

$$\hat{n}_f \bullet \hat{\xi}_i = -\hat{e}_z \bullet (\hat{\xi}_r - \hat{\xi}_i) \frac{\sqrt{1 + s^2}}{2}.$$

Putting this and Eq. (7.2) into Eq. (7.1) we arrive at

$$\vec{s} = \hat{e}_z - \frac{\hat{\xi}_r - \hat{\xi}_i}{\hat{e}_z \bullet (\hat{\xi}_r - \hat{\xi}_i)}, \quad (7.3)$$

which is the fundamental relationship between the slope vector \vec{s} and the incident and reflected directions. It can be used in many ways: (1) given $\hat{\xi}_i$ and $\hat{\xi}_r$ it provides \vec{s} , which with Eq. (7.2) gives the surface normal \hat{n}_f required to effect the reflection from $\hat{\xi}_i$ to $\hat{\xi}_r$; (2) given $\hat{\xi}_i$ and \hat{n}_f , Eq. (7.2) gives \vec{s} and Eq. (7.3) gives $\hat{\xi}_r$, the direction of the reflected photon; and (3) given $\hat{\xi}_r$ and \hat{n}_f , Eq. (7.2) gives \vec{s} and Eq. (7.3) gives $\hat{\xi}_i$.

The angle of incidence χ of the light on the facet is needed to determine the facet's reflectivity from the Fresnel equations. It can be found in two ways. First, by virtue of the fact that $\hat{\xi}_i$, $\hat{\xi}_r$ and \hat{n}_f all lie in a plane: $\hat{\xi}_i \bullet \hat{\xi}_r = \cos(\pi - 2\chi) = -\cos 2\chi$. Alternatively, we note that in Eq. (7.1) $\hat{n}_f \bullet \hat{\xi}_i = \cos \chi$, so we have $|\hat{\xi}_r - \hat{\xi}_i| = 2 \cos \chi$. Thus given $\hat{\xi}_i$ and $\hat{\xi}_r$ we can find χ .

Going forward we will indicate the unit vectors by their polar (θ) and azimuthal (ϕ) angles: $\hat{\xi}_i = (\theta_i, \phi_i)$; $\hat{\xi}_r = (\theta_r, \phi_r)$; and $\hat{n}_f = (\theta_n, \phi_n)$, so

$$\begin{aligned} \hat{\xi}_i &= \hat{e}_x \cos \phi_i \sin \theta_i + \hat{e}_y \sin \phi_i \sin \theta_i + \hat{e}_z \cos \theta_i, \\ \hat{\xi}_r &= \hat{e}_x \cos \phi_r \sin \theta_r + \hat{e}_y \sin \phi_r \sin \theta_r + \hat{e}_z \cos \theta_r, \\ \hat{n}_f &= \hat{e}_x \cos \phi_n \sin \theta_n + \hat{e}_y \sin \phi_n \sin \theta_n + \hat{e}_z \cos \theta_n. \end{aligned} \quad (7.4)$$

We can now compute the radiance reflected in the direction $\hat{\xi}_r$ from a patch of the water surface containing many facets, as follows (refer to Figure 7.5). The power falling on the facet due to an incident radiance $L(\hat{\xi}_i)$ propagating in the direction $\hat{\xi}_i$ in a range of solid angles $\Delta\Omega(\hat{\xi}_i)$ is just $L(\hat{\xi}_i)a_j \cos \chi \Delta\Omega(\hat{\xi}_i)$, where a_j is the actual surface area of the facet. A fraction of the power r_f , where r_f is the Fresnel reflectivity of the facet for the given incident angle χ , is reflected in the direction $\hat{\xi}_r$ in a solid angle of the same magnitude as $\Delta\Omega(\hat{\xi}_i)$, i.e., $\Delta\Omega(\hat{\xi}_i) = \Delta\Omega(\hat{\xi}_r)$. The radiance from the surface in the direction $\hat{\xi}_r$ is (Chapter 2)

$$L(\hat{\xi}_r) = \frac{\text{Power toward } \hat{\xi}_r \text{ in } \Delta\Omega(\hat{\xi}_r)}{A_H |\cos \theta_r| \Delta\Omega(\hat{\xi}_r)},$$

where A_H is the (horizontal) surface area of the water viewed by the radiometer. For one facet, e.g., the j^{th} , the radiance is

$$L_j(\hat{\xi}_r) = \frac{r_f(\chi) L(\hat{\xi}_i) a_j \cos \chi \Delta\Omega(\hat{\xi}_i)}{A_H |\cos \theta_r| \Delta\Omega(\hat{\xi}_r)}. \quad (7.5)$$

Note that the two solid angles are the same and so they cancel out of the final result. We now write the facet area a_j in terms of its projection onto the horizontal water surface a_{H_j} , i.e., $a_j = a_{H_j} / \cos \theta_n$. The total radiance is the sum of the radiance from each facet, i.e., the sum over j , but the only quantity in the sum that depends on j is a_{H_j} . Thus, the total radiance is

$$L(\hat{\xi}_r) = \sum_j L_j(\hat{\xi}_r) = \sum_j \frac{r_f(\chi) L(\hat{\xi}_i) a_{H_j} \cos \chi}{A_H |\cos \theta_r| \cos \theta_n} = \frac{r_f(\chi) L(\hat{\xi}_i) a_H \cos \chi}{A_H |\cos \theta_r| \cos \theta_n},$$

where a_H is the total projected area of all facets in the area A_H oriented so as to reflect light from $\hat{\xi}_i$ to $\hat{\xi}_r$. The quantity a_H/A_H is the fraction of the surface with the correct orientation, that is, it is the probability that the slope \vec{s} or the normal \hat{n}_f has the correct orientation to reflect light from $\hat{\xi}_i$ to $\hat{\xi}_r$. We denote this as $P(\vec{s})$, the probability that facets within A_H have the required orientation. If we write the probability $P(\vec{s}) = p(s_x, s_y) ds_x ds_y$, where $p(s_x, s_y)$ is the probability density of \vec{s} , then

$$L(\hat{\xi}_r) = r_f(\chi) L(\hat{\xi}_i) \frac{\cos \chi}{|\cos \theta_r| \cos \theta_n} p(s_x, s_y) ds_x ds_y.$$

If the incident radiance comes from several directions, each of which require a different \vec{s} , we need to sum over these as well. Then finally,

$$L(\hat{\xi}_r) = \frac{1}{|\cos \theta_r|} \int \int r_f(\chi) L(\hat{\xi}_i) \frac{\cos \chi}{\cos \theta_n} p(s_x, s_y) ds_x ds_y. \quad (7.6)$$

To carry out the integration indicated in Eq. (7.6) we use the relationship between \vec{s} and $\hat{\xi}_i$ given $\hat{\xi}_r$: Eq. (7.3). In terms of the angles, Eq. (7.3) yields

$$\begin{aligned} s_x &= \frac{(\hat{\xi}_r - \hat{\xi}_i)_x}{\hat{e}_z \bullet (\hat{\xi}_r - \hat{\xi}_i)} = \frac{\cos \phi_r \sin \theta_r - \cos \phi_i \sin \theta_i}{\cos \theta_r - \cos \theta_i} \\ s_y &= \frac{(\hat{\xi}_r - \hat{\xi}_i)_y}{\hat{e}_z \bullet (\hat{\xi}_r - \hat{\xi}_i)} = \frac{\sin \phi_r \sin \theta_r - \sin \phi_i \sin \theta_i}{\cos \theta_r - \cos \theta_i}. \end{aligned} \quad (7.7)$$

In some cases it is more convenient to integrate over all incident directions $\hat{\xi}_i$ than over \vec{s} . To change the integration variables from s_x and s_y to θ_i and ϕ_i , we need to compute the Jacobian determinant of the transformation, Eqs. (7.7):

$$ds_x ds_y = |J(\vec{s}, \hat{\xi}_i)| d\theta_i d\phi_i,$$

where

$$J(\vec{s}, \hat{\xi}_i) = \begin{vmatrix} \partial s_x / \partial \theta_i & \partial s_x / \partial \phi_i \\ \partial s_y / \partial \theta_i & \partial s_y / \partial \phi_i \end{vmatrix} = \left(\frac{\partial s_x}{\partial \theta_i} \right) \left(\frac{\partial s_y}{\partial \phi_i} \right) - \left(\frac{\partial s_y}{\partial \theta_i} \right) \left(\frac{\partial s_x}{\partial \phi_i} \right).$$

From Eqs. (7.7) all of the indicated partial derivatives can be computed in a straightforward (but very tedious) manner. The final result is

$$|J| = \frac{[1 - \hat{\xi}_i \bullet \hat{\xi}_r]}{[\hat{e}_z \bullet (\hat{\xi}_i - \hat{\xi}_r)]^3} \sin \theta_i, \quad (7.8)$$

and noting that $\sin \theta_i d\theta_i d\phi_i = d\Omega(\hat{\xi}_i)$, we have

$$ds_x ds_y = \frac{[1 - \hat{\xi}_i \bullet \hat{\xi}_r]}{[\hat{e}_z \bullet (\hat{\xi}_i - \hat{\xi}_r)]^3} d\Omega(\hat{\xi}_i).$$

Using a trigonometric identity, $[1 - \hat{\xi}_i \bullet \hat{\xi}_r] = 1 + \cos 2\chi = 2 \cos^2 \chi$, and taking the scalar (dot) product of \hat{e}_z with Eq. (7.1) yields

$$\hat{e}_z \bullet (\hat{\xi}_i - \hat{\xi}_r) = 2(\hat{e}_z \bullet \hat{n}_f)(\hat{n}_f \bullet \hat{\xi}_i) = 2 \cos \theta_n \cos \chi, \quad (7.9)$$

so

$$ds_x ds_y = \frac{1}{4 \cos^3 \theta_n \cos \chi} d\Omega(\hat{\xi}_i).$$

Equation (7.6) can then be rewritten

$$L(\hat{\xi}_r) = \frac{1}{|\cos \theta_r|} \int \int r_f(\chi) L(\hat{\xi}_i) \frac{1}{4 \cos^4 \theta_n} p(\vec{s}) d\Omega(\hat{\xi}_i). \quad (7.10)$$

7.3.2.1 Reflected Direct Solar Radiance: Sun Glitter

If the water surface is illuminated solely by the solar beam (i.e., the Sun in a black sky) then the incident radiance is simply

$$L(\hat{\xi}_i) = F_0 \delta^2(\hat{\xi}_i - \hat{\xi}_0) = F_0 \delta(\cos \theta_i - \cos \theta_0) \delta(\phi_i - \phi_0)$$

where F_0 is the extraterrestrial solar irradiance and $\hat{\xi}_0$ is the propagation direction of the solar beam. Inserting this into Eq. (7.10) yields

$$L(\hat{\xi}_r) = \frac{r_f(\chi) F_0}{4 |\cos \theta_r| \cos^4 \theta_n} p(\vec{s}), \quad (7.11)$$

and suggests a method for finding $p(\vec{s})$.

7.3.2.2 The “Road to Happiness”

We now have the tools required to understand the origin of the triangular glitter pattern shown schematically in Figure 7.2. Consider an observer standing, say, on the deck of a ship looking at the water in the general direction of the setting Sun, as depicted in Figure 7.6. The observer, at a height H above the surface, is looking at a spot at F on the surface. Light is propagating in a direction $\hat{\xi}_r$ after being reflected at F from the direction $\hat{\xi}_i$. We assume that the photons propagating in the incident direction $\hat{\xi}_i$ are all from the Sun in a black sky, and take the azimuth of the solar beam to be $\phi_i = 0$. Thus, Eqs. (7.4) becomes

$$\begin{aligned} \hat{\xi}_i &= \hat{e}_x \sin \theta_i + \hat{e}_z \cos \theta_i, \\ \hat{\xi}_r &= \hat{e}_x \cos \phi_r \sin \theta_r + \hat{e}_y \sin \phi_r \sin \theta_r + \hat{e}_z \cos \theta_r, \end{aligned}$$

and Eq. (7.3) yields

$$\begin{aligned} s_x &= -\frac{(\cos \phi_r \sin \theta_r - \sin \theta_i)}{(\cos \theta_r - \cos \theta_i)} \\ s_x &= -\frac{\sin \phi_r \sin \theta_r}{(\cos \theta_r - \cos \theta_i)}. \end{aligned}$$

Noting that the equation of the tilt of the surface normal is $\cos \theta_n = (1 + s^2)^{-1/2}$, we have that $s = (s_x^2 + s_y^2)^{1/2} = \tan \theta_n$. Putting these together yields the angle ϕ_r :

$$\cos \phi_r = \frac{\sin^2 \theta_r + \sin^2 \theta_i - (\cos \theta_r - \cos \theta_i)^2 \tan^2 \theta_n}{2 \sin \theta_r \sin \theta_i}.$$

The interpretation of this equation is that for a given tilt of a facet on the water surface θ_n , ϕ_r is the azimuth of propagation of the reflected photon for it to be seen by the observer

given θ_r and θ_i .⁴ Now, the observer is looking down toward the surface (at F) in order to see this photon. The position on the surface from which the reflection takes place is easily seen to be

$$\begin{aligned}x &= -H \tan(\pi - \theta_r) \cos \phi_r \\y &= -H \tan(\pi - \theta_r) \sin \phi_r.\end{aligned}$$

Thus, for a given θ_r , θ_i , and θ_n the position of the spot on the water surface from which the solar beam is reflected has these coordinates (x, y) , i.e., the observer would see a bright spot at (x, y) , if the correct \hat{n}_f azimuth angle is available. Note the observer is at $(x, y, z) = (0, 0, -H)$.

To provide a concrete example of the use of these relationships, consider an observer standing at $H = 1$ m on the deck of ship. Let the solar zenith angle be 70° , i.e., $\theta_i = 70^\circ$, and assume that all azimuths of the normal to a surface facet are equally probable, i.e., all ϕ_n 's are equally probable, so whatever azimuth of the normal is required to reflect the light toward the observer, it is always available. Then, for given values of θ_n , the loci of points on the water surface as seen by the observer are given in Figure 7.7. Thus, for a fixed θ_n , the observer will see two approximately straight lines of bright spots progressing toward the horizon (one on the negative y -axis, as shown, and an identical line on the positive y -axis). The larger the surface tilt (θ_n) the larger the angle this straight line will make with the x axis. So, if only a single surface tilt (θ_n) were available, but all azimuth angles ϕ_n were equally probable, the observer would see a “V”-shaped line of flashes (as the correct ϕ_n 's occur) extending from the ship to the horizon. If all values of θ_n are possible up to some maximum value, then the interior of the “V”-shape will fill in and the resulting pattern of flashes will be similar to that shown schematically in Figure 7.2.

7.3.2.3 Estimation of $p(\vec{s})$ (Cox and Munk Analysis)

If we examine the glitter pattern for the Sun, i.e., measure $L_r(\hat{\xi}_r)$ for a given $\hat{\xi}_r$ and a fixed solar beam in the direction $\hat{\xi}_0$ with spherical coordinates (θ_0, ϕ_0) , we can find the, as yet, unknown probability density $p(\vec{s})$ in the following manner. First the variable χ is found from

$$\cos 2\chi = -\hat{\xi}_r \bullet \hat{\xi}_0 = -(\cos \phi_r \cos \phi_0 + \sin \phi_r \sin \phi_0) \sin \theta_r \sin \theta_0 + \cos \theta_r \cos \theta_0.$$

Next, θ_n is determined from Eq. (7.9), i.e.,

$$\cos \theta_n = \frac{\hat{e}_z \bullet (\hat{\xi}_0 - \hat{\xi}_r)}{2 \cos \chi} = \frac{\cos \theta_0 - \cos \theta_r}{2 \cos \chi}.$$

⁴Clearly, the azimuth angle of the surface normal ϕ_n must have the correct value for the reflection toward the observer to take place. If not, the observer sees no reflected light propagating in the $\hat{\xi}_r$ direction.

Finally, the slope components s_x and s_y can be found from Eq. (7.7):⁵

$$s_x = \frac{\cos \phi_r \sin \theta_r - \cos \phi_0 \sin \theta_0}{\cos \theta_r - \cos \theta_0}$$

$$s_y = \frac{\sin \phi_r \sin \theta_r - \sin \phi_0 \sin \theta_0}{\cos \theta_r - \cos \theta_0}$$

Note that the radiance in Eq. (7.11) is a function of time, i.e., in any given viewing direction the radiance will fluctuate as the facet at that point assumes various orientations, only a narrow range of which will produce a flash. However, if we use long (compared to the period of the capillary waves) integration times, we can measure the average radiance, and thus, $p(\vec{s})$. Alternatively, we can examine the water surface from a great distance away, so that a large number of facets are within the spatial resolution of the radiometer, e.g., a satellite-borne sensor viewing the water surface from space, and can obtain a similar average radiance and $p(\vec{s})$.⁶

Using out-of-focus photographs of the water surface from aircraft (equivalent to a spatial average) Cox and Munk estimated the probability density $p(s_x, s_y)$ to be approximately Gaussian, i.e.,

$$p(s_x, s_y) \approx \frac{1}{2\pi\tilde{\sigma}_x\tilde{\sigma}_y} \exp \left[-\frac{1}{2} \left(\frac{s_x}{\tilde{\sigma}_x} \right)^2 - \frac{1}{2} \left(\frac{s_y}{\tilde{\sigma}_y} \right)^2 \right], \quad (7.12)$$

where the slope variances $\tilde{\sigma}_x^2$ and $\tilde{\sigma}_y^2$ are related to the wind velocity \vec{W} .⁷ In Eq. (7.12), the x -axis is oriented so that the vector \hat{e}_x is parallel to the wind vector \vec{W} . With this

⁵Another useful set of relationships for the slope components can be found from Eq. (7.2) written in the form $\hat{n}_f = (\hat{e}_z - \vec{s})(\hat{n}_f \bullet \hat{e}_z) = (\hat{e}_z - \vec{s}) \cos \theta_n$. Then, $\hat{n}_f \bullet \hat{e}_x = \cos \phi_n \sin \theta_n = -s_x \cos \theta_n$ and $\hat{n}_f \bullet \hat{e}_y = \sin \phi_n \sin \theta_n = -s_y \cos \theta_n$, and

$$s_x = -\cos \phi_n \tan \theta_n$$

$$s_y = -\sin \phi_n \tan \theta_n.$$

The angle ϕ_n can be found from the law of reflection, $(\hat{n}_f = \hat{\xi}_0 - \hat{\xi}_r)/2 \cos \chi$, and Eqs. (7.4):

$$\cos \phi_n = \frac{\cos \phi_0 \sin \theta_0 - \cos \phi_r \sin \theta_r}{2 \cos \chi \sin \theta_n}$$

$$\sin \phi_n = \frac{\sin \phi_0 \sin \theta_0 - \sin \phi_r \sin \theta_r}{2 \cos \chi \sin \theta_n}.$$

⁶The so-called *ergodic theorem* of statistical physics demands that these two averages (time and spatial) are equivalent.

⁷Note that this is normalized in the sense that

$$\int_{-\infty}^{+\infty} ds_y \int_{-\infty}^{+\infty} p(s_x, s_y) ds_x = 1.$$

orientation, $\tilde{\sigma}_x$ is usually called the up-wind standard deviation $\tilde{\sigma}_u$ and $\tilde{\sigma}_y$ the cross-wind standard deviation $\tilde{\sigma}_c$. In using this equation, it is vitally important to remember that the various ϕ 's in Eq. (7.4) are *measured relative to the wind direction*. If we replace the surface slopes in Eq. (7.12) by the angles θ_n and ϕ_n , we have⁸

$$p(\theta_n, \phi_n) \approx \frac{1}{2\pi\tilde{\sigma}_u\tilde{\sigma}_c} \exp \left[- \left(\frac{\cos^2 \phi_n}{2\tilde{\sigma}_u^2} + \frac{\sin^2 \phi_n}{2\tilde{\sigma}_c^2} \right) \tan^2 \theta_n \right]. \quad (7.13)$$

In these equations, the empirical values of $\tilde{\sigma}_u$ and $\tilde{\sigma}_c$ were found to be

$$\begin{aligned} \tilde{\sigma}_c^2 &= 0.003 + 1.92 \times 10^{-3} W \\ \tilde{\sigma}_u^2 &= 0.000 + 3.16 \times 10^{-3} W, \end{aligned}$$

where W is in m/s.⁹

⁸Note here, that this probability density is still normalized; however, when we make the transformation from (s_x, s_y) to (θ_n, ϕ_n) we also need to transform from $ds_x ds_y$ to $d\theta_n d\phi_n$ which requires computation of the Jacobian of the transformation, i.e.,

$$ds_x ds_y = |J| d\theta_n d\phi_n,$$

where

$$J = \frac{\partial s_x}{\partial \theta_n} \frac{\partial s_y}{\partial \phi_n} - \frac{\partial s_y}{\partial \theta_n} \frac{\partial s_x}{\partial \phi_n}.$$

Using the relationships in Footnote 5 yields $J = \sec^2 \theta_n \tan \theta_n$, so the corresponding normalization integral is

$$\int_0^{2\pi} d\phi_n \int_0^{\pi/2} p(\theta_n, \phi_n) \sec^2 \theta_n \tan \theta_n d\theta_n = 1.$$

⁹There is considerable debate as to the validity of the values assigned to these parameters. [Shaw and Churnside \[1997\]](#) have directly measured $\tilde{\sigma}_u$ using a scanning-laser glint meter. Their results showed a strong dependence of $\tilde{\sigma}_u$ on the atmospheric stability. The atmospheric stability is characterized by the Richardson number Ri given by

$$Ri = g \frac{(T_a - T_w)}{T_w W^2},$$

where T_a and T_w are, respectively, the air and water temperatures ($^{\circ}\text{C}$), and g is the gravitational constant (9.8 m/s^2). The atmosphere is stable when $Ri > 0$ and unstable when $Ri < 0$. They combined their measurements with those of [Hwang and Shemdin \[1988\]](#) and developed the relationship between σ_u and Ri :

$$\begin{aligned} \frac{\tilde{\sigma}_u^2}{\tilde{\sigma}_{cm}^2} &= 1.42 - 2.8 Ri & \text{for} & \quad -0.23 < Ri < 0.27 \\ \frac{\tilde{\sigma}_u^2}{\tilde{\sigma}_{cm}^2} &= 0.65 & \text{for} & \quad Ri > 0.27. \end{aligned}$$

[Cox and Munk \[1954\]](#) collected most of their data for positive stability, thus for unstable atmospheres, $\tilde{\sigma}_u$ is considerably larger than that suggested by their equations. It is expected that $\tilde{\sigma}_c$ behaves in a manner similar to $\tilde{\sigma}_u$ in respect to its dependence on stability. It is important to note that a larger $\tilde{\sigma}$ implies a more diffuse glitter pattern, i.e., it extends farther from the specular point (the point at which Sun light would be reflected from a *flat* surface toward the sensor), but with smaller radiance near the specular

To demonstrate the effect of the wind direction on the slope distribution, we compare $p(\theta_n, 0^\circ)$ with $p(\theta_n, 90^\circ)$:

$$\frac{p(\theta_n, 0^\circ)}{p(\theta_n, 90^\circ)} = \exp \left[- \left(\frac{1}{2\tilde{\sigma}_u^2} - \frac{1}{2\tilde{\sigma}_c^2} \right) \tan^2 \theta_n \right],$$

and introducing the measured $\tilde{\sigma}$'s we find for a wind speed of 5 m/s, the values of the above ratio for $\theta_n = 5^\circ, 10^\circ, 15^\circ$, and 20° are 1.06, 1.28, 1.78, and 2.90, respectively. Thus, the wind direction is relatively unimportant for facets with small tilt angles, but more important for large tilt angles. The effect of the wind direction decreases with increasing wind speed, e.g., for $\theta_n = 20^\circ$ and $W = 20$ m/s the ratio is 1.74 compared to 2.90 for $W = 5$ m/s.

In Figure 7.8 we provide an example of the radiance reflected from a wind-ruffled water surface as a function of θ_r and ϕ_r . The source is the Sun in a black sky with a zenith angle of 30° , and lying in the x - z plane. The azimuth of $\hat{\xi}_r$ is “phi” (in Deg.) measured in the x - y plane relative to \hat{e}_x . The wind vector \vec{W} is in the x -direction, i.e., has *only* an \hat{e}_x component, so the plane containing the Sun’s rays and the z -axis also contains \vec{W} . Thus, for “phi = 0”, the projection $\hat{\xi}_r$ in the x - y plane is parallel to \vec{W} , while for “phi = 90° ”, it is perpendicular to \vec{W} . The wind speed is 7.5 m/s. The angle $\pi - \theta_r$ is the angle that $\hat{\xi}_r$ makes with $-\hat{e}_z$, which points *out* of the water. Note that as ϕ_r increases from 0° the maximum of the Sun glitter moves from near the specular point (the point where the Sun’s image would be for a *smooth* surface; here at $\pi - \theta_r = 30^\circ$) to approximately $\pi - \theta_r = 0^\circ$.

Figure 7.9 provides a second example of radiance reflected from a wind-ruffled water surface. The source is again the Sun in a black sky with a zenith angle of 30° . The wind is parallel to \hat{e}_x . For the filled symbols the plane containing $\hat{\xi}_i$, $\hat{\xi}_r$, and the surface normal also contains \vec{W} , while for the open symbols, it is perpendicular to \vec{W} . The wind speed is 7.5 m/s. We see that when the plane of $\hat{\xi}_r$ and $\hat{\xi}_i$ is parallel to the wind vector, the reflected radiance is somewhat higher than when the plane is perpendicular to \vec{W} .

For demonstration purposes, or to simplify computations (as below) the dependence of $p(\theta_n, \phi_n)$ on ϕ_n is sometimes ignored. In this case we need to determine what to take for

point. In contrast to direct measurements of the surface slope statistics, Ebuchi and Kizu [2002] combined directly observed glitter patterns (from a geostationary satellite) with satellite estimates of the wind speed and direction from space-borne radar scatterometers. They assumed that the apparent radiance of the surface in the visible is proportional to $p(s_x, s_y)$, and on this basis, derived the surface slope parameters. Their resulting $\tilde{\sigma}_c$ agreed well with that measured by Cox and Munk [1954]; however, their $\tilde{\sigma}_u$ showed a considerably *weaker* dependence on W . Considering that most of their measurements were in the tropics, where the atmosphere is expected to be unstable, their conclusions are *opposite* to Shaw and Churnside [1997] and Hwang and Shemdin [1988]. Ebuchi and Kizu [2002] attribute this to the likelihood that the direct measurements were made under conditions in which the waves were *growing* with the wind, whereas in their measurements the waves were in *equilibrium* with the wind, and therefore, represent *average* conditions. In our opinion the question of the most realistic values for $\tilde{\sigma}_c$ and $\tilde{\sigma}_u$ remains open.

$\tilde{\sigma}$. One possibility, is to first write

$$p(\theta_n, \phi_n) \approx C \exp \left[-\frac{\tan^2 \theta_n}{2\tilde{\sigma}_{\phi_n}^2} \right],$$

where

$$\frac{1}{2\tilde{\sigma}_{\phi_n}^2} = \frac{\cos^2 \phi_n}{2\tilde{\sigma}_u^2} + \frac{\sin^2 \phi_n}{2\tilde{\sigma}_c^2},$$

and C is a new normalization constant. Then take $\phi_n = 45^\circ$ (i.e., midway between the upwind and crosswind directions) and use

$$\tilde{\sigma}_{45^\circ}^2 = \frac{2\tilde{\sigma}_c^2\tilde{\sigma}_u^2}{\tilde{\sigma}_c^2 + \tilde{\sigma}_u^2} \triangleq \tilde{\sigma}^2 \approx 2.39 \times 10^{-3}W$$

as the wave slope variance *independent of direction*. Often, as we do in Chapter 6, $\tilde{\sigma}^2$ is taken to be the the Cox-Munk value of $\tilde{\sigma}_u^2 + \tilde{\sigma}_c^2 = 0.003 + 5.12 \times 10^{-3}W \triangleq \tilde{\sigma}^2$. In the literature this is sometimes referred to as the wave slope variance for an *omnidirectional* wind. The normalization constant C is easily found by requiring (see Footnote 8)

$$\int_0^{2\pi} d\phi_n \int_0^\infty p(\theta_n, \phi_n) \sec^2 \theta_n \tan \theta_n d\theta_n = 1.$$

It is $2\pi\tilde{\sigma}^2$, so if we choose to ignore the wind direction, we take

$$p(\theta_n, \phi_n) = \frac{1}{2\pi\tilde{\sigma}^2} \exp \left[-\frac{\tan^2 \theta_n}{2\tilde{\sigma}^2} \right] \quad \text{or} \quad p(\vec{s}) = \frac{1}{2\pi\tilde{\sigma}^2} \exp \left[-\frac{s_x^2 + s_y^2}{2\tilde{\sigma}^2} \right]. \quad (7.14)$$

A word of caution: In their paper, Cox and Munk used the distribution in Eq. (7.13) to estimate $\tilde{\sigma}_u$ and $\tilde{\sigma}_c$; however, in an earlier section of their paper they used the distribution

$$p(\vec{s}) = \frac{1}{\pi\sigma^2} \exp \left[-\frac{s_x^2 + s_y^2}{\sigma^2} \right],$$

to assess the influence of skylight on their measurements. This has caused confusion, especially for this author. Thus, in this work, we use different symbols for the slope variance depending on which form of $p(\vec{s})$ is being used: $\tilde{\sigma}$ and σ . These are related by $2\tilde{\sigma}^2 = \sigma^2$. Thus, while $\tilde{\sigma}^2 = 0.003 + 5.12 \times 10^{-3}W$ for an omnidirectional wind, $\sigma^2 = 0.006 + 10.24 \times 10^{-3}W$. Except in the present chapter, in most of the examples in this work, σ is used rather than $\tilde{\sigma}$.

7.3.3 Refraction of Light by the Wind-Ruffled Water Surface

We can describe the refraction of light through the interface between two media by the vector form of Snell's law:

$$m_t \hat{\xi}_t = m_i \hat{\xi}_i + (m_t \cos \chi_t - m_i \cos \chi_i) \hat{n}_{i \rightarrow t}, \quad (7.15)$$

where m_i , χ_i , and $\hat{\xi}_i$ are, respectively, the refractive index, the angle of incidence, and the propagation direction of the incident ray, m_t , χ_t , and $\hat{\xi}_t$ respectively, the refractive index, the angle of incidence, and the propagation direction of the transmitted ray, and $\hat{n}_{i \rightarrow t}$ is the unit normal to the surface directed from the incident side toward the transmitted side of the interface. The angles χ_i and χ_t are the angles ($< 90^\circ$) between $\hat{\xi}_i$ and $\hat{n}_{i \rightarrow t}$ and between $\hat{\xi}_t$ and $\hat{n}_{i \rightarrow t}$, respectively. They are related through Snell's law: $m_i \sin \chi_i = m_t \sin \chi_t$. Equation (7.15) is the refraction equivalent of the law of reflection Eq. (7.1). We simplify the notation by writing Eq. (7.1) in the form

$$m_t \hat{\xi}_t - m_i \hat{\xi}_i = k(\chi_i) \hat{n}_{i \rightarrow t},$$

with $k(\chi_i) \triangleq (m_t \cos \chi_t - m_i \cos \chi_i)$.

We now consider air to water refraction by a facet at the surface (Figure 7.10). Here, $\hat{n}_{i \rightarrow t}$ is \hat{n}_f . Note that Eq. (7.2) still provides \hat{n}_f in terms of the slope vector \vec{s} , so

$$m_t \hat{\xi}_t - m_i \hat{\xi}_i = k(\chi_i) \frac{\hat{e}_z - \vec{s}}{\sqrt{1 + s^2}},$$

and taking the scalar product of this with \hat{e}_z , we have

$$\hat{e}_z \bullet (m_t \hat{\xi}_t - m_i \hat{\xi}_i) = \frac{k(\chi_i)}{\sqrt{1 + s^2}},$$

so,

$$\vec{s} = \hat{e}_z - \frac{[m_t \hat{\xi}_t - m_i \hat{\xi}_i]}{\hat{e}_z \bullet [m_t \hat{\xi}_t - m_i \hat{\xi}_i]} = \hat{e}_z - \frac{[(m_t/m_i) \hat{\xi}_t - \hat{\xi}_i]}{\hat{e}_z \bullet [(m_t/m_i) \hat{\xi}_t - \hat{\xi}_i]}. \quad (7.16)$$

This is the analog of Eq. (7.3) for refraction. Note that the only difference between this and Eq. (7.3) is replacement of $\hat{\xi}_r$ with $(m_t/m_i) \hat{\xi}_t$.

As with reflection, we will compute the radiance transmitted in the direction $\hat{\xi}_t$ from a patch of the water surface containing many facets. In complete analogy to Eq. (7.5), the radiance transmitted by the j^{th} facet is

$$L_j(\hat{\xi}_t) = \frac{t_f(\chi_i) L(\hat{\xi}_i) a_j \cos \chi_i}{A_H |\cos \theta_t|} \frac{\Delta \Omega(\hat{\xi}_i)}{\Delta \Omega(\hat{\xi}_t)}, \quad (7.17)$$

where in Eq. (7.5) we have replaced the Fresnel reflectance r_f by the Fresnel transmittance t_f in Eq. (7.17), as well as θ_r by θ_t , and $\hat{\xi}_r$ by $\hat{\xi}_t$. Now in Eq. (7.5) the two solid angles were equal, but here they are not. In fact, we showed in Chapter 2 that

$$m_i^2 \cos \chi_i d\Omega(\hat{\xi}_i) = m_t^2 \cos \chi_t d\Omega(\hat{\xi}_t),$$

so

$$L_j(\hat{\xi}_t) = \frac{m_t^2 t_f(\chi_i) L(\hat{\xi}_i) a_j \cos \chi_t}{m_i^2 A_H |\cos \theta_t|},$$

and writing $a_j = a_{H_j} / \cos \theta_n$, summing over j , and identifying $\sum_j a_{H_j} / A_H$ as $P(\vec{s}) = p(\vec{s}) ds_x ds_y$, the probability of the slope \vec{s} in the area A_H of the water surface, the final result is

$$L_j(\hat{\xi}_t) = \frac{m_t^2}{m_i^2} t_f(\chi_i) L(\hat{\xi}_i) \frac{\cos \chi_t}{\cos \theta_n \cos \theta_t} P(\vec{s}).$$

If the incident radiance comes from a range of directions, the total transmitted radiance is

$$L(\hat{\xi}_t) = \frac{1}{\cos \theta_t} \frac{m_t^2}{m_i^2} \int \int t_f(\chi_i) L(\hat{\xi}_i) \frac{\cos \chi_t}{\cos \theta_n} p(s_x, s_y) ds_x ds_y, \quad (7.18)$$

where for the air-water interface $m_t^2/m_i^2 = m_w^2$.¹⁰ Note that the essential difference between Eq. (7.18) and Eq. (7.6) is the replacement of $r_f(\chi) \cos \chi$ in the latter by $(m_t/m_i)^2 t_f(\chi_i) \cos \chi_t$ in the former (and θ_r by θ_t).

Finally as before, it is sometimes easier to integrate over $d\theta_i d\phi_i$ rather than $ds_x ds_y$. To do this we again need to compute a Jacobian $J(\vec{s}, \hat{\xi}_i)$, but this time the \vec{s} and $\hat{\xi}_i$ are related by Eq. (7.16) rather than (7.3). The final result can differ from Eqs. (7.8) only in the appearance of the various m 's. As in the case of Eq. (7.8) the direct calculation is tedious; however, after having carried out the detailed derivation of Eq. (7.8) it is easy to see where the various m 's need to be placed in the final result, which is

$$|J(\vec{s}, \hat{\xi}_i)| = \left| \frac{-m_i^3 + m_i[m_i \hat{\xi}_i \bullet m_t \hat{\xi}_t]}{\{\hat{e}_z \bullet [m_i \hat{\xi}_i - m_t \hat{\xi}_t]\}^3} \sin \theta_i \right| = \frac{m_i^2}{k(\chi_i)^2} \frac{\cos \chi_i}{\cos^3 \theta_n} \sin \theta_i.$$

Therefore,

$$L(\hat{\xi}_t) = \frac{m_t^2}{\cos \theta_t} \int \int t_f(\chi) L(\hat{\xi}_i) \frac{\cos \chi_t \cos \chi_i}{k^2 \cos^4 \theta_n} p(\vec{s}) d\Omega(\hat{\xi}_i). \quad (7.19)$$

¹⁰Note that for a flat surface, $p(s_x, s_y) = \delta(s_x) \delta(s_y)$, so in this case the integral gives

$$L(\hat{\xi}_t) = \frac{m_t^2}{m_i^2} t_f(\chi_i) L(\hat{\xi}_i) = m_w^2 t_f(\chi_i) L(\hat{\xi}_i),$$

where we have used the fact that $\theta_n = 0$ so $\chi_t = \theta_t$ and χ_i and χ_t are related by Snell's law. Thus, we have recovered the well-known m_w^2 law for transmitted radiance (Chapter 2).

or, if we let $m = m_t/m_i$, and $k' = k/m_i = (m_t/m_i) \cos \chi_t - \cos \chi_i = m \cos \chi_t - \cos \chi_i$,

$$L(\hat{\xi}_t) = \frac{m^2}{\cos \theta_t} \int \int t_f(\chi_i) L(\hat{\xi}_i) \frac{\cos \chi_t \cos \chi_i}{k'^2 \cos^4 \theta_n} p(\vec{s}) d\Omega(\hat{\xi}_i).$$

When the source of $L(\hat{\xi}_i)$ is the Sun in a black sky, i.e., $L(\hat{\xi}_i) = F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0)$, the resultant subsurface radiance is

$$L(\hat{\xi}_t) = \frac{m^2}{\cos \theta_t} t_f(\chi_i) \frac{\cos \chi_t \cos \chi_i}{k'^2 \cos^4 \theta_n} p(\vec{s}) F_0. \quad (7.20)$$

7.3.4 Requirements for Conservation of Energy Across the Water Surface

Let the water surface be illuminated by the Sun in a black sky. The reflected and transmitted irradiances from the Sun are given by

$$\begin{aligned} E_{\text{Reflected}} &\triangleq E_u^+(\hat{\xi}_i) = \int_{\hat{\xi}_r \bullet \hat{e}_z < 0} L(\hat{\xi}_r) |\cos \theta_r| d\Omega(\hat{\xi}_r), \\ E_{\text{Transmitted}} &\triangleq E_d^-(\hat{\xi}_i) = \int_{\hat{\xi}_t \bullet \hat{e}_z > 0} L(\hat{\xi}_t) \cos \theta_t d\Omega(\hat{\xi}_t), \end{aligned} \quad (7.21)$$

where the subscripts u and d mean “up” and “down” while the superscripts $+$ and $-$ mean above and below the interface, respectively and the argument $\hat{\xi}_i$ on the irradiances is a reminder that the Sun light is incident in that direction. The reflected and transmitted radiances, $L(\hat{\xi}_r)$ and $L(\hat{\xi}_t)$ are provided by Eqs. (7.11) and (7.20). Using these,

$$\begin{aligned} E_u^+(\hat{\xi}_i) &= F_0 \int_{\hat{\xi}_r \bullet \hat{e}_z < 0} \frac{r_f(\chi)}{\cos^4 \theta_n} p(\vec{s}) d\Omega(\hat{\xi}_r) \\ E_d^-(\hat{\xi}_i) &= F_0 \int_{\hat{\xi}_t \bullet \hat{e}_z > 0} m_w^2 t_f(\chi_i) \frac{\cos \chi_t \cos \chi_i}{k'^2 \cos^4 \theta_n} p(\vec{s}) d\Omega(\hat{\xi}_t). \end{aligned} \quad (7.22)$$

In these equations it turns to be more convenient to integrate over s_x and s_y rather than the solid angles. This requires computing the Jacobians $J(\vec{s}, \hat{\xi}_r)$ and $J(\vec{s}, \hat{\xi}_t)$. These are computed in the same manner as before,¹¹ and the results are

$$|J(\vec{s}, \hat{\xi}_r)| = \frac{1}{4 \cos^3 \theta_n \cos \chi_i} \sin \theta_r$$

¹¹From experience computing the earlier Jacobians, it should be clear that for the transmitted case, the difference between $J(\vec{s}, \hat{\xi}_i)$ and $J(\vec{s}, \hat{\xi}_t)$ is the simple replacement of m_i with m_t , and possibly an irrelevant sign change.

and

$$|J(\vec{s}, \hat{\xi}_t)| = \frac{m_w^2 \cos \chi_t}{k'^2 \cos^3 \theta_n} \sin \theta_t.$$

Therefore

$$ds_x ds_y = \frac{1}{4 \cos^3 \theta_n \cos \chi_i} d\Omega(\hat{\xi}_r)$$

and

$$ds_x ds_y = \frac{m_w^2 \cos \chi_t}{k'^2 \cos^3 \theta_n} d\Omega(\hat{\xi}_t).$$

Finally,

$$\begin{aligned} E_u^+(\hat{\xi}_i) &= F_0 \int_{\hat{\xi}_r \cdot \hat{e}_z < 0} \frac{r_f(\chi_i) \cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y \\ E_d^-(\hat{\xi}_i) &= F_0 \int_{\hat{\xi}_t \cdot \hat{e}_z > 0} \frac{t_f(\chi_i) \cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y. \end{aligned} \quad (7.23)$$

Adding the integrals and recognizing that $r_f(\chi_i) + t_f(\chi_i) = 1$, we find

$$E_u^+(\hat{\xi}_i) + E_d^-(\hat{\xi}_i) = F_0 \int \frac{\cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y.$$

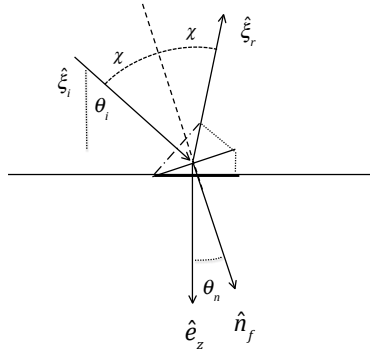
The left-hand-side of this equation is the irradiance (on a horizontal plane) propagating *away* from the interface, and recognizing that $F_0 \cos \theta_i$ is the irradiance propagating *toward* the interface, conservation of energy across the interface requires

$$E_u^+(\hat{\xi}_i) + E_d^-(\hat{\xi}_i) = F_0 \cos \theta_i$$

This is only possible if

$$\frac{1}{\cos \theta_i} \int \frac{\cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y = 1. \quad (7.24)$$

There is a simple geometric explanation for Eq. (7.24). In the diagram below, let



the longer horizontal line represent the area A_H and the heavy horizontal line a_{H_j} , the horizontal area associated with all of the facets oriented as shown, with a slope \vec{s}_j . Projected normal to $\hat{\xi}_i$ this facet has area $a_{H_j} \cos \chi_j / \cos \theta_{n_j}$, represented by the chain dashed line $(-\cdot-\cdot-\cdot-)$. Then similarly, the projected area of the surface on a plane normal to $\hat{\xi}_i$ is $A_H \cos \theta_n$. Noting that $a_{H_j}/A_H = P(\vec{s}_j)$, and that $A_H \cos \theta_n$ must equal the sum of all of the facets sloped so as to receive light from $\hat{\xi}_i$, i.e., $\sum_j a_{H_j} \cos \chi_j / \cos \theta_{n_j}$, we have

$$A_H \cos \theta_i = \sum_j a_{H_j} \cos \chi_j / \cos \theta_{n_j} \quad \text{or} \quad \cos \theta_i = \sum_j P(\vec{s}_j) \cos \chi_j / \cos \theta_{n_j}.$$

Thus, Eq. (7.24) is simply a manifestation of the fact that solar photons propagating toward A_H must strike *some* facet.

Is Eq. (7.24) satisfied when $p(s)$ is given by the Cox-Munk distribution, i.e., Eq. (7.12)? Let's calculate the integral and find out. For simplicity, we will use the omnidirectional form of $p(\vec{s})$, Eq. (7.14). But first, we have been a little sloppy in not putting limits on the integration over s_x and s_y . In reality, some facets are oriented in such a manner that $\chi_i > 90^\circ$, and the cannot contribute to the radiances and should be left out of the integrations. To properly set the limits, we note that, since we are using the omnidirectional form of $p(\vec{s})$, ϕ_n is not relevant, and we can orient the $(x-y)$ coordinate system in any manner we desire, i.e., take the most convenient value of ϕ_n . So let's take \hat{n}_f to be in the $x-z$ plane, i.e., $\phi_n = 0$. In this case, $\hat{\xi}_i$ and $\hat{\xi}_r$ are also in the $x-z$ plane as shown in Figure 7.11. Now, in order for light to be reflected by a facet, $\chi < 90^\circ$ or $\cos \chi > 0$. In the figure it is clear that $\cos \chi_i = \cos(\theta_n + \theta_i)$, which leads to

$$\frac{\cos \chi_i}{\cos \theta_n} = \frac{\cos(\theta_n + \theta_i)}{\cos \theta_n} = \cos \theta_i - \tan \theta_n \sin \theta_i,$$

but if $\phi_n = 0$, $\tan \theta_n = -s_x$, so

$$\frac{\cos \chi_i}{\cos \theta_n} = \cos \theta_i + s_x \sin \theta_i,$$

and the requirement that $\cos \chi_i > 0$ implies that $s_x > -\cot \theta_i$. Thus, in the integrals above, s_x goes from $-\cot \theta_i$ to $+\infty$ and s_y from $-\infty$ to $+\infty$, and the integral we need to compute, which we will call the inverse of $S(\theta_i)$, is

$$\frac{1}{S(\theta_i)} \triangleq \frac{1}{\cos \theta_i} \int \frac{\cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y = \int_{-\infty}^{+\infty} ds_y \int_{-\cot \theta_i}^{+\infty} (1 + s_x \tan \theta_i) p(\vec{s}) ds_x,$$

or inserting $p(\vec{s})$,

$$\begin{aligned} \frac{1}{S(\theta_i)} &= \frac{1}{2\pi\tilde{\sigma}^2} \int_{-\infty}^{+\infty} \exp \left[- \left(\frac{s_y^2}{2\tilde{\sigma}^2} \right) \right] ds_y \int_{-\cot \theta_i}^{+\infty} (1 + s_x \tan \theta_i) \exp \left[- \left(\frac{s_x^2}{2\tilde{\sigma}^2} \right) \right] ds_x \\ &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \int_{-\cot \theta_i}^{+\infty} (1 + s_x \tan \theta_i) \exp \left[- \left(\frac{s_x^2}{2\tilde{\sigma}^2} \right) \right] ds_x \end{aligned}$$

It is shown in Appendix 1 to this chapter that the value of this integral is

$$\frac{1}{S(\theta_i)} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\cot \theta_i}{\sqrt{2}\tilde{\sigma}} \right) + \frac{\tilde{\sigma} \tan \theta_i}{\sqrt{2\pi}} \exp \left(-\frac{\cot^2 \theta_i}{2\tilde{\sigma}^2} \right) \right],$$

where $\operatorname{erf}(x)$ is called the error function and is defined through

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp[-t^2] dt.$$

It is easy to see that

$$\frac{1}{S(\theta_i)} \leq 1$$

where the equality occurs only for $\theta_i = 0$. Thus, if we replace $p(\vec{s})$ by $p(\vec{s})S(\theta_i)$ Eqs. (7.21)-(7.23), the results satisfy energy conservation.¹² The quantity $S(\theta_i)$ is called the “shadowing factor” for light incident from $\hat{\xi}_i$.

It would appear that we have some inconsistency here, as the measurements made by Cox and Munk would have had to include $S(\theta_i)$, so the use Gaussian form of $p(\vec{s})$ may not be correct. However, most of their measurements were made for $\theta_i < 30^\circ$ and often $< 20^\circ$. For $\tilde{\sigma} = 0.2$ ($W \approx 7.5$ m/s) and $\theta_i = 30^\circ$, $\cot \theta_i / \sqrt{2}\tilde{\sigma} = 6.14$, and $S \approx 1$. (Actually $1 - \operatorname{erf}(6) \approx 2 \times 10^{-17}$ so S is *very* close to 1.) So we will assume henceforth that p is gaussian and that Eq. (7.14) is valid.¹³

7.3.5 Reflection and Refraction of Diffuse Radiance by the Water Surface

We now turn to the reflection and transmission of diffuse light, e.g., sky light and upwelling subsurface radiance. These can be computed using Eqs. (7.10) and (7.19) given the incident radiance, and replacing $p(\vec{s})$ by $p(\vec{s})S(\hat{\xi}_i)$. The integrals must in general be computed numerically; however, some simplification occurs when the incident radiance is totally diffuse, i.e., when $L(\hat{\xi}_i)$ is independent of $\hat{\xi}_i$. When the incident radiance is not totally diffuse, the reciprocity principle can be a useful guide to understanding the rough-surface effects. We start by applying the reciprocity relationship to the general case of propagation across the interface and then examine reflection and transmission.

¹²The difficulty that we face here, and which does not arise with a proper model of the water surface, is that we are assuming that the reflected and refracted radiance depends only on the surface slope at a given point, and is in no way dependent on the *elevation* variations of the surface. In reality, certain facets that are properly orientated so they could contribute to, say, reflected radiance are not illuminated by the source because they are in the shadows of other parts of the surface. The “fix” to conserve energy that we have constructed here should be considered to be a first-order approximation to account for the fact that the surface is being modeled in an overly simple manner.

¹³Measurements of the surface slope distribution by other means also yield a gaussian distribution similar to Eq. (7.12).

7.3.5.1 General Considerations: Reciprocity

The reciprocity principle was stated in Chapter 2, where we considered two radiative transfer problems on the same medium. The medium has a volume V bounded by a surface S . The (*outward*) normal to S is \hat{n} . Let any point on the boundary (surface) of the medium be indicated by the position vector $\vec{\rho}$. In Problem 1 the radiance incident on the boundary from the outside is $L_1(\vec{\rho}, \hat{\xi}, \lambda)$ and the internal sources are specified by $Q_1(\vec{r}, \hat{\xi}, \lambda)$. Likewise, in Problem 2 the radiance incident on the boundary from the outside is $L_2(\vec{\rho}, \hat{\xi}, \lambda)$. The inherent optical properties of the medium are identical in both problems. If there are no internal sources ($Q = 0$ in Chapter 2) then

$$\int_S dS \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| \left[\frac{L_1(\vec{\rho}, \hat{\xi}, \lambda) L_2(\vec{\rho}, -\hat{\xi}, \lambda)}{m^2(\vec{\rho})} - \frac{L_1(\vec{\rho}, -\hat{\xi}, \lambda) L_2(\vec{\rho}, \hat{\xi}, \lambda)}{m^2(\vec{\rho})} \right] d\Omega(\hat{\xi}) = 0. \quad (7.25)$$

In this equation, $L_1(\vec{\rho}, -\hat{\xi}, \lambda)$ and $L_2(\vec{\rho}, -\hat{\xi}, \lambda)$ are the radiances *exiting* the medium at the boundaries, and the $\hat{\xi} \bullet \hat{n} < 0$ on the solid angle integral on the left-hand-side indicates that the integration is restricted to directions *into* the medium.

To apply this we look at Figure 7.12, where the water surface contained within a volume bounded by two surfaces “A” for just “above” and “B” for just “below” the interface. Any point on the corresponding bounding surface is designated by the vector $\vec{\rho}_A$ and $\vec{\rho}_B$. The unit normals \hat{n}_A and \hat{n}_B are directed up and down, respectively, in the figure. The water surface itself can be flat or ruffled by the wind (the usual case). The boundary radiances (radiances *into* the volume) in case (a) on the left are $L_1(\vec{\rho}_A, \hat{\xi})$ in the form of a collimated beam, i.e., $L_1(\vec{\rho}_A) = F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0)$, for Problem 1 and a perfectly general $L_2(\vec{\rho}_A, \hat{\xi})$ for Problem 2. Both are independent the horizontal position on the surface. The radiances incident on boundary B are both zero. In this case, (suppressing the λ 's) the reciprocity principle states

$$\int_S dS \int_{\hat{\xi} \bullet \hat{n}_A < 0} |\hat{\xi} \bullet \hat{n}_A| \left[F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0) L_2(\vec{\rho}_A, -\hat{\xi}) - L_1(\vec{\rho}_A, -\hat{\xi}) L_2(\vec{\rho}_A, \hat{\xi}) \right] d\Omega(\hat{\xi}) = 0,$$

Because the radiances are independent of position on the boundary integral over S , just yields just the surface area of whatever sized volume we are examining. Therefore,

$$F_0 |\hat{\xi}_0 \bullet \hat{n}_A| L_2(\vec{\rho}_A, -\hat{\xi}_0) = \int_{\hat{\xi} \bullet \hat{n}_A < 0} |\hat{\xi} \bullet \hat{n}_A| L_1(\vec{\rho}_A, -\hat{\xi}) L_2(\vec{\rho}_A, \hat{\xi}) d\Omega(\hat{\xi}) \quad (7.26)$$

Note that $L_1(\vec{\rho}_A, -\hat{\xi}, \lambda)$ is the radiance reflected from the interface in the direction $-\hat{\xi}$ when collimated radiance is incident on the boundary in the direction $\hat{\xi}_0$, and $L_2(\vec{\rho}_A, -\hat{\xi}_0)$ is the radiance reflected from the interface in the direction $-\hat{\xi}_0$, when $L_2(\vec{\rho}_A, \hat{\xi})$ is incident on the surface.

It's illustrative to consider the case of a flat (smooth) interface. We know that the reflected radiance in Problem 1 is in the form of a collimated beam propagating in the direction $\hat{\xi}_r$, in fact $L_1(\rho_A, \hat{\xi}) = r_f(\chi_r)F_0\delta^{(2)}(\hat{\xi} - \hat{\xi}_r)$ (Figure 7.1). Then, $L_1(\rho_A, -\hat{\xi}) = r_f(\chi_r)F_0\delta^{(2)}(-\hat{\xi} - \hat{\xi}_r) = r_f(\chi_r)F_0\delta^{(2)}(\hat{\xi} + \hat{\xi}_r)$, and inserting this into Eq. (7.26) and executing the Dirac delta function yields

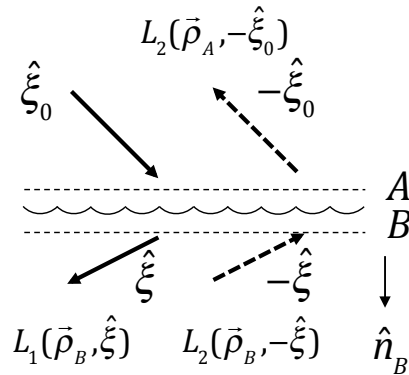
$$F_0|\hat{\xi}_0 \bullet \hat{n}_A|L_2(\vec{\rho}_A, -\hat{\xi}_0) = F_0|\hat{\xi}_r \bullet \hat{n}_A|r(\chi_r)L_2(\vec{\rho}_A, -\hat{\xi}_r),$$

The two dot products are equal, so $L_2(\vec{\rho}_A, -\hat{\xi}_0) = r(\chi_r)L_2(\vec{\rho}_A, -\hat{\xi}_r)$, i.e., radiance propagating in the direction $-\hat{\xi}_r$ is reflected into direction $-\hat{\xi}_0$ and reduced in magnitude by the Fresnel reflectance. This example provides a qualitative way of understanding what will happen when the surface is ruffled by the wind. In that case the radiance reflected in Problem 1, $L_1(\rho_A, -\hat{\xi})$, is no longer a Dirac delta function (it is given by Eq. (7.11)), but it is a sharply peaked function of direction around $-\hat{\xi}_r$ above, i.e., the “glitter pattern” similar to Figures 7.8 and 7.9. Then $L_2(\vec{\rho}_A, -\hat{\xi}_0)$ in Eq. (7.26) will receive radiance from a region of directions near $-\hat{\xi}_r$ rather than a single direction.

Now, consider Case (b) in Figure 7.12. Problem 1 is the same as above; however, for Problem 2, radiance is incident from below the interface, given by $L_2(\rho_B, \hat{\xi})$ with no radiance incident from above the interface. In this case,

$$\begin{aligned} F_0|\hat{\xi}_0 \bullet \hat{n}_A|L_2(\vec{\rho}_A, -\hat{\xi}_0) &= \frac{1}{m_w^2} \int_{\hat{\xi} \bullet \hat{n}_B < 0} |\hat{\xi} \bullet \hat{n}_B| L_1(\vec{\rho}_B, -\hat{\xi}) L_2(\vec{\rho}_B, \hat{\xi}) d\Omega(\hat{\xi}) \\ &= \frac{1}{m_w^2} \int_{\hat{\xi} \bullet \hat{n}_B > 0} |\hat{\xi} \bullet \hat{n}_B| L_1(\vec{\rho}_B, \hat{\xi}) L_2(\vec{\rho}_B, -\hat{\xi}) d\Omega(\hat{\xi}), \end{aligned} \quad (7.27)$$

where in the second line we have replaced $\hat{\xi}$ with $-\hat{\xi}$, which means the integration is now over directions $\hat{\xi}$ exiting the lower boundary. For clarity we provide a schematic in the figure below. Note that here $L_1(\vec{\rho}_B, \hat{\xi})$ is the radiance below the surface due to the collimated



beam propagating in the direction $\hat{\xi}_0$ (Problem 1) transmitted through the interface and propagating in the direction $\hat{\xi}$ (solid arrows), and the desired $L_2(\vec{\rho}_A, -\hat{\xi}_0)$ is the radiance resulting from $L_2(\vec{\rho}_B, -\hat{\xi})$ transmitted through the interface from below (dashed arrows). In the case of a flat interface, the solar beam is refracted from $\hat{\xi}_0$ above the surface to $\hat{\xi}_{0w}$ below the surface, yielding

$$L_1(\vec{\rho}_B, \hat{\xi}) = F_0 \frac{|\hat{\xi}_0 \bullet \hat{n}_A|}{|\hat{\xi}_{0w} \bullet \hat{n}_A|} t_f(\hat{\xi}_0 \rightarrow \hat{\xi}_{0w}) \delta^{(2)}(\hat{\xi} - \hat{\xi}_{0w}),$$

so Eq. (7.27) gives

$$F_0 |\hat{\xi}_0 \bullet \hat{n}_A| L_2(\vec{\rho}_A, -\hat{\xi}_0) = F_0 \frac{t_f(\hat{\xi}_0 \rightarrow \hat{\xi}_{0w})}{m_w^2} |\hat{\xi}_0 \bullet \hat{n}_A| L_2(\vec{\rho}_B, -\hat{\xi}_{0w}),$$

or

$$L_2(\vec{\rho}_A, -\hat{\xi}_0) = \frac{t_f(\hat{\xi}_0 \rightarrow \hat{\xi}_{0w})}{m_w^2} L_2(\vec{\rho}_B, -\hat{\xi}_{0w}),$$

the familiar formula radiance transmittance across the air-water interface (Chapter 2).

7.3.5.2 Reflection of Sky Radiance

The reflected sky radiance, can be computed using Eq. (7.10) with $p(\vec{s})$ replaced by $p(\vec{s})S(\hat{\xi}_i)$, or by utilizing Eq. (7.26). If we approximate the sky radiance by an isotropic radiance, i.e., $L_2(\vec{\rho}_A, \hat{\xi}) = L_0$, then Eq. (7.26) becomes

$$F_0 |\hat{\xi}_0 \bullet \hat{n}| \frac{L(\vec{\rho}_A, -\hat{\xi}_0)}{L_0} = \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| L_1(\vec{\rho}_A, -\hat{\xi}, \lambda) d\Omega(\hat{\xi}) = E_u^+(\hat{\xi}_0),$$

or

$$\frac{L(\vec{\rho}_A, -\hat{\xi}_0)}{L_0} \triangleq r_+(-\hat{\xi}_0) = \frac{E_u^+(\hat{\xi}_0)}{F_0 |\hat{n} \bullet \hat{\xi}_0|}. \quad (7.28)$$

In words: given a uniform sky radiance L_0 , the radiance reflected in the direction $-\hat{\xi}_0$ is the same as the irradiance reflected from the surface divided by the incident irradiance when the source is a collimated beam incident in the direction $\hat{\xi}_0$. This is valid for more realistic models of the rough surface, e.g., models that use the full spectrum of waves to synthesize random surfaces, and provides a convenient method for computing $r_+(-\hat{\xi}_0)$. As we shall see below, it is particularly useful when Monte Carlo techniques are employed to compute $r_+(-\hat{\xi}_0)$.¹⁴ Figure 7.13 provides $r_+(-\hat{\xi}_0)$ computed by Monte Carlo means to be described

¹⁴It is easy to show that Eq. (7.6) satisfies this result. When $L(\hat{\xi}_r)$ is a constant L_0 , this equation yields

$$\frac{L(\vec{\rho}_A, \hat{\xi}_r)}{L_0} = \frac{1}{|\cos \theta_r|} \int r_f(\chi_i) \frac{\cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y.$$

later in this chapter. In the figure, the “Viewing angle” is $\pi - \theta_r$, i.e., the polar angle of $\hat{\xi}_r$ measured with respect to $-\hat{e}_z$. Note the the reflectance closely follows the Fresnel reflectivity until viewing angles of about 60° (see Figure 7.14). This is manifest in the fact that for a given viewing angle, the radiance in the direction $\hat{\xi}_r$ results from reflection from directions close the vector $\hat{\xi}_i$ that would be required for a *flat* surface, i.e., $L_1(\vec{\rho}_A, -\hat{\xi}, \lambda)$ in Eq. (7.26) is a sharply peaked function around the direction $-\hat{\xi}_0$.

For larger angles the reflectivity is significantly smaller than Fresnel (Figure 7.14). This accounts in part for the existence of the horizon in directions away from the Sun. If the surface was smooth, as one looked at the water surface closer and closer to the horizon, there would be no variation in the magnitude of the radiance as the viewing angle varied from just less to just greater than 90° , i.e., there would be no visible horizon. If the surface is rough, Figure 7.13 shows that the radiance would be smaller viewing the water near 90° than the sky, i.e., the water would be darker than the sky.¹⁵

7.3.5.3 Transmittance of Upwelling Subsurface Radiance (L_u)

We now turn to the transmittance of upwelling subsurface radiance through the water-air interface. To do this we use the result for Case (b) in Figure 7.12, i.e., Eq. (7.27). If we assume the subsurface upwelling radiance is uniform, e.g., $L_2(\vec{\rho}_B, \hat{\xi}) = L_0$, then Eq. (7.27) reads

$$\frac{L_2(\vec{\rho}_A, -\hat{\xi}_0)}{L_0} = \frac{1}{m_w^2} \frac{\int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| L_1(\vec{\rho}_B, -\hat{\xi}) d\Omega(\hat{\xi})}{F_0 |\hat{\xi}_0 \bullet \hat{n}|},$$

In this integration $\hat{\xi}_r$ is fixed, and as \vec{s} varies it determines $\hat{\xi}_i$ through Eqs. (7.3). Now, compare this to Eq. (7.23):

$$E_u^+(\hat{\xi}_i) = F_0 \int_{\hat{\xi}_r \bullet \hat{e}_z < 0} \frac{r_f(\chi_i) \cos \chi_i}{\cos \theta_n} p(\vec{s}) ds_x ds_y.$$

In this integral, $\hat{\xi}_i$ is fixed and as \vec{s} varies it determines $\hat{\xi}_r$, so the same relationship between $\hat{\xi}_i$, $\hat{\xi}_r$ and \vec{s} is extant. If the value of the latter is $E_u^+(\hat{\xi}_i)/F_0$, the value of the former must be $E_u^+(\hat{\xi}_r)/F_0$. Thus,

$$\frac{L(\vec{\rho}_A, \hat{\xi}_r)}{L_0} = \frac{E_u^+(\hat{\xi}_r)}{F_0 |\cos \theta_r|}.$$

¹⁵ Another reason for the horizon is that the sky radiance is usually higher near the horizon than higher in the sky (limb brightening). Looking toward the horizon one sees facets that face the observer as well as those that are level. These facing facets reflect light from higher in the sky (smaller radiance) and with smaller incident angles (χ_i) (lower reflectivity). Both of these effects also account for the horizon. In addition, even for a smooth surface, the curvature of the Earth and limb brightening would combine to produce the visible horizon.

but the integral above is the downward irradiance just beneath the surface in Problem 1: $E_d^-(\hat{\xi}_0)$. Thus if we define the transmittance of the interface from below to be

$$t_-(-\hat{\xi}_0) \triangleq \frac{m_w^2 L_2(\vec{\rho}_A, -\hat{\xi}_0)}{L_0},$$

then

$$t_-(-\hat{\xi}_0) = \frac{E_d^-(\hat{\xi}_0)}{F_0|\hat{\xi}_0 \bullet \hat{n}|}. \quad (7.29)$$

Noting that $E_d^-(\hat{\xi}_0) + E_u^+(-\hat{\xi}_0) = F_0|\hat{\xi}_0 \bullet \hat{n}|$, we have the interesting result

$$t_-(-\hat{\xi}_0) + r_+(-\hat{\xi}_0) = 1 :$$

across the air-water interface, the sum of the reflectance of uniform radiance incident from above and the transmittance of uniform radiance incident from below is unity. We have already computed $r_+(-\hat{\xi}_0)$ as a function of W (Figure 7.13), so $t_-(-\hat{\xi}_0)$ follows immediately. For completeness this is provided in Figure 7.15. The transmittance factor for a viewing angle in air of 60° or less is within 0.5% of the Fresnel transmittance for flat (smooth) interface. We have seen in Chapter 6 that, although the angular distribution of upwelling radiance just beneath the water surface is not uniform, it does not deviate significantly from being uniform. Therefore, it is safe to assume that the transmittance of such radiance through the interface is given by the Fresnel transmittance for a flat (smooth) interface.

7.3.6 Incorporation of Wind-Ruffled Surface into Radiative Transfer

In this section, we discuss incorporating the wind-ruffled surface into radiative transfer codes used in remote sensing. We start by sketching how traditional codes, e.g., successive order of scattering, are modified by the rough surface. Then we discuss in a more detailed manner the inclusion of rough surface effects in Monte Carlo codes, as such a code is used to study the influence of surface roughness on the apparent optical properties (AOPs) of natural waters in Chapter 6.

7.3.6.1 General Methods

The addition of reflecting and transmitting boundaries to radiative transfer problems was described generally in Chapter 2. For a 1-d problem, i.e., a problem in which the properties of the medium and surface are independent on horizontal (x - y) position, if the radiance is

incident from a direction $\hat{\xi}_i$ the radiances reflected in the direction $\hat{\xi}_r$ and transmitted in the direction $\hat{\xi}_t$ are, respectively, given by

$$L(\hat{\xi}_r) = \int_{\text{All } \hat{\xi}_i} r(\hat{\xi}_i \rightarrow \hat{\xi}_r) L(\hat{\xi}_i) d\Omega(\hat{\xi}_i) \quad (7.30)$$

and

$$L(\hat{\xi}_t) = \int_{\text{All } \hat{\xi}_i} t(\hat{\xi}_i \rightarrow \hat{\xi}_t) L(\hat{\xi}_i) d\Omega(\hat{\xi}_i), \quad (7.31)$$

where r and t are the reflection and transmittance functions. Compare Eq. (7.30) to Eq. (7.10) with $p \rightarrow pS$:

$$L(\hat{\xi}_r) = \frac{1}{|\cos \theta_r|} \int \int r_f(\chi) L(\hat{\xi}_i) \frac{1}{\cos^4 \theta_n} p(\vec{s}) S(\hat{\xi}_i) d\Omega(\hat{\xi}_i).$$

This implies that for an atmosphere bounded by a wind-ruffled surface,

$$r(\hat{\xi}_i \rightarrow \hat{\xi}_t) = \frac{1}{|\cos \theta_r|} r_f(\chi) \frac{1}{\cos^4 \theta_n} p(\vec{s}) S(\hat{\xi}_i).$$

Similarly,

$$t(\hat{\xi}_i \rightarrow \hat{\xi}_t) = \frac{m_t^2}{\cos \theta_t} t_f(\chi) \frac{\cos \chi_t \cos \chi_i}{k^2 \cos^4 \theta_n} p(\vec{s}) S(\hat{\xi}_i).$$

Now, usually integrals such as this are discretized, e.g., as in Chapter 2, $L(z, \hat{\xi}) = L(z, \theta, \phi)$ might be Fourier-analyzed in ϕ and discretized using Gaussian quadratures in θ . Because of the form of these equations, it is clear that r and t are even functions of $\phi - \phi_w$, where ϕ_w is the azimuth of the wind and can be expanded in a Fourier cosine series in $\phi - \phi_w$. However, in the absence of the surface, the radiance can be expanded in a Fourier cosine series in $\phi - \phi_0$, where ϕ_0 is the solar azimuth. Thus, the addition of the surface to the atmosphere destroys the symmetry relative to the solar beam. However, if the wind is taken to be omnidirectional, then r and t are even functions of $\phi_i - \phi_r$ (just like the scattering phase function) and the symmetry is retained. The beauty of this symmetry is that when the radiance and reflectance are expanded in Fourier series, i.e.,

$$\begin{aligned} L(z, \theta_i, \phi_i) &= L^{(0)}(\theta_i) + 2 \sum_{\ell} L^{(\ell)}(\theta_i) \cos \ell(\phi_i - \phi_0) \\ r(\theta_i, \phi_i \rightarrow \theta_r, \phi_r) &= t^{(0)}(\theta_i \rightarrow \theta_r) + 2 \sum_{\ell} r^{(\ell)}(\theta_i \rightarrow \theta_r) \cos \ell(\phi_i - \phi_r), \end{aligned}$$

and inserted into the above equations, the Fourier orders completely decouple:

$$L^{(\ell)}(\theta_r) = 2\pi \int_0^{\pi/2} r^{(\ell)}(\theta_i \rightarrow \theta_r) L^{(\ell)}(\theta_i) \sin \theta_i d\theta_i,$$

with the result that the problem separates into individual problems for each Fourier coefficient $L^{(\ell)}(z, \theta)$. These individual problems can be solve by methods described in Chapter 2. Because of this separation, most treatments of rough surfaces (other than Monte Carlo) assume an omnidirectional wind. Further discussion of this problem would take us too far afield; however, a reference dealing with it in detail is given in the Bibliographic Notes.

7.3.6.2 Monte Carlo Methods

The evaluation of integrals using Monte Carlo methods is thoroughly discussed in the Mathematical Appendix (Chapter 14). There it is shown that one method for estimating the value of the multidimensional integral

$$I \triangleq \int_a^b \int_c^d \int_e^f \cdots f(x, y, z, \cdots) p(x, y, z, \cdots) dx dy dz \cdots$$

is to treat $p(x, y, z, \cdots)$ as a probability density and sample $x_i, y_i, z_i, \cdots, i = 1$, to N , from

$$\frac{p(x, y, z, \cdots)}{\int_a^b \int_c^d \int_e^f \cdots p(x, y, z, \cdots) dx dy dz \cdots}.$$

The estimator for I is then

$$I \approx \left[\frac{1}{N} \sum_{i=1}^N f(x_i, y_i, z_i, \cdots) \right] \int_a^b \int_c^d \int_e^f \cdots p(x, y, z, \cdots) dx dy dz \cdots \quad (7.32)$$

If it is too difficult to sample from p , then we can sample from $\tilde{p}(x, y, z, \cdots)$ and use as the estimator

$$I \approx \left[\frac{1}{N} \sum_{i=1}^N f(x_i, y_i, z_i, \cdots) w(x_i, y_i, z_i, \cdots) \right] \int_a^b \int_c^d \int_e^f \cdots \tilde{p}(x, y, z, \cdots) dx dy dz \cdots, \quad (7.33)$$

where

$$w(x_i, y_i, z_i, \cdots) = \frac{p(x_i, y_i, z_i, \cdots)}{\tilde{p}(x_i, y_i, z_i, \cdots)}.$$

We are now ready to consider a rough water surface. In the case of reflection from the surface in a given direction $\hat{\xi}_r$, recall Eq. (7.6):

$$L(\hat{\xi}_r) = \frac{1}{|\cos \theta_r|} \int \int r_f(\chi_i) L(\hat{\xi}_i) \frac{\cos \chi_i}{\cos \theta_n} S(\theta_i) p(s_x, s_y) ds_x ds_y,$$

where we have replace p by Sp to conserve energy. To estimate $L(\hat{\xi}_r)$ by Monte Carlo means, we sample \vec{s} from $p(\vec{s})$, which is already normalized, i.e., $\int p(s_x, s_y) ds_x ds_y = 1$. The determined \vec{s} provides $\hat{\xi}_i$, which allows computation of χ_i , θ_i and $S(\theta_i)$. For this particular sample, call it the j^{th} sample, the quantity corresponding to $f(x_j, y_j, z_j, \dots)$ in Eqs. (7.32) or (7.33) is then

$$\frac{1}{|\cos \theta_r|} \left[r_f(\chi_i) L(\hat{\xi}_i) \frac{\cos \chi_i}{\cos \theta_n} S(\theta_i) \right]_j.$$

Repeat this procedure N_j times and the Monte Carlo estimate is

$$L(\hat{\xi}_r) = \frac{1}{|\cos \theta_r|} \frac{1}{N_j} \sum_{j=1}^{N_j} r_f(\chi_{ij}) L(\hat{\xi}_{ij}) \frac{\cos \chi_{ij}}{\cos \theta_{nj}} S(\theta_{ij}), \quad (7.34)$$

where a second subscript j has been added to each variable to indicate that each sample of \vec{s} , i.e., \vec{s}_j , produces a different $\hat{\xi}_i$, now referred to as $\hat{\xi}_{ij}$ — the incident direction for the j^{th} sample, etc. This estimate of $L(\hat{\xi}_r)$ becomes better and better as N_j increases.¹⁶

Now a Monte Carlo radiative transfer code doesn't deal with radiance, it deals with individual photons, so how do we translate what we have done thus far into numbers of photons? Recall from Chapter 2 that the number of photons N_P associated with a radiance $L(\hat{\xi}_r)$ leaving an (horizontal) area of surface A_S in a solid angle $\Delta\Omega(\hat{\xi}_r)$ is

$$N_P(\hat{\xi}_r) = C_P L(\hat{\xi}_r) |\cos \theta_r| A_S \Delta\Omega(\hat{\xi}_r),$$

where C_P is a constant. The equation for $N_P(\hat{\xi}_i)$ is similar with the subscripts r replaced by i . Thus, if we partition the unit sphere into a set of solid angles, all of which are the same size $\Delta\Omega$, then Eq. (7.34) becomes

$$N_P(\hat{\xi}_r) = \frac{1}{N_j} \sum_{j=1}^{N_j} \frac{1}{\cos \theta_{ij}} r_f(\chi_{ij}) \frac{\cos \chi_{ij}}{\cos \theta_{nj}} S(\theta_{ij}) N_P(\hat{\xi}_{ij}) = \frac{1}{N_j} \sum_{j=1}^{N_j} w(\hat{\xi}_{ij} \rightarrow \hat{\xi}_r) N_P(\hat{\xi}_{ij}),$$

with

$$w(\hat{\xi}_{ij} \rightarrow \hat{\xi}_r) \triangleq \frac{1}{\cos \theta_{ij}} r_f(\chi_{ij}) \frac{\cos \chi_{ij}}{\cos \theta_{nj}} S(\theta_{ij}).$$

But in a radiative transfer code, we don't *a priori* know $N_P(\hat{\xi}_i)$ for all $\hat{\xi}_i$. That will be one of the results of the simulation. So, how do we incorporate these ideas into a radiative transfer code? As described earlier in Chapter 2, photons are followed through their various interactions, for each of which the probability of the various outcomes are known. Let one

¹⁶Typically, the percent error in the final result is proportional to $\sqrt{N_j}$, so for a given N_j , to decrease the error by a factor of 10 one must increase N_j by a factor of 100.

of these interactions be with the surface. The photon has a direction $\hat{\xi}_i$. We choose \vec{s} from $p(\vec{s})$ and this gives $\hat{\xi}_r$ and the various angles from which we compute $w(\hat{\xi}_i \rightarrow \hat{\xi}_r)$. We then propagate the photon in the direction $\hat{\xi}_r$ but with a “weight” of w . In the end we do not accumulate photon numbers in the various detectors, but we *accumulate photon weights*. This is similar to the way the effect of ω_0 was treated in Chapter 2. In the case of ω_0 , in a given interaction instead of destroying the photon if it was absorbed, the photon is retained but its weight is reduced by a factor of ω_0 , the probability that it survived the interaction, so after n interactions its weight is ω_0^n . Here, if the weight of the photon entering the surface interaction propagating in the direction $\hat{\xi}_i$ is W , it will be $w(\hat{\xi}_i \rightarrow \hat{\xi}_r) \times W$ after the interaction with the surface, propagating in the $\hat{\xi}_r$ direction. But wait, don’t photons ever penetrate the surface? Of course, the transmitted photons have a weight

$$w(\hat{\xi}_i \rightarrow \hat{\xi}_t) = \frac{1}{\cos \theta_i} t_f(\chi_i) \frac{\cos \chi_i}{\cos \theta_n} S(\theta_i).$$

Actually, rather than following two photons, in the codes used here, at the interaction with the surface $r_f(\chi_i)$ is used to determine whether the photon is reflected or transmitted: from a random number ρ with uniform distribution on $0 \rightarrow 1$, reflection occurs if $\rho < r_f$, and transmission otherwise. So actually, in either case the weight is modified by

$$w(\hat{\xi}_i \rightarrow \hat{\xi}_f) = \frac{\cos \chi_i}{\cos \theta_i \cos \theta_n} S(\theta_i),$$

where $\hat{\xi}_f$ is the final direction.

In our code, we do not choose \vec{s} directly from $p(\vec{s})$. Rather, we choose θ_n and ϕ_n from $p(\theta_n, \phi_n)$. Recall (Footnote 8),

$$p(\vec{s}) ds_x ds_y = p(\theta_n, \phi_n) \sec^2 \theta_n \tan \theta_n d\theta_n d\phi_n,$$

and since θ_n and ϕ_n are independent random variables, $p(\theta_n, \phi_n) = p_\theta(\theta_n)p_\phi(\phi_n)$. Assuming an omnidirectional wind, from Eq. (7.14)

$$p(\theta_n, \phi_n) = \frac{1}{2\pi\tilde{\sigma}^2} \exp \left[-\frac{\tan^2 \theta_n}{2\tilde{\sigma}^2} \right],$$

but since all ϕ_n ’s are equally probable, $p_\phi(\phi_n)$ is uniform, so we take

$$p_\theta(\theta_n) = \frac{1}{\tilde{\sigma}^2} \exp \left[-\frac{\tan^2 \theta_n}{2\tilde{\sigma}^2} \right],$$

$$p_\phi(\phi_n) = \frac{1}{2\pi}.$$

Then, given a uniformly distributed random number ρ_{ϕ_n} on the interval $0 \rightarrow 1$,

$$\rho_{\phi_n} = \int_0^{\phi_n} \frac{1}{2\pi} d\phi_n \implies \phi_n = 2\pi\rho_{\phi_n}.$$

Likewise, given a uniformly distributed random number ρ_{θ_n} on the interval $0 \rightarrow 1$,

$$\rho_{\theta_n} = \frac{1}{\tilde{\sigma}^2} \int_0^{\theta_n} \exp \left[-\frac{\tan^2 \theta_n}{2\tilde{\sigma}^2} \right] \sec^2 \theta_n \tan \theta_n d\theta_n = 1 - \exp \left[-\frac{\tan^2 \theta_n}{2\tilde{\sigma}^2} \right],$$

so

$$\tan \theta_n = \tilde{\sigma} \sqrt{-2 \ell n(1 - \rho_{\theta_n})}.$$

However, we are not quite finished. Although all values of $\tan \theta_n \geq 0$ are possible, not all values of ϕ_n are. For a facet to intercept the incident photon, we must have $\hat{n}_f \bullet \hat{\xi}_i > 0$, i.e., those values of ϕ_n for which $\hat{n}_f \bullet \hat{\xi}_i \leq 0$, are not possible. Thus, values of ϕ_n that *are possible* are given by

$$\begin{aligned} \hat{n}_f \bullet \hat{\xi}_i &= \cos \phi_i \cos \phi_n \sin \theta_i \sin \theta_n + \sin \phi_i \sin \phi_n \sin \theta_i \sin \theta_n + \cos \theta_i \cos \theta_n \\ &= \cos(\phi_n - \phi_i) \sin \theta_i \sin \theta_n + \cos \theta_i \cos \theta_n \\ &\geq 0 \end{aligned}$$

or

$$\cos(\phi_n - \phi_i) \geq -\cot \theta_i \cot \theta_n \quad \implies \quad \phi_n - \phi_i \leq \arccos(-\cot \theta_i \cot \theta_n) \triangleq \phi_n^0.$$

If this condition is satisfied, then the value obtained for ϕ_n is acceptable, otherwise a new value must be chosen.¹⁷ This of course means that we are not actually sampling from the true distribution of ϕ_n , i.e., $p_\phi = 1/2\pi$, but rather we are sampling from $\tilde{p}_\phi = 1/(2\pi - \phi_n^0)$. Thus, we must include an additional weight in the estimate: $w(\phi_n^0) = (2\pi - \phi_n^0)/2\pi$. If a photon goes into an interaction with the surface with a weight W_i , it exits the interaction with a weight

$$W_f = w(\phi_n^0) w(\hat{\xi}_i \rightarrow \hat{\xi}_f) W_i = \frac{(2\pi - \phi_n^0)}{2\pi} \frac{\cos \chi_i}{\cos \theta_i \cos \theta_n} S(\theta_i) W_i.$$

Note that it is possible for a photon going toward the water surface ($\hat{\xi}_i \bullet \hat{e}_z > 0$) to still be going downward after it is *reflected*, i.e., $\hat{\xi}_r \bullet \hat{e}_z > 0$. If this is the case, then let $\hat{\xi}_r$ become the new $\hat{\xi}_i$ and the photon interact with the surface a second time. In this manner the code will take into account some of the effects of multiple reflections, which were *not* addressed in the earlier sections.

The Monte Carlo code described above was that used to compute the results in Figure 7.13 by utilizing Eq. (7.28), removing the atmosphere, and rendering the water totally absorbing. In this manner the surface effects were isolated from the other radiative transfer processes. However, caution should be exercised: because the elevation of the water surface is not modeled, the shadowing of one surface element by another *is not considered at all*.

¹⁷When a sample is not acceptable and a new one is chosen, this process is known as “rejection sampling.”

This can lead to errors for large values of θ_i , especially when $\tilde{\sigma}$ is large as well. The results with this code are qualitatively comparable to those that treat the surface in a more realistic manner, and quantitatively comparable for $\theta_i \lesssim 60^\circ$.

Unfortunately, the correct way of treating the wind ruffled surface in a radiative transfer model, although straightforward in principle, is too computer intensive. One would need to develop many realizations of a randomly rough surface based on the energy spectrum of waves, use each of these realizations as the actual water surface for a Monte Carlo simulation, and then average the results of the individual Monte Carlo simulations.

7.3.7 The Normalized Water-Leaving Radiance

A simple application of influence of the surface occurs in the computation of the *normalized water-leaving radiance*. Let the upwelling radiance just beneath the water surface, propagating in the upward direction $\hat{\xi}'_v$, be $L_u(\hat{\xi}'_v, \hat{\xi}'_s)$, where solar photons are propagating in the direction $\hat{\xi}'_s$ *in the water*.¹⁸ If the surface is flat, the radiance transmitted through the interface – the *water-leaving radiance* – $L_w(\hat{\xi}_v, \hat{\xi}_s)$ is given by

$$L_w(\hat{\xi}_v, \hat{\xi}_s) = \frac{t_f(\hat{\xi}_v, \hat{\xi}'_v)}{m_w^2} L_u(\hat{\xi}'_v, \hat{\xi}'_s),$$

where $\hat{\xi}_v$ and $\hat{\xi}'_v$ are related by Snell's law and t_f is the Fresnel transmittance (water to air). An observer looking down at the water surface in the direction $-\hat{\xi}_v$ would measure $L_w(\hat{\xi}_v, \hat{\xi}_s)$. In Chapter 6 we studied L_u and its relationship to the IOPs in detail. There we showed that

$$L_u(\hat{\xi}'_v, \hat{\xi}'_s) = \frac{t_f(\hat{\xi}_s, \hat{\xi}'_s) E_d^+(\hat{\xi}_s)}{1 - rR(\hat{\xi}'_s)} \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right),$$

where E_d^+ is the downwelling irradiance incident the water surface, $\hat{\xi}_s$ the propagation direction of the solar beam (in air), $r \approx 0.48$ and R is the irradiance reflectance just beneath the surface. The vectors $\hat{\xi}_s$ and $\hat{\xi}'_s$ are also related by Snell's law. The combination L_u/E_d was the quantity directly related to the IOPs (also called R/Q). Therefore,

$$L_w(\hat{\xi}_v, \hat{\xi}_s) = \left[\frac{t_f(\hat{\xi}_v, \hat{\xi}'_v) t_f(\hat{\xi}_s, \hat{\xi}'_s)}{m_w^2 (1 - rR(\hat{\xi}'_s))} \right] E_d^+(\hat{\xi}_s) \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right).$$

The quantity in the square brackets is only weakly dependent on the optical properties of the water through the dependence of $R(\hat{\xi}'_s)$ on the IOPs. Typically $0 \leq R \leq 0.1$, so the

¹⁸Here we indicate the direction of the scattered radiance ($\hat{\xi}_v$) and the solar beam ($\hat{\xi}_s$) as arguments in L_w (and L_u) because later we will try to relate measurements of L_w made by observers looking in different directions with illumination at different solar zenith angles.

quantity $1 - rR(\hat{\xi}'_s)$ varies only from 0.952 to 1, and can be replaced by an average value with an error that is $\lesssim 3\%$. Thus, we assume that

$$\Re(\hat{\xi}_v, \hat{\xi}_s) \triangleq \frac{t_f(\hat{\xi}_v, \hat{\xi}'_v)t_f(\hat{\xi}_s, \hat{\xi}'_s)}{m_w^2(1 - rR(\hat{\xi}'_s))}$$

depends only on $\hat{\xi}_s$ and $\hat{\xi}_v$.¹⁹ Now, we multiply and divide by the extraterrestrial solar irradiance at the Sun-Earth distance equal to 1 Astronomical Unit (AU), \bar{F}_0 , to yield²⁰

$$L_w(\hat{\xi}_v, \hat{\xi}_s) = \Re(\hat{\xi}_s, \hat{\xi}_v) \bar{F}_0 \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right) \frac{E_d^+(\hat{\xi}_s)}{\bar{F}_0}.$$

This defines the normalized water-leaving radiance $[L_w]_N$ though

$$L_w(\hat{\xi}_v, \hat{\xi}_s) = [L_w(\hat{\xi}_s, \hat{\xi}_v)]_N \frac{E_d^+(\hat{\xi}_s)}{\bar{F}_0}, \quad (7.35)$$

so

$$[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N \triangleq \Re(\hat{\xi}_v, \hat{\xi}_s) \bar{F}_0 \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right) = \Re(\hat{\xi}_v, \hat{\xi}_s) \bar{F}_0 R_L(\hat{\xi}'_v, \hat{\xi}'_s),$$

where $R_L(\hat{\xi}'_v, \hat{\xi}'_s)$ is the remote sensing ratio (Chapter 6) generalized to viewing directions other than nadir. Thus, $[L_w(\hat{\xi}_s, \hat{\xi}_v)]_N$ consists of two parts, one of which (\Re) depends only on the air-water interface (ignoring the weak dependence on R).

How does this carry over to the case where the water surface is ruffled by the wind? We saw in Chapter 6 that L_u depends only weakly on $\hat{\xi}'_v$. So in the first approximation we will assume that L_u is independent of $\hat{\xi}'_v$, i.e., the upwelling radiance below the water surface is totally diffuse. Then the results of the Section 7.3.5.3 show that the transmittance of totally diffuse radiance through the interface is governed by $t_-(-\hat{\xi}_v)$. So one must replace $t_f(\hat{\xi}_v, \hat{\xi}'_v)$ by $t_-(-\hat{\xi}_v)$ in \Re to account for the surface roughness. However, from Figure 7.15 we see that for a viewing direction $\hat{\xi}_v$ such that $\pi - \theta_v \lesssim 60^\circ$, there is virtually no difference between $t_-(-\hat{\xi}_v)$ and $t_f(\hat{\xi}_v, \hat{\xi}'_v)$. What about the other transmittance in \Re ? This is the irradiance transmittance of the solar beam, but this is precisely what was computed as $t_-(-\hat{\xi}_0)$ in Eq. (7.29), and used to compute the first factor in \Re . Thus, for a roughened surface we can write

$$\Re(\hat{\xi}_v, \hat{\xi}_s) = \frac{t_-(-\hat{\xi}_v)t_-(-\hat{\xi}_s)}{m_w^2(1 - rR(\hat{\xi}'_s))}. \quad (7.36)$$

¹⁹It is easy to reduce the error in \Re due to R through an iterative process where one first takes a nominal R and computes $L_u/E_d = R/Q$. Then Q is estimated yielding a new R , etc. This can cut the error in \Re by at least a factor of 2 by using a nominal value for Q .

²⁰The extraterrestrial solar irradiance at the Sun-Earth distance equal to 1 AU, \bar{F}_0 , is about 0.07% less than the extraterrestrial solar irradiance averaged over a year and indicated by $\langle F_0 \rangle$ in Chapter 8. Also, it is about 0.03% less than the value of F_0 at the mean distance from the Earth to the Sun.

Again, $t_-(-\hat{\xi}_s) \approx t_f(\hat{\xi}_s, \hat{\xi}'_s)$ for $\theta_s \lesssim 60^\circ$. Thus, summarizing, for $\theta_s \lesssim 60^\circ$, \Re can be computed with high accuracy assuming that the surface is flat (smooth).

Why do we call $[L_w(\hat{\xi}_s, \hat{\xi}_v)]_N$ the *normalized* water-leaving radiance? Normalized in what way? We know that

$$E_d^+(\hat{\xi}_s) = \frac{\bar{F}_0 \cos \theta_s}{a_\oplus^2} t_E(\hat{\xi}_s),$$

where $t_E(\hat{\xi}_s)$ is the transmittance of irradiance from the top of the atmosphere to the water surface, and a_\oplus is the Earth-Sun distance in astronomical units (AU). Putting this into Eq. (7.35), we have

$$L_w(\hat{\xi}_v, \hat{\xi}_s) = [L_w(\hat{\xi}_v, \hat{\xi}_s)]_N \frac{1}{a_\oplus^2} \cos \theta_s t_E(\hat{\xi}_s),$$

or

$$[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N = L_w(\hat{\xi}_v, \hat{\xi}_s) \frac{a_\oplus^2}{\cos \theta_s t_E(\hat{\xi}_s)}. \quad (7.37)$$

Note that if the Sun is at the zenith ($\hat{\xi}_s = \hat{e}_z$) at $a_\oplus = 1$ AU, and in the absence of the atmosphere ($t_E(\hat{\xi}_s) = 1$), the actual water-leaving radiance propagating in toward the zenith ($\hat{\xi}_v = -\hat{e}_z$) is equal to the normalized water-leaving radiance, i.e., $L_w(-\hat{e}_z, \hat{e}_z) = [L_w(-\hat{e}_z, \hat{e}_z)]_N$.

We can compare the normalized water-leaving radiances for two different Sun-viewing geometries:

$$\frac{[L_w(\hat{\xi}_{v2}, \hat{\xi}_{s2})]_N}{[L_w(\hat{\xi}_{v1}, \hat{\xi}_{s1})]_N} = \frac{\Re(\hat{\xi}_{v2}, \hat{\xi}_{s2}) R_L(\hat{\xi}'_{v2}, \hat{\xi}'_{s2})}{\Re(\hat{\xi}_{v1}, \hat{\xi}_{s1}) R_L(\hat{\xi}'_{v1}, \hat{\xi}'_{s1})}. \quad (7.38)$$

The \Re ratio in this equation is easily calculated, while the R_L ratio can be modeled, e.g., from Chapter 6, we have to a good approximation $R_L = QSSA'_L$, so

$$\frac{R_L(\hat{\xi}'_{v2}, \hat{\xi}'_{s2})}{R_L(\hat{\xi}'_{v1}, \hat{\xi}'_{s1})} = \frac{QSSA'_L(\hat{\xi}'_{v2}, \hat{\xi}'_{s2})}{QSSA'_L(\hat{\xi}'_{v1}, \hat{\xi}'_{s1})}.$$

Ideally the remote sensor provides $L_w(\hat{\xi}_v, \hat{\xi}_s)$, which when given an estimate of $t_E(\hat{\xi}_s)$, provides $[L_w(\hat{\xi}_s, \hat{\xi}_v)]_N$. However, most field measurements of L_u are taken at a different Sun angle and are made at nadir, i.e., at $\hat{\xi}_v^{\text{Field}} = -\hat{e}_z$, and $\hat{\xi}'_s = \hat{\xi}_s^{\text{Field}}$. So, to compare field measurements of $[L_w]_N$ to remotely sensed measurements, modeling of the dependence of R_L on $\hat{\xi}_v$ and $\hat{\xi}_s$ is required. Such modeling is often referred to as the “BRDF effect.”

Later we will find it useful (especially in atmospheric correction, Chapter 10) to replace all radiances L by a reflectance ρ according to

$$\rho \triangleq \frac{\pi L}{F_0 \cos \theta_s},$$

where F_0 is the extraterrestrial solar irradiance *at the time* L is measured, i.e., the *instantaneous* F_0 (not to be confused with \bar{F}_0 , the value of F_0 at time mean Sun-Earth distance). How does ρ_w relate to $[L_w]_N$? Consider Eq. (7.37) and replace L_w by ρ_w . The result is

$$[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N = \rho_w(\hat{\xi}_v, \hat{\xi}_s) \frac{F_0 a_\oplus^2}{\pi t_E(\hat{\xi}_s)}.$$

Noting that $F_0 a_\oplus^2 = \bar{F}_0$, this becomes

$$[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N = \rho_w(\hat{\xi}_v, \hat{\xi}_s) \frac{\bar{F}_0}{\pi t_E(\hat{\xi}_s)},$$

and if we define the normalized *water-leaving reflectance*, $[\rho_w(\hat{\xi}_v, \hat{\xi}_s)]_N$, according to

$$\rho_w(\hat{\xi}_v, \hat{\xi}_s) = [\rho_w(\hat{\xi}_v, \hat{\xi}_s)]_N \pi t_E(\hat{\xi}_s), \quad (7.39)$$

we have

$$[\rho_w(\hat{\xi}_v, \hat{\xi}_s)]_N = \frac{\pi [L_w(\hat{\xi}_v, \hat{\xi}_s)]_N}{\bar{F}_0}. \quad (7.40)$$

It is important to note that $[\rho_w(\hat{\xi}_v, \hat{\xi}_s)]_N$ and $[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N$ are related by \bar{F}_0 , while $\rho_w(\hat{\xi}_v, \hat{\xi}_s)$ and $L_w(\hat{\xi}_v, \hat{\xi}_s)$ are related by the instantaneous F_0 . Recalling,

$$[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N = \Re(\hat{\xi}_v, \hat{\xi}_s) \bar{F}_0 \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right),$$

we see that

$$[\rho_w(\hat{\xi}_v, \hat{\xi}_s)]_N = \pi \Re(\hat{\xi}_v, \hat{\xi}_s) \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right). \quad (7.41)$$

While we are on this subject, recall that the *remote sensing reflectance*, R_{rs} was defined as

$$R_{rs}(\hat{\xi}_v, \hat{\xi}_s) \triangleq \frac{L_w(\hat{\xi}_v, \hat{\xi}_s)}{E_d^+}.$$

Using this it is easy to show that

$$R_{rs}(\hat{\xi}_v, \hat{\xi}_s) = \frac{[\rho_w(\hat{\xi}_v, \hat{\xi}_s)]_N}{\pi},$$

and so

$$R_{rs}(\hat{\xi}_v, \hat{\xi}_s) = \Re(\hat{\xi}_v, \hat{\xi}_s) \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right).$$

7.4 Whitecaps

When waves “break” on the water surface they inject air into the water to produce a rack of foam that is usually referred to as a *whitecap*.²¹ On a windy day one often sees a more or less uniform distribution of small white patches on the water surface. These patches are highly reflective and can supply a reflected radiance that directly competes with the water-leaving radiance we are interested in measuring. One can get a rough idea of the magnitude of the radiance produced by whitecaps by (1) measuring the fraction (f_{wc}) of the water surface covered by whitecaps, (2) estimating the reflectance (r_{wc}) of a single whitecap and (3) assuming the individual whitecaps are *lambertian* reflectors. With these quantities and assumptions, $L_{wc} = r_{wc} f_{wc} E_d^+ / \pi$, where E_d^+ is the irradiance incident on the water surface. For the Sun in a back sky, $E_d^+ = F_0 \cos \theta_0$, so

$$\frac{L_{wc}}{F_0} = \frac{r_{wc} f_{wc}}{\pi} \cos \theta_0.$$

It is obvious that the fraction of the sea surface covered with whitecaps is related to the wind speed W . Monahan and O’Muircheartaigh [1980] summarized the very noisy relationship between f and W as

$$f = 2.95 \times 10^{-6} W^{3.52},$$

where W is in m/s measured 10 meters above the water surface.²² It is also observed that there are essentially no whitecaps for $W \leq 6$ m/s. For the spectral range in the visible, the effective whitecap reflectance is almost independent of λ and has been determined by Koepke [1982] to be $\sim 22\%$. These included both fresh whitecaps as well as decaying whitecaps. Thus, an estimate of the whitecap radiance is

$$\frac{L_{wc}}{F_0} = \frac{2.95 \times 10^{-6} W^{3.52} \times 0.22}{\pi} \cos \theta_0 = 2.07 \times 10^{-7} W^{3.52} \cos \theta_0 \quad (7.42)$$

The water-leaving radiance near 500 nm $L_w \sim (2.5 \times 10^{-3}) F_0$, so this suggests that at $W \approx 13$ to 14 m/s the water-leaving radiance and the whitecap radiance are comparable in magnitude.

As whitecaps have the potential of adding upwelling radiance just above the water surface comparable magnitude to the desired L_w , it is important to try to obtain radiometric data

²¹The same process creates what is usually termed the *marine* or *maritime* aerosol.

²²Later Monahan and O’Muircheartaigh [1986] modified the expression to

$$f = 1.95 \times 10^{-5} W^{2.55} \exp[-0.0861 \Delta T],$$

which included the influence of the atmospheric stability. The quantity ΔT is the air-sea temperature difference ($\Delta T = T_A - T_W$), with $\Delta T > 0$ implying a stable atmosphere.

of actual oceanic whitecaps, and determine the effect of the wind speed. In particular, it is critical to understand the spectral dependence of whitecap reflectance. The author and co-workers approached this by constructing a ship-based radiometer for observing whitecaps while a ship is on station or underway. The radiometer, suspended from a boom off the bow of the ship, continuously viewed a spot about 12 cm in diameter on the sea surface, continuously measuring a radiance L_s . A video image, from a TV camera mounted along side of the radiometer to visually observe the water surface, was used to reject Sun glitter. A second radiometer on the deck of the ship recorded the incident irradiance E_d^+ . The radiance of the surface measured by the radiometer was recorded as a function of time (~ 7 samples/sec). This radiance consists of background radiance (L_b) from whitecap-free areas (the predominant situation) and a much higher radiance ($L_{wc} + L_b$) whenever a portion of a whitecap was in the field of view of the radiometer. After determining the radiance of the whitecap-free areas (L_b , essentially the “baseline” of the radiance, and in the absence of skylight and Sun glitter, L_w), and subtracting it from the entire record, we are left with the time-average radiance due to the whitecaps,

$$\langle L_{wc} \rangle = \langle L_s \rangle - \langle L_b \rangle.$$

The associated reflectance (the *remote-sensing augmented reflectance*, $RSAR$) is defined by

$$RSAR = \frac{\pi \langle L_{wc} \rangle}{E_d^+}. \quad (7.43)$$

The $RSAR$ is the irradiance reflectance of whitecaps (over and above the background) assuming that they are Lambertian reflectors.²³ The radiometer was accompanied by a meteorological package to provide the speed of the wind relative to the ship (and other, possibly relevant, parameters) and a GPS unit to provide the absolute speed of the ship. Combining these yielded W . The whitecap radiometer recorded in 10 nm bands centered at 6 wavelengths: 410, 510, 550, 670, 750, and 860 nm, and the downward surface irradiance was measured in 5 bands, also 10 nm wide, centered at 410, 510, 550, 670, and 860 nm.

An example of two whitecaps passing under the radiometer is shown in Figure 7.16. The 96 consecutive samples shown are acquired over a period of ~ 15 seconds. In this example a large whitecap suddenly breaks in view of the radiometer with thick white foam (sample point 11) reaching a peak reflectance of $\sim 55\%$. Six traces are plotted representing the six radiometer channels. The lower trace corresponds to the 860 nm reflectance. The thick foam is temporarily replaced by a region of submerged bubbles and less thick foam (\sim sample points 13, 14, 15) and some thick foam comes into view again at sample point 17. At sample point 20 and 21 a thin layer of foam passes followed by the decaying thicker foam to about sample point 35. Sample points from about 35 to 55 show the reflectance of

²³Rewriting Eq. (7.40) using Eq. (7.35) yields $[\rho_w]_N = \pi[L_w]_N/\bar{F}_0 = \pi L_w/E_d^+$. Comparison with Eq. (7.43) shows that $RSAR$, like $[\rho_w]_N$, is a similarly *normalized* reflectance.

thinning residual foam. From 60 to about 75 the reflectance of the foam free water surface is shown and is suddenly followed by another whitecap of smaller magnitude (sample point 76) and continues to decay out to about sample point 96. The data clearly suggest a significant fall in the near-infrared (NIR) reflectance of whitecaps.

From 1 to 13 November 1996, the whitecap radiometer was operated on a cruise from Manzanillo, Mexico to Honolulu, Hawaii. This cruise provided whitecap data under conditions of steady winds (the trades) of essentially unlimited duration and fetch. Unfortunately, analysis of the data was not as straightforward as expected. Under clear skies it is still very difficult to separate whitecaps from Sun glint events. Thus, the analysis was performed only under overcast conditions. Furthermore, the determination of the “baseline” reflectance is critical to the analysis and proved to be difficult as well.

The analysis for estimating $RSAR$ is described in detail in [Moore et al. \[2000\]](#). The dependence of $RSAR$ at 410 nm on wind speed is provided in Figure 7.17 along with the results of earlier investigations and parameterizations. The black triangles (joined by a vertical line) are the results of two different methods of data analysis (establishing the baseline). The larger (lower) black triangles are believed to be the better analysis of the data. For these,

$$RSAR \sim 3 \times 10^{-6} W^{2.55}.$$

There was no discernible spectral variation of $RSAR$ in the visible; however, the $RSAR$ was significantly lower at 860 nm than at 410 nm. (Figure 7.18). Although the data are very noisy, they suggest that

$$RSAR(860) \sim 0.85 \times RSAR(410),$$

for $RSAR(410) \lesssim 0.06$. This reduction of $RSAR$ in the NIR was similar to, but less than, that observed in the surface zone [[Frouin et al., 1996](#)] and in ship wakes [[Moore et al., 1998](#)]. Combining all of the observations, the $RSAR$ of whitecaps is estimated to be

$$RSAR(\lambda) = S(\lambda) \times 3 \times 10^{-6} W^{2.55}, \quad (7.44)$$

where $S(\lambda)$ is a spectral reflectance factor for whitecaps taken to be unity in the visible, 0.925 at 750 nm and 0.85 at 860 nm. While this formula fit our data reasonably well, it could well be an underestimate at higher wind speeds. The spread of the various measurements in Figure 7.17 suggests that the standard deviation in the $RSAR$ may be approximately equal to or somewhat greater than the $RSAR$ itself.

7.5 Numerical Comparisons of Surface Effects

We close out this chapter on surface optics by providing a quantitative example that compares the various effects of the water surface to the signal that we are attempting to measure (remotely): the normalized water-leaving radiance. Consider a sensor on a satellite in orbit moving in a south to north direction (with the sub-satellite track on the surface being a line of constant longitude) at the spring or fall equinox. In this case, the solar zenith angle (θ_i) is the latitude of the sensor, and the solar azimuth (ϕ_i) is zero. The sensor views the surface in one of two modes: (1) it is aimed in the nadir direction so $\theta_r = 180^\circ$; and (2) it is in a “tilted”²⁴ mode in which $\theta_r = 160^\circ$ and $\phi_r = 180^\circ$. Figure 7.19 compares the radiance that the sensor would view (in the absence of atmospheric attenuation) due to direct Sun glitter, reflected sky radiance²⁵ and whitecaps, to the signal of interest: the normalized water-leaving radiance $[L_w]_N$. The sensor is assumed to be similar to most space-borne ocean color instruments having a field of view represented by an approximately 1 km \times 1 km area on the water surface, so that the radiance due to Sun glitter represents an average over a very large number of facets. Note that the quantity plotted on the abscissa is radiance divided by F_0 . The plot is for a wavelength of 500 nm. The water is assumed to have a chlorophyll concentration of less than 0.2 mg/m³, and the wind is assumed to be omnidirectional.

The left panel in Figure 7.19 is for a nadir-viewing sensor (mode 1). Here we see that the radiance due to whitecaps is about an order magnitude lower than $[L_w]_N$, so whitecaps would not be a significant factor in the determination of $[L_w]_N$; however, note that the whitecap contribution increases rapidly with increasing wind speed. Similarly, for latitudes less than about 40° , the reflected skylight is significantly lower than $[L_w]_N$. In contrast, the direct Sun glitter is approximately a full to half an order of magnitude greater than $[L_w]_N$ for latitudes of 30° ($W = 5$ m/s) and 40° ($W = 10$ m/s). Retrieving $[L_w]_N$ in the presence of such strong Sun glitter would be a significant challenge. In the tilted configuration (mode 2, the right panel in Figure 7.19) one sees that the range of contamination of the desired $[L_w]_N$ by Sun glitter is considerably reduced extending the range (of latitudes) over which retrievals might be possible. The message here is that the contamination by Sun glitter (near the position where the specular image of the Sun would appear for a flat surface) is so large that useful results there are nearly impossible, and every effort must be made

²⁴The first ocean color sensor, the Coastal Zone Color Scanner (CZCS), was equipped with the capability to tilt the scan plane 20° forward and aft of nadir so as to avoid some of the Sun glitter. Mode (1) here is equivalent to the center of the CZCS scan when it is in the untilted mode, while mode (2) here equivalent to the center of the CZCS scan when it is tilted fully forward. The SeaWiFS sensor had the same capability.

²⁵For the reflected sky radiance in the figure, for simplicity we have assumed that $r_+(-\hat{\xi}_0) = r_f(-\hat{\xi}_0)$, the simple Fresnel reflectivity at the angle appropriate for a flat surface. The error in doing so is easily determined by noting that over the range $0 \leq \chi_i \leq 60^\circ$, $|r_+(-\hat{\xi}_0) - r_f(-\hat{\xi}_0)| \leq 0.004$. Since $L_{\text{sky}}/F_0 \lesssim 0.02$ over this range, the error in the reflected component is $\lesssim 8 \times 10^{-5}$, significantly less than $[L_w]_N/F_0$.

to avoid such regions of $\hat{\xi}_r$. Fortunately, since most ocean color instruments scan roughly from east to west, even in regions where there is significant glitter contamination at nadir, there will be regions along the scan for which the glitter contamination will be negligible.

7.6 Summary

In this chapter we have examined the transmission of radiance through, and reflection from, the air-water interface. The simple relationships developed for a flat surface become complex, but tractable, for a wind-ruffled interface. Modeling the latter as a set of plane facets enables the estimation of the distribution of surface slopes, which turns out to be approximately gaussian with a variance proportional to the wind speed. Applying the basic laws of refraction and reflection along with the Fresnel equations to the individual facets allowed a description of the propagation of radiance across such a surface. The most important results are: (1) the computation of the contribution of directly-reflected Sun light from the water surface — the glitter pattern given by Eq. (7.11); and (2) demonstration that the transmittance of a uniform $L_u(\hat{\xi})$ distribution incident on the interface from the water side is accurately provided by the Fresnel transmittance for a *flat* interface, as long the viewing angle (from above the interface) is less than about 60° . These results are the product of a first-order theory, in which photons interact with the water surface only once, and for which one portion of the surface does not shadow another — conditions that are valid for low wind speeds. In a brief section we provided a summary of what is presently known about the effects of whitecaps their relation to wind speed.

We also developed the idea of normalized water-leaving radiance (and reflectance), which will be the ultimate radiometric quantity derived from remote sensing, and used as inputs to algorithms for water constituents.

7.7 Appendix 1. Evaluation of an Integral

In this appendix we compute the integrals involved in

$$\frac{1}{S(\theta_i)} = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \int_{-\cot\theta_i}^{+\infty} (1 + s_x \tan\theta_i) \exp\left[-\left(\frac{s_x^2}{2\tilde{\sigma}^2}\right)\right] ds_x$$

This is of the form

$$I = \int_{-c}^{\infty} \exp(-bx^2) dx + \int_{-c}^{\infty} ax \exp(-bx^2) dx \triangleq I_1 + I_2$$

where a , b , and c are constants. First consider I_1 :

$$\begin{aligned}
 I_1 &= \int_{-c}^{\infty} \exp(-bx^2) dx = \int_{-c}^0 \exp(-bx^2) dx + \int_0^{\infty} \exp(-bx^2) dx \\
 &= \int_{-c}^0 \exp(-bx^2) dx + \frac{1}{2} \sqrt{\frac{\pi}{b}} \\
 &= \int_0^{+c} \exp(-bx^2) dx + \frac{1}{2} \sqrt{\frac{\pi}{b}} \\
 &= \frac{1}{2} \sqrt{\frac{\pi}{b}} \operatorname{erf}(\sqrt{bc}) + \frac{1}{2} \sqrt{\frac{\pi}{b}} \\
 &= \frac{1}{2} \sqrt{\frac{\pi}{b}} [1 + \operatorname{erf}(\sqrt{bc})],
 \end{aligned}$$

where $\operatorname{erf}(x)$ is the error function defined through

$$\operatorname{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt.$$

The integral I_2 is simple:

$$\begin{aligned}
 I_2 &= \int_{-c}^{\infty} ax \exp(-bx^2) dx = -\frac{a}{2b} \int_{-c}^{\infty} d[\exp(-bx^2)] \\
 &= \frac{a}{2b} [\exp(-bx^2)] \Big|_{-c}^{\infty} = \frac{a}{2b} \exp(-bc^2).
 \end{aligned}$$

Noting that $a = \tan \theta_i$, $b = 1/2\tilde{\sigma}^2$, and $c = \cot \theta_i$, yields the desired relationship:

$$\frac{1}{S(\theta_i)} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\cot \theta_i}{\sqrt{2\tilde{\sigma}}} \right) + \frac{\tilde{\sigma} \tan \theta_i}{\sqrt{2\pi}} \exp \left(-\frac{\cot^2 \theta_i}{2\tilde{\sigma}^2} \right) \right].$$

7.8 Appendix 2. Polarization of Sun Glitter

Here, we will briefly develop the polarization of reflected Sun glitter. The Mueller matrix for Fresnel reflection from the (flat) air-water interface was given earlier in Chapter 1:

$$\mathbf{r}_f(\theta_i) = \begin{pmatrix} r_+ & r_- & 0 & 0 \\ r_- & r_+ & 0 & 0 \\ 0 & 0 & r_{33} & 0 \\ 0 & 0 & 0 & r_{33} \end{pmatrix},$$

with

$$r_{\pm} = \frac{1}{2} \left[\frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} \pm \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} \right],$$

$$r_{33} = -\frac{\tan(\theta_i - \theta_t) \sin(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t) \sin(\theta_i + \theta_t)},$$

and $m_w \sin \theta_t = \sin \theta_i$. In analogy to the scalar case in the text, the reflected Stokes vector $\mathbf{I}_r(\hat{\xi}_r)$ for unpolarized incident solar radiation (Sun in a black sky) is

$$\mathbf{I}(\hat{\xi}_r) = \frac{p(\vec{s})}{4|\cos \theta_r| \cos^4 \theta_n} \mathbf{R}(-\alpha_r) \mathbf{r}_f(\chi) \mathbf{R}(-\alpha_i) \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where $\mathbf{R}(\psi)$ is the rotation matrix,

$$\mathbf{R}(\psi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\psi & -\sin 2\psi & 0 \\ 0 & \sin 2\psi & \cos 2\psi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The angles α_r and α_i are shown in Figure 7.20. The “−” sign in the arguments of the rotation matrices results from the fact that the rotations in Figure 7.20 are *clockwise* looking toward the light rather than counterclockwise (Chapter 1). We note that because the incident radiation is unpolarized, $\mathbf{R}(\alpha_i)$ is irrelevant, so

$$\mathbf{I}(\hat{\xi}_r) = \begin{pmatrix} I_r \\ Q_r \\ U_r \\ V_r \end{pmatrix} = \frac{p(\vec{s})F_0}{4|\cos \theta_r| \cos^4 \theta_n} \begin{pmatrix} r_+(\chi) \\ r_-(\chi) \cos(-2\alpha_r) \\ r_-(\chi) \sin(-2\alpha_r) \\ 0 \end{pmatrix},$$

where the law of sines applied to the spherical triangle in Figure 7.20 gives

$$\sin \alpha_r = \frac{\sin \theta_i \sin(\phi_r - \phi_i)}{\sin 2\chi},$$

and the angle χ is given by $\cos 2\chi = |\hat{\xi}_r \bullet \hat{\xi}_i|$. The degree of polarization is

$$P = \frac{\sqrt{Q_r^2 + U_r^2 + V_r^2}}{I_r} = \left| \frac{r_-(\chi)}{r_+(\chi)} \right|,$$

Noting that r_- is negative, we can write

$$\begin{aligned}
 \mathbf{I}(\hat{\xi}_r) &= K \begin{pmatrix} r_+(\chi) \\ r_-(\chi) \cos(-2\alpha_r) \\ r_-(\chi) \sin(-2\alpha_r) \\ 0 \end{pmatrix} = Kr_+ \begin{pmatrix} 1 \\ -P \cos(-2\alpha_r) \\ -P \sin(-2\alpha_r) \\ 0 \end{pmatrix} = Kr_+ \begin{pmatrix} 1 \\ P \cos(-2\alpha_r \pm \pi) \\ P \sin(-2\alpha_r \pm \pi) \\ 0 \end{pmatrix} \\
 &= Kr_+ \begin{pmatrix} 1-P \\ 0 \\ 0 \\ 0 \end{pmatrix} + Kr_+ \begin{pmatrix} P \\ P \cos(-2\alpha_r \pm \pi) \\ P \sin(-2\alpha_r \pm \pi) \\ 0 \end{pmatrix} \\
 &= \mathbf{I}_{Ur} + \mathbf{I}_{Pr},
 \end{aligned}$$

where \mathbf{I}_{Ur} and \mathbf{I}_{Pr} are, respectively the unpolarized and polarized components of \mathbf{I}_r , and

$$K \triangleq \frac{p(\vec{s})F_0}{4|\cos \theta_r| \cos^4 \theta_n}.$$

Now, since \mathbf{I}_{Pr} is in the standard form for linearly polarized light, i.e.,

$$\mathbf{I} = \begin{pmatrix} I \\ I \cos(2\gamma) \\ I \sin(2\gamma) \\ 0 \end{pmatrix},$$

and we identify γ as specifying the direction polarization (Chapter 1), the angle of polarization γ is given by

$$2\gamma = -2\alpha_r \pm \pi \implies \gamma = \pm \frac{\pi}{2} - \alpha_r.$$

Noting that $-\pi/2$ is the same angle as $+3\pi/2$, we see that the two values of γ refer to the same polarization direction:

$$\gamma = \frac{\pi}{2} - \alpha_r \quad \text{or} \quad \gamma = \frac{3\pi}{2} - \alpha_r.$$

Interestingly, neither P nor γ depend on the surface roughness. They depend only on the directions $\hat{\xi}_r$ and $\hat{\xi}_i$; however, the magnitudes of both Q_r and U_r do depend on $p(\vec{s})$. For most directions $\hat{\xi}_r$ and $\hat{\xi}_i$, the surface must be rough to get any reflected light at all. The polarization angle γ is easy to explain. For reflection from a dielectric, the light with \vec{E} perpendicular to the plane of incidence-reflection is more strongly reflected than that with \vec{E} parallel to the plane of incidence-reflection. So when unpolarized light is reflected from a tilted facet, the reflected light will have a larger \vec{E} perpendicular to the plane formed by $\hat{\xi}_r$ and $\hat{\xi}_i$ than in the plane parallel to it. Thus, examination of the reflected light with a

plane polarization analyzer will yield a maximum of transmission when the pass direction of the analyzer is perpendicular to the plane formed by $\hat{\xi}_r$ and $\hat{\xi}_i$, or making an angle $\gamma = \pi/2 - \alpha_r$ with the vertical ($\pm\hat{e}_z$).

It should be noted that these results apply to first order interactions with the surface. Should a photon reflect two or more times from the surface, e.g., an incident photon strikes the surface again after being reflected by a facet, its polarization will not follow the equations developed in this appendix.

7.9 Bibliographic Notes

7.2 Optical Effects of a Flat (Smooth) Water Surface

The material here was taken from Chapters 1 and 2.

7.3 Optical Effects of a Wind-Ruffled Water Surface

The material in much of this Section was developed by the author drawing heavily from [Cox and Munk \[1954\]](#), with some of the vector relationships (Eqs. (7.2) and (7.3)) inspired by [Walker \[1994\]](#).

7.3.2 Reflection of Light by a Wind-Ruffled Surface

In this section we alluded to the rigorous method of dealing with reflection and refraction from a random rough surface that involves use of the energy spectrum of waves. This is discussed in [Preisendorfer \[1976f\]](#), [Preisendorfer and Mobley \[1986\]](#) and [Mobley \[1994\]](#), and culminated in [Mobley \[2016\]](#), which treats the reflection light from water based on the wave energy spectrum, in its entirety, and from which the limitations of first-order theory developed in the text can be assessed. However, it still remains to carry out the complete radiative transfer problem – coupled water, surface, and atmosphere. This will likely require the next generation of super computers.

7.3.2.3 Estimation of $p(\vec{s})$ (Cox and Munk)

There is considerable debate as to the validity of the values assigned to the [Cox and Munk \[1954\]](#) parameters; however, there is little debate about the correctness of the basic formulation, i.e., the gaussian nature of $p(\vec{s})$, the [Shaw and Churnside \[1997\]](#) direct measurements using a scanning-laser glint meter demonstrated this, as did [Duntley \[1954\]](#) by a completely different method. Even the satellite measurements of glitter by [Ebuchi and Kizu \[2002\]](#) confirmed this basic distribution. Only the relationship of the parameters to

wind speed is controversial: the conclusions of Ebuchi and Kizu [2002] are *opposite* to those of Shaw and Churnside [1997] (and Hwang and Shemdin [1988]). Ebuchi and Kizu [2002] attribute this to the likelihood that the direct measurements were made under conditions in which the waves were *growing* with the wind, whereas in their measurements the waves were in *equilibrium* with the wind, and therefore, represent *average* conditions. Interestingly, the surface realizations based in wave energy spectra in Mobley [2016] show surface slope variances in reasonable agreement with Cox and Munk [1954]. The relationship of surface slope to wind speed is therefore still an open question, which may be answered as more complete energy spectra become available.

7.3.4 Requirements for Conservation of Energy Across the Water Surface

This section was developed completely by the author; however, ideas concerning the “shadowing” of a rough surface developed by Saunders [1967] helped crystalize the development.

7.3.5.1 General Considerations: Reciprocity

The form of the reciprocity principle used here was developed in Yang and Gordon [1997] and applied to radiative transfer across a rough surface in Gordon [2005].

7.3.6.1 General Methods

The methods described here are developed in Nakajima and Tanaka [1983] for scalar radiative transfer and in Gordon and Wang [1992] for vector theory. For the most general case, i.e., the wind direction and the solar beam neither parallel nor antiparallel, see Mobley [2018].

7.4 The Normalized Water-Leaving Radiance

The concept of a normalized water-leaving radiance was first introduced by Gordon and Clark [1981], used to in the processing of CZCS imagery [Gordon et al., 1988, 1983], and later discussed in considerable detail by Morel and Gentili [1996].

In this section it was assumed that $L_u(\hat{\xi}'_v, \hat{\xi}'_s)$ was independent of $\hat{\xi}'_v$. Gordon [1999] examined the case where L_u is not totally diffuse, but varies slowly with $\hat{\xi}'_v$. By expanding L_u in a power series in $\hat{\xi}'_v$, it was shown that

$$L_w(\hat{\xi}_v, \hat{\xi}_s) = \frac{t_-(-\hat{\xi}_s)}{m_w^2} L_u(\hat{\xi}'_c, \hat{\xi}'_s)$$

where $\hat{\xi}_c$ is a direction that is very close to, but not identical to, $\hat{\xi}'_v$. For a surface ruffled by an omnidirectional wind, it was shown that the polar angles θ'_c and θ'_v differed by less

than 4° for viewing angles $\pi - \theta_v \leq 75^\circ$ and $W \leq 16$ m/s, and that the difference decreases rapidly for smaller viewing angles. Thus, \Re is unchanged if L_u is not totally diffuse, but the water-leaving radiance is characteristic of a slightly different $\hat{\xi}'_v$ in the water. Examples in Chapter 6 should convince the reader that the variation in L_u when the viewing angle changes by $\lesssim 4^\circ$ is insignificant.

7.5 Whitecaps

Our knowledge of whitecap coverage and its relationship to water and meteorological parameters W and ΔT owes much to Monahan and co-workers [Monahan, 1969, 1971; Monahan and O’Muircheartaigh, 1980, 1986; Stabenro and Monahan, 1986]. His work, accomplished largely through photography was centered in determining the fractional coverage. Data presented by Bortkovskii [Bortkovskii, 1987] appears to show a dependence of coverage on the water temperature itself; however, Monahan and O’Muircheartaigh [1986] pointed out that it is not clear if the effect is real or due to the fact that the cold water measurements may not be representative because the seas may not have been fully developed in regions where such measurements have been made. For fully developed seas, $f_{wc} \sim W^\eta$, and for non-fully developed seas the value of η is reduced. Thus, the value of the exponent depends in some manner on the “duration” of the wind in the open ocean (where there is no fetch limitation).

Concerning the reflective properties, Whitlock et al. [1982] measured the reflectivity of laboratory-produced foam suggesting that the maximum reflectivity was $\sim 55\%$. Koepke [1984] estimated their *effective reflectivity*, accounting for the decay in reflectivity as the foam rack thins in an aging whitecap to be $\sim 22\%$. Frouin et al. [1996] measured the reflectance of breaking waves in the surf zone and were the first to suggest that there might be a significant decrease in the NIR reflectance of whitecaps. The measurements using the whitecap radiometer are described in detail in Moore et al. [1998, 2000].

7.10 Figures

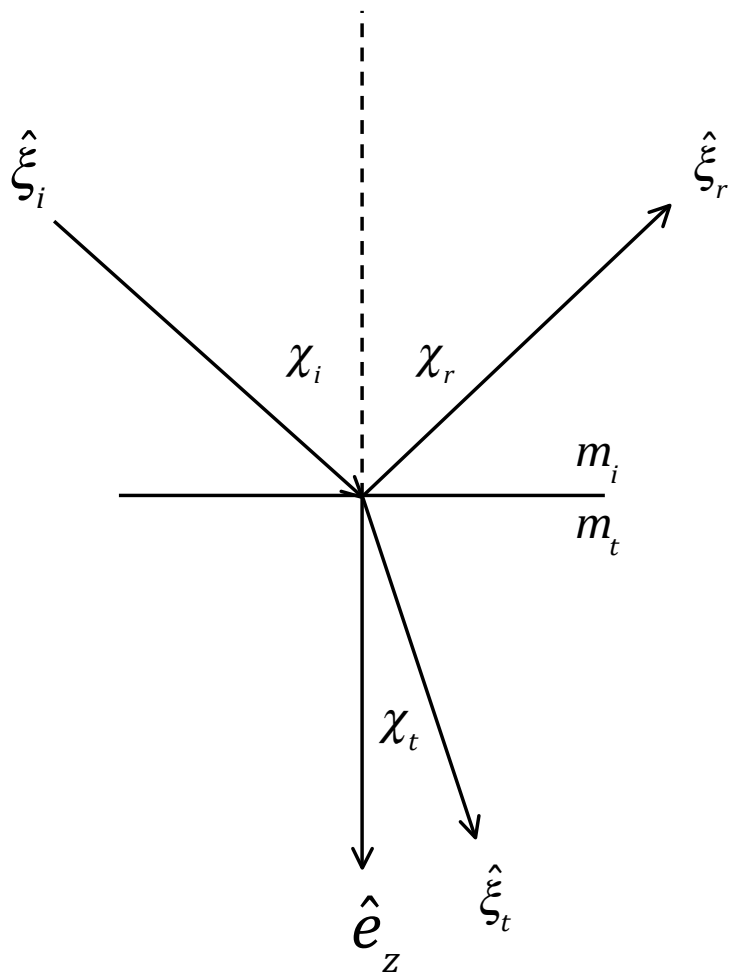


Figure 7.1: Refraction and reflection of light by an interface. The unit vectors $\hat{\xi}_i$, $\hat{\xi}_r$ and $\hat{\xi}_t$ are, respectively, the propagation directions of the incident, the reflected direction and the transmitted electromagnetic waves. The unit vector \hat{e}_z is the surface normal, and $\hat{\xi}_i$, $\hat{\xi}_r$, $\hat{\xi}_t$ and \hat{e}_z all lie in the same plane. The refractive indices on the two sides of the interface are m_i and m_t as shown. The law of reflection is $\chi_r = \chi_i$ or in vector form $\hat{\xi}_r = \hat{\xi}_i - 2\hat{e}_z(\hat{e}_z \cdot \hat{\xi}_i)$ and the law of refraction is $m_t \sin \chi_t = m_i \sin \chi_i$ or in vector form $m_t \hat{\xi}_t = m_i \hat{\xi}_i + (m_t \cos \chi_t - m_i \cos \chi_i)\hat{e}_z$. The drawing here is for $m_i < m_t$, so $\chi_t < \chi_i$.

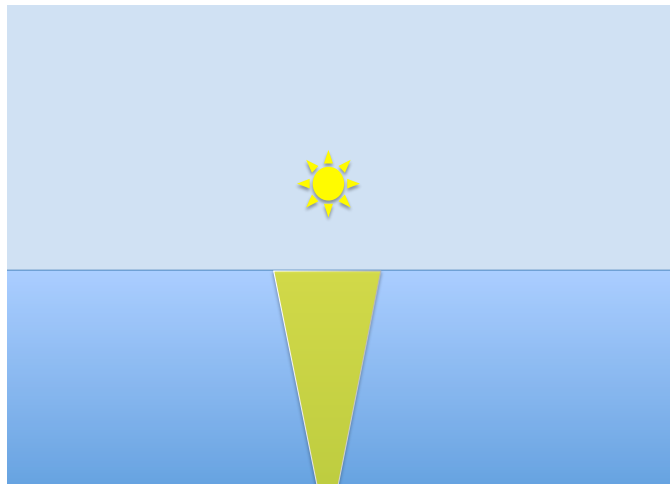


Figure 7.2: Schematic showing what an observer might see from the deck of a ship looking out over the sea surface toward the horizon in the direction of the rising or setting sun. The triangular patch is the reflection of the sun from the rough surface. The horizontal line between the light blue (sky) and dark blue (water) is the horizon.

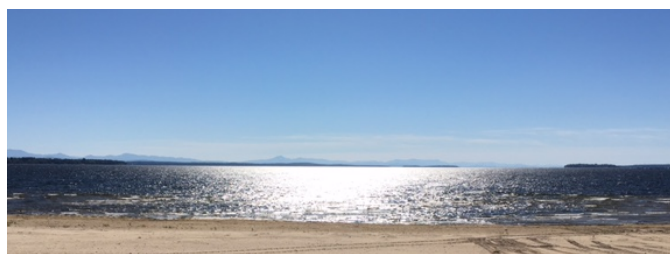


Figure 7.3: Photograph of Sun glitter taken September 21, 2016 from the shore of Lake Champlain. Here the solar zenith angle was approximately 52° .

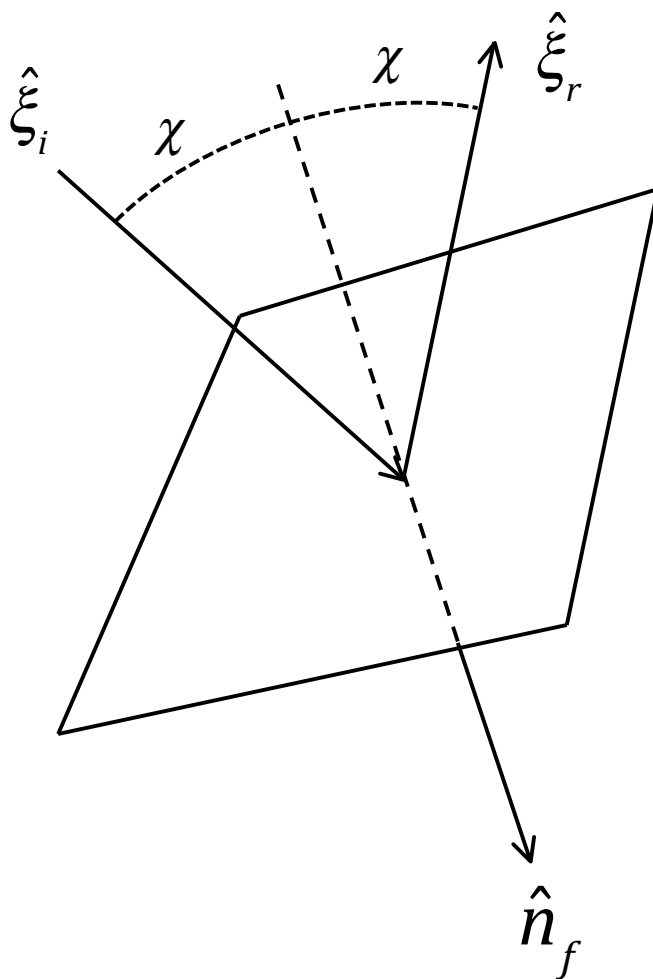


Figure 7.4: Schematic showing a facet on the water surface, the normal to the facet \hat{n}_f , and the incident ($\hat{\xi}_i$) and the reflected ($\hat{\xi}_r$) directions. The angle of incidence (and reflectance) on the facet is χ (also called χ_i in the text). The vectors \hat{n}_f , $\hat{\xi}_i$ and $\hat{\xi}_r$ all lie in the same plane. The z axis (and \hat{e}_z) is vertically downward in the drawing and is *not* in the plane of the other three vectors.

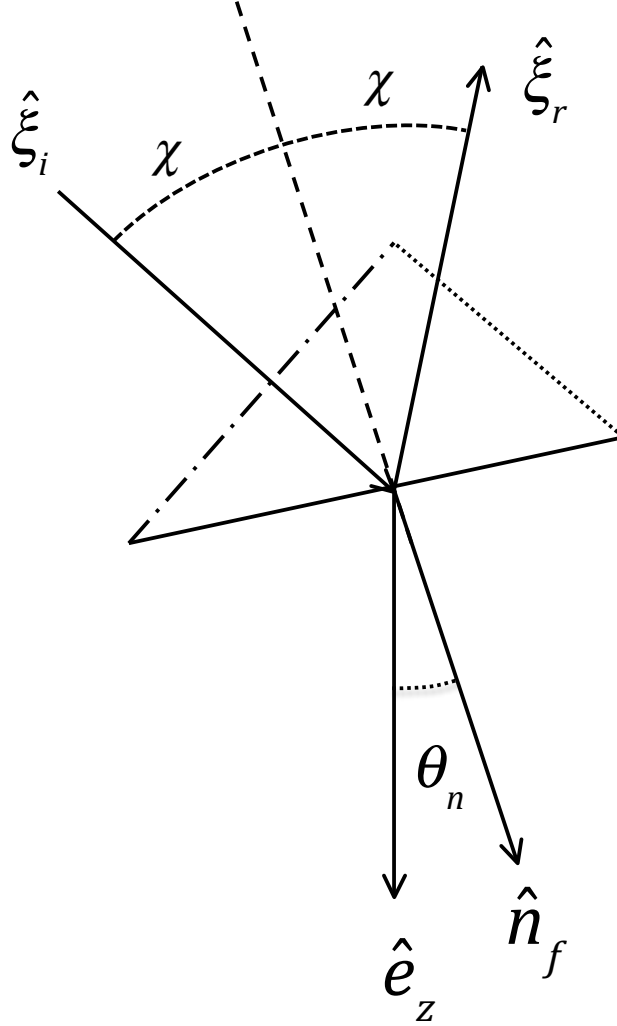


Figure 7.5: Schematic showing a single facet. In contrast to Figure 7.4, the plane of the drawing is the plane containing the normal to the facet \hat{n}_f , the incident ($\hat{\xi}_i$) and the reflected ($\hat{\xi}_r$) directions. The unit vector \hat{e}_z is *not* in this plane, but makes an angle θ_n with the normal. The area of the facet is a_i . The dash-dot-dash line ($-\cdot-\cdot-$) represents a plane perpendicular to the incident direction, and is drawn so that its area is the projected area (in the incident direction) of the facet, i.e., this projected area is $a_i \cos \chi$.

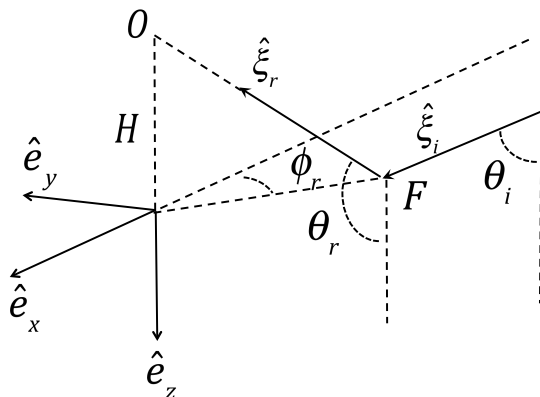


Figure 7.6: Schematic showing an observer (O) at a height (H) above the water surface looking toward the surface so as to observe a reflected photon propagating in the direction $\hat{\xi}_r$, after being reflected from a properly oriented facet at F . The coordinates (x, y) of F are given by $x = -H \tan(\pi - \theta_r) \cos \phi_r$ and $y = -H \tan(\pi - \theta_r) \sin \phi_r$.

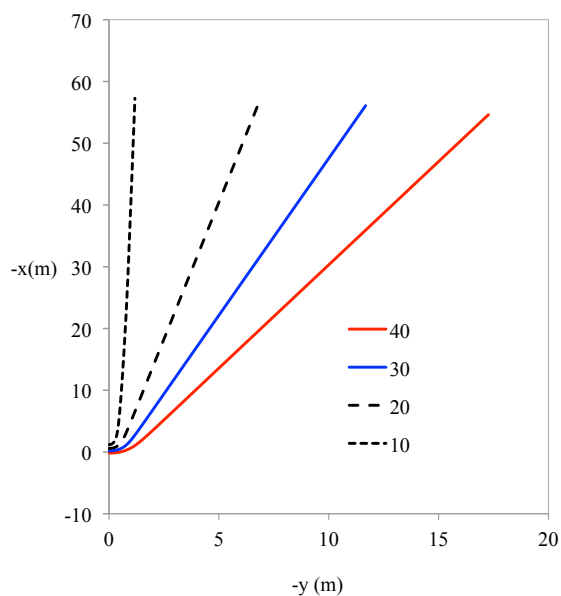


Figure 7.7: Position in Figure 7.6 from which the reflected image of the sun will appear for a given θ_n assuming that all ϕ_n 's are equally probable. There is an equivalent set of lines having positive values of y , leading to a “V”-shaped figure for a given value of θ_n (in degrees). Here the solar zenith angle $\theta_i = 70^\circ$ and the observer is located at $x = y = 0$, and 1 m above the water surface.

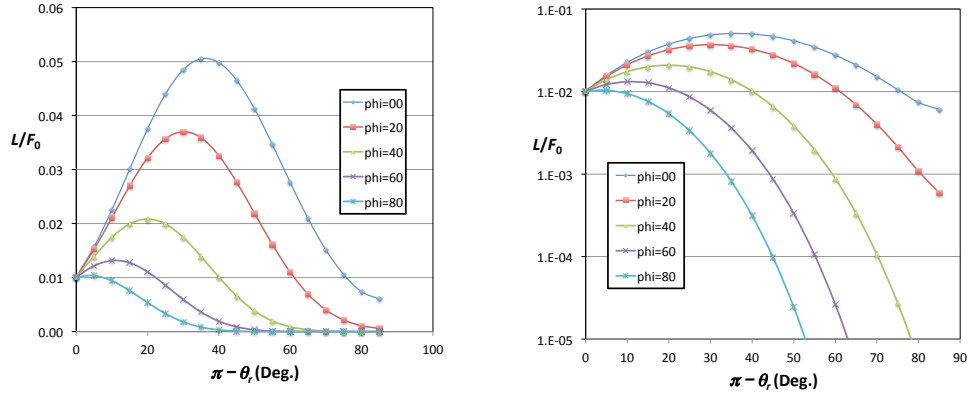


Figure 7.8: Radiance reflected from a wind-ruffled water surface. The source is the sun in a black sky with a zenith angle of 30° ($\hat{\xi}_i$ is in the x - z plane). The azimuth of $\hat{\xi}_r$ is “phi” (in degrees) measured in the x - y plane relative to \hat{e}_x . The wind vector \vec{W} is in the x -direction, i.e., has *only* an \hat{e}_x component. Thus, for “phi = 0”, the projection $\hat{\xi}_r$ in the x - y plane is parallel to \vec{W} , while for “phi = 90° ”, it is perpendicular to \vec{W} . The wind speed is 7.5 m/s. The angle $\pi - \theta_r$ is the angle that $\hat{\xi}_r$ makes with the *outward* normal, i.e., surface normal pointing *out* of the water. Left Panel: Linear scale. Right panel: Log scale.

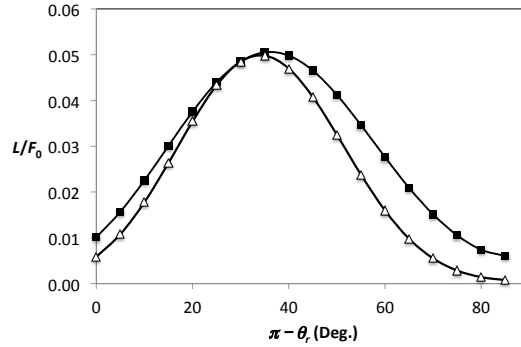


Figure 7.9: Radiance reflected from a wind-ruffled water surface. The source is the sun in a black sky with a zenith angle of 30° . The wind is parallel to \hat{e}_x . For the filled symbols the plane containing $\hat{\xi}_i$, $\hat{\xi}_r$, and \hat{e}_z also contains \vec{W} , while for the open symbols, it is perpendicular to \vec{W} . The wind speed is 7.5 m/s. The angle $\pi - \theta_r$ is the angle that $\hat{\xi}_r$ makes with the *outward* normal, i.e., surface normal pointing *out* of the water ($-\hat{e}_z$).

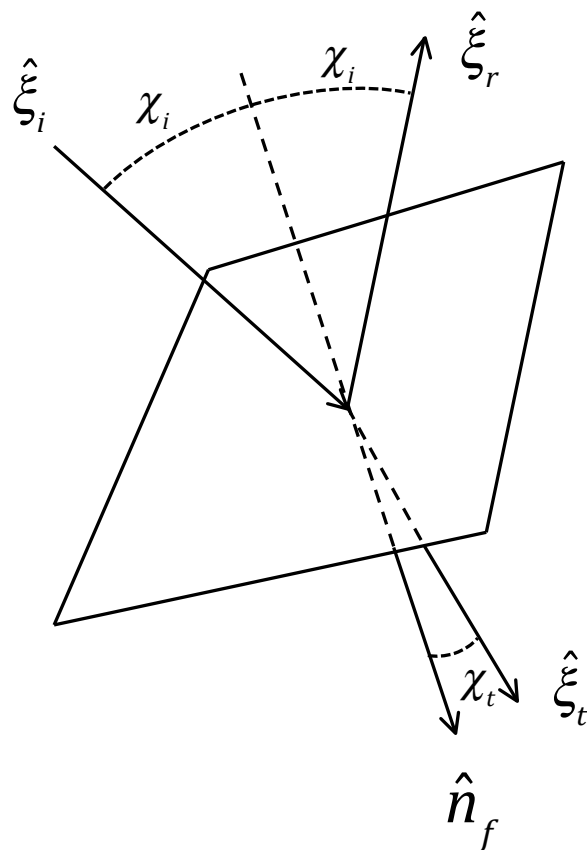


Figure 7.10: Schematic showing a facet on the water surface, the normal to the facet \hat{n}_f , the incident ($\hat{\xi}_i$), the reflected ($\hat{\xi}_r$) and the refracted (transmitted, $\hat{\xi}_t$) directions. The angle of incidence (and reflectance) on the facet is χ_i and the angle of refraction is χ_t . The z axis (and \hat{e}_z) is vertically downward in the drawing, but not in the plane containing $\hat{\xi}_i$, $\hat{\xi}_r$, $\hat{\xi}_t$ and \hat{n}_f .

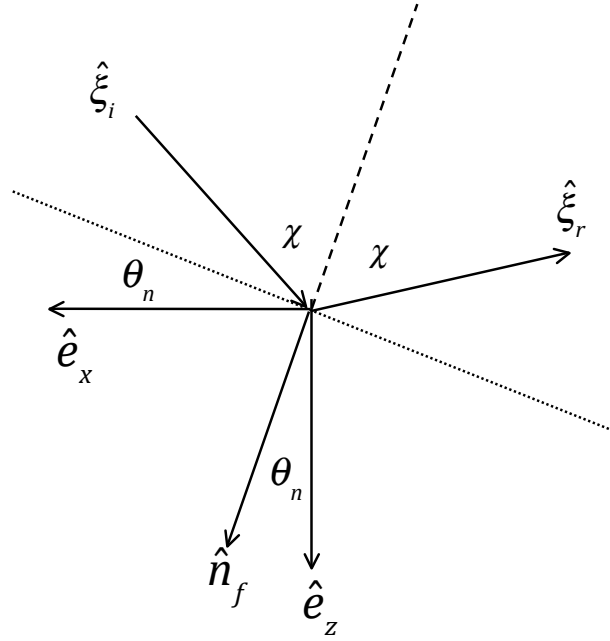


Figure 7.11: Schematic showing a facet (dotted line for clarity) aligned so that its normal n_f is in the x - z plane. All of the vectors shown are also in the same x - z plane.

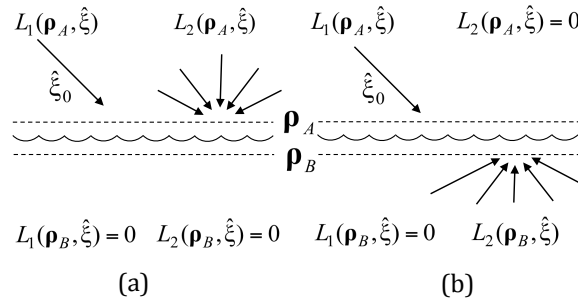


Figure 7.12: Schematic two cases (a) and (b) for which the reciprocity principle is applied for radiances L_1 and L_2 shown incident on the surfaces ρ_A and ρ_B that straddle the interface. (From [Gordon \[2005\]](#))

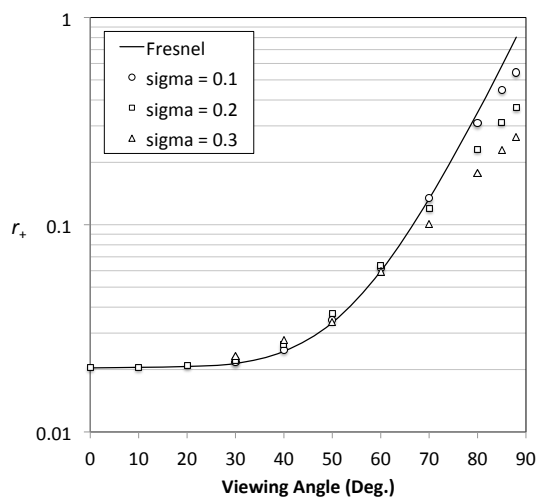


Figure 7.13: Reflection factor $r_+(\hat{\xi}_r)$ for totally diffuse sky radiance as a function of the viewing angle $(\pi - \theta_r)$ for various values of $\tilde{\sigma}$ and an omnidirectional wind. The line labeled “Fresnel” is $r_f(\pi - \theta_r)$ and corresponds to a smooth surface.

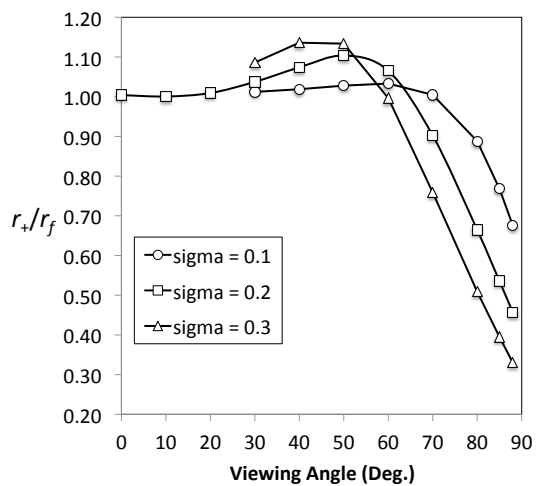


Figure 7.14: Reflection factor $r_+(\hat{\xi}_r)/r_f$ for totally diffuse sky radiance as a function of the viewing angle $(\pi - \theta_r)$ for various values of $\tilde{\sigma}$ and an omnidirectional wind.

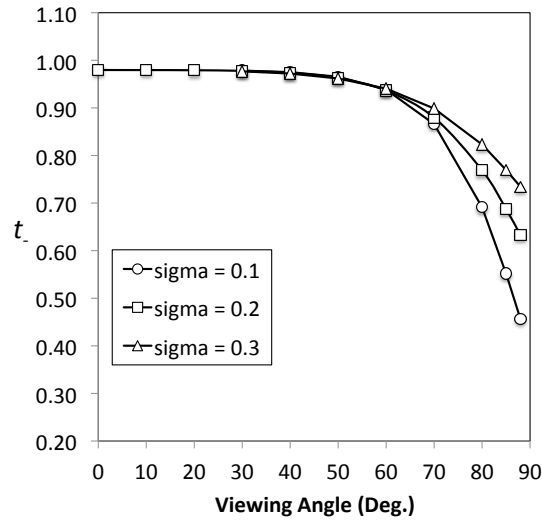


Figure 7.15: Transmittance factor $t_+(\hat{\xi}_r)$ for totally diffuse upwelling subsurface radiance as a function of the viewing angle $(\pi - \theta_r)$ in air for various values of $\tilde{\sigma}$ and an omnidirectional wind.

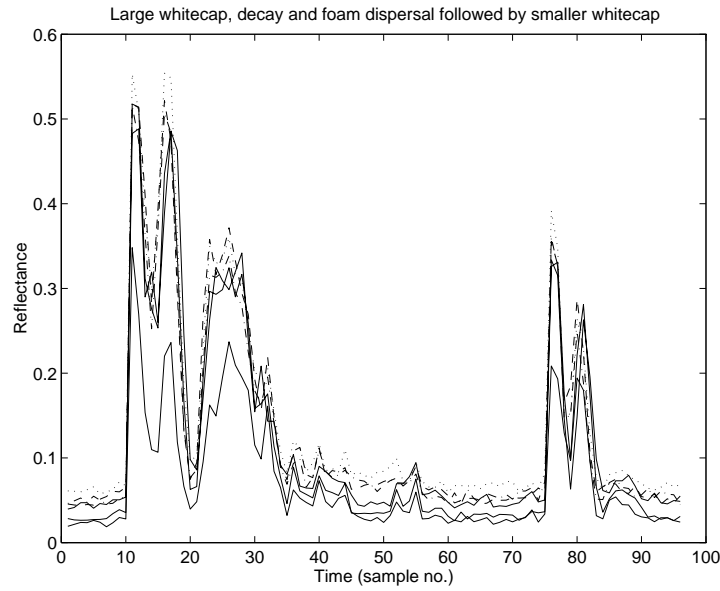


Figure 7.16: An approximately 15 second record of the reflectance of two whitecaps passing within the field of view of the radiometer. The lowest line corresponds to 860 nm. Redrawn from [Moore et al. \[2000\]](#).

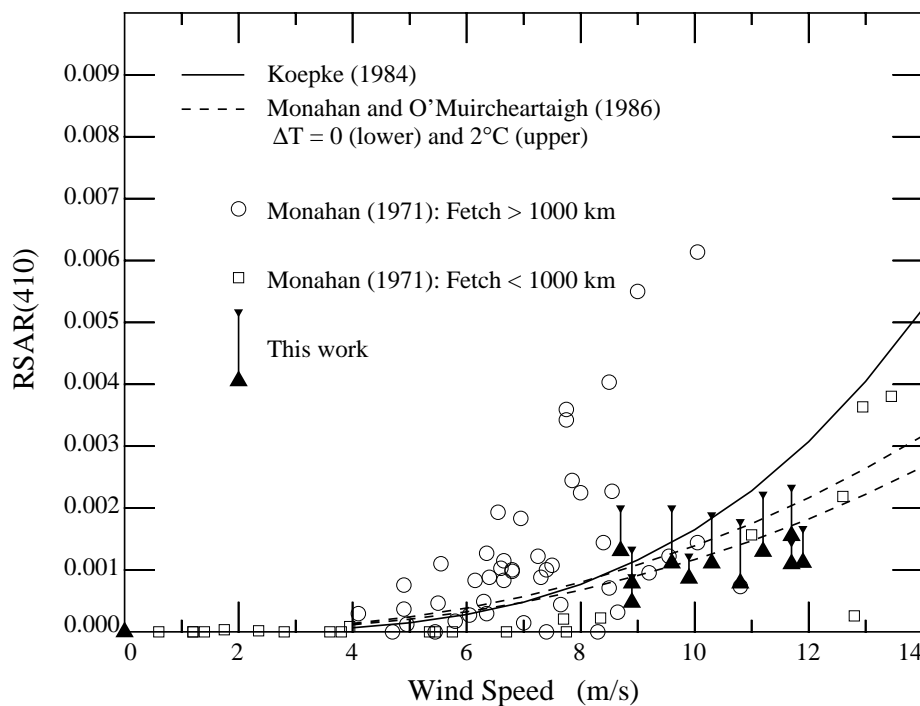


Figure 7.17: Remote-sensing augmented reflectance of whitecaps at 410 nm. The small and large triangles are from Moore et al. [2000] and correspond to two methods of analyzing the whitecap measurements. The open symbols are the Monahan [1971] fractional coverage multiplied by 0.155, the value of the Koepke [1984] effective whitecap reflectance of 0.22 minus 0.065, in order to convert from reflectance to augmented reflectance. The solid line is the formula developed from Koepke's work (Eq. (7.42)) with the overall reflectance reduced by 1/3 to better the Moore et al. [2000] radiometer data: $RSAR \sim 1.6 \times 10^{-7} W^{3.52}$. The dashed lines use the Monahan and O'Muircheartaigh [1986] model for a neutrally stable ($\Delta T = 0$) and an unstable ($\Delta T = 2^\circ\text{C}$) atmosphere to provide fractional coverage for use in computing the augmented reflectance: $RSAR = 3 \times 10^{-6} W^{2.55} \exp(0.861 \times \Delta T)$. From Moore et al. [2000].

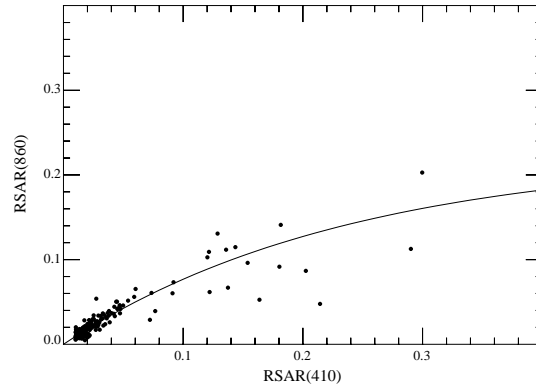


Figure 7.18: Relationship between RSAR at 865 nm and 410 nm (from [Moore et al. \[2000\]](#)).

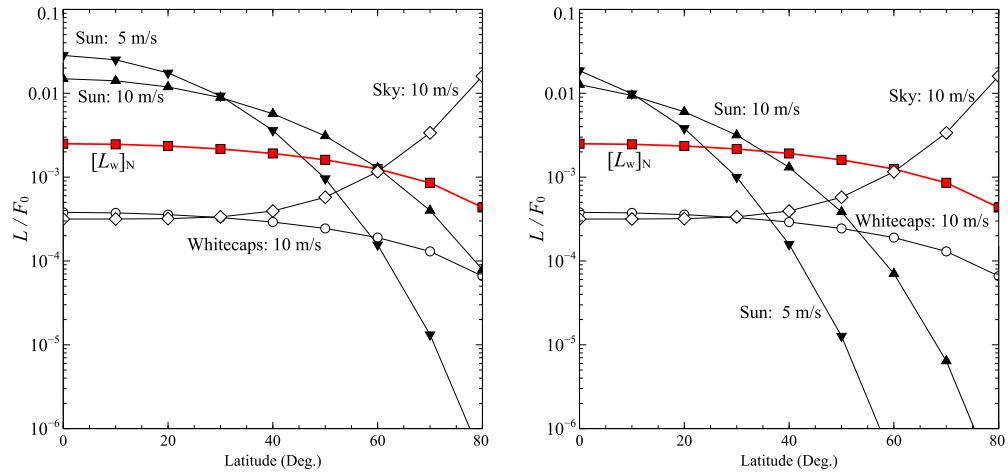


Figure 7.19: Comparison of the radiances leaving the water surface seen by a sensor looking down from a satellite as a function of latitude at the equinox. Solid squares (red curve) are for the normalized water-leaving radiance at 500 nm. Downward and upward pointing triangles are the Sun glitter radiance for omnidirectional wind speeds of 5 and 10 m/s, respectively. Open circles are the whitecap radiance $W = 10$ m/s (there are no whitecaps for $W = 5$ m/s). The open diamonds are the reflected sky radiance for $W = 10$ m/s. Left panel: nadir viewing ($\theta_r = 180^\circ$). Right panel: tilted viewing ($\theta_r = 160^\circ$, $\phi_r = 180^\circ$).

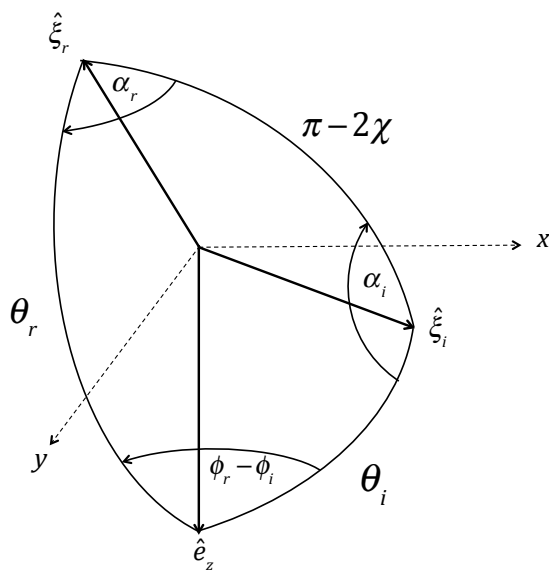


Figure 7.20: Light propagating along $\hat{\xi}_i$ is reflected in the direction $\hat{\xi}_r$ from a rough water surface. The angle between the planes formed by $(\hat{\xi}_r, \hat{e}_z)$ and $(\hat{\xi}_i, \hat{e}_z)$ is $\phi_r - \phi_i$. The angle between the planes formed by $(\hat{\xi}_r, \hat{e}_z)$ and $(\hat{\xi}_r, \hat{\xi}_i)$ is α_r . The angle between the planes formed by $(\hat{\xi}_i, \hat{e}_z)$ and $(\hat{\xi}_r, \hat{\xi}_i)$ is α_i . The angle α_i is positive if the rotation of the plane $(\hat{\xi}_i, \hat{e}_z)$ into the plane $(\hat{\xi}_i, \hat{\xi}_r)$ is counterclockwise looking toward the light (i.e., in the direction $-\hat{\xi}_i$). Likewise, the angle α_r is positive if the rotation of the plane $(\hat{\xi}_i, \hat{\xi}_r)$ into the plane $(\hat{\xi}_r, \hat{e}_z)$ is counterclockwise looking toward the light (i.e., in the direction $-\hat{\xi}_r$). So both α_r and α_i are entered as $-\alpha_i$ and $-\alpha_r$ in the rotation matrices based on this drawing (clockwise rotations), because the computation via the law of sines will yield positive numbers for each. The angles θ_r , θ_i , and $\pi - 2\chi$ are great circles on the unit sphere on which the ends of the vectors $\hat{\xi}_r$, $\hat{\xi}_i$, and \hat{e}_z fall.

Chapter 8

The Sun and Moon

8.1 Introduction

In the chapter we examine the physical properties of the Sun and the Moon that are important in remote sensing.

8.2 The Sun

The Sun is the source for the radiation reflected from the ocean-atmosphere system. To interpret the reflected radiance, we need to know how to predict the position of the sun in sky and understand the spectrum and strength of the incident solar radiation. This is the subject of this section. As the motion of the Sun in the sky is determined by the motion of the Earth around the Sun, we start by examining the motion of the Earth around the Sun and the rotation of the Earth about its own axis. Next, we use this information to learn how to predict the position of the Sun in the sky at any location and time. Finally, we provide the solar spectrum and the dependence of the extraterrestrial solar irradiance on time.

8.2.1 Motion of the Earth relative to the Sun and Fixed Stars

The Earth moves around the Sun in an elliptical orbit with the Sun at one of the two foci of the ellipse. This orbit is shown in Figure 8.1. The period of the orbit is one year or 365.2422 twenty-four hour days. The position of the Earth relative to the Sun is specified by the vector \vec{r} . This vector makes an angle θ with respect to the orbit perihelion (position of smallest distance between Earth and Sun). The perihelion of the orbit is near January 3 of each year. The semi-major and semi-minor axes of the ellipse are a and b , respectively. The parameter ε is the eccentricity of the ellipse and has a value of approximately 0.0167. The equation of the orbit is

$$r = \frac{a(1 - \varepsilon^2)}{1 + \varepsilon \cos \theta}. \quad (8.1)$$

Thus, due to the small value of ε the orbit is almost circular: at $\theta = 0$, $r = a(1 - \varepsilon)$ (the perihelion) and $\theta = 180^\circ$, $r = a(1 + \varepsilon)$ (the aphelion). The length of the semi-major axis is $\sim 149,598$ km, which is 1 astronomical unit (AU). The semi-minor axis has length $a\sqrt{1 - \varepsilon^2}$, so a and b differ by only about 0.01%, i.e., b is only 0.01% less than 1 AU. The average distance from the Earth to the Sun is

$$\langle r \rangle \triangleq \frac{1}{2\pi} \int_0^{2\pi} r d\theta = b.$$

Conservation of angular momentum leads to Kepler's second law, which states that as the Earth moves in its orbit the \vec{r} sweeps out equal areas in equal times. Thus, the Earth must move faster than average in the vicinity of the perihelion and slower in the vicinity of the aphelion.¹

As the Earth orbits around the Sun, it also rotates around its own axis. This rotation is depicted schematically in Figure 8.2. The direction of the rotation axis is fixed (on decadal time scales) relative to the fixed stars and in our epoch is directed near the north star (Polaris).² The angle between the axis of rotation and the plane of the Earth's orbit (the ecliptic) is 23.43° (δ_N). From the figure it is clear that at different times of the year the Sun's rays are normal to the Earth's surface at different latitudes: in Northern Hemisphere Summer these latitudes are north of the Equator, while in Winter they are south of the Equator. At the positions marked "Spring" and "Fall" (the vernal and autumnal equinox, respectively) the Sun rays are normal to the Earth's surface at the Equator. In Figure 8.2

¹The angular momentum \vec{L} of the Earth with respect to the origin of coordinates at the Sun is $\vec{L} = \vec{r} \times m\vec{v}$, where \vec{v} is the Earth's velocity. Its rate of change is $d\vec{L}/dt = \vec{r} \times m d\vec{v}/dt = \vec{r} \times \vec{F}$, where \vec{F} is the force on the Earth. Since the force is parallel to \vec{r} , $d\vec{L}/dt = 0$ and \vec{L} is constant. The magnitude of \vec{L} is $mr^2 d\theta/dt$, so $r^2 d\theta/dt$ is constant and $d\theta/dt$ is inversely proportional to r^2 . The area swept out by the radius vector when θ changes by $d\theta$ is $dA = r^2 d\theta/2$, so dA/dt is constant.

²The rotation axis actually moves over the surface of a cone centered on the center of the Earth with a period of 25,772 years.

the orbital position marked “Winter” (winter solstice) occurs near December 21; about 13 days before the orbit perigee is reached. As viewed from the Earth, the Sun reaches its highest point in the sky near local noon. In the Northern Hemisphere this occurs at the orbital position marked “Summer.” At all other times during the year, the solar zenith angle is larger at noon.

The declination of the Sun or solar declination is the latitude at which the Sun’s rays are normal to the Earth’s surface. This is relatively easy to compute. Figure 8.3 provides the geometry relevant to the computation. In the figure, the shaded area is in the plane of the Earth’s orbit (x - y plane). The vector \hat{r} is in the direction of \vec{r} in Figure 8.1 and is the direction of propagation of the Sun’s radiation. The x -axis is directed toward the perihelion of the orbit, \hat{w} is toward the position of the Earth at winter solstice, and the z -axis is normal to the ecliptic. The unit vector \hat{N} is the axis of rotation and the angle δ_N is inclination of the Earth’s rotation axis to the plane of the orbit. It is in the plane containing \hat{w} and the z axis. The angles θ_w and θ_E are both defined a positive in the diagram. The problem is to find α , the angle between \hat{N} and \hat{r} . Clearly, $\delta_N - \pi/2 \leq \alpha \leq \delta_N + \pi/2$. The computation of α is not difficult: $\cos \alpha = \hat{N} \bullet \hat{r}$. The vectors \hat{r} and \hat{N} are given by

$$\begin{aligned}\hat{r} &= \hat{e}_x \cos \theta_E + \hat{e}_y \sin \theta_E, \\ \hat{N} &= -\hat{e}_x \cos \theta_w \sin \delta_N + \hat{e}_y \sin \theta_w \sin \delta_N + \hat{e}_z \cos \delta_N.\end{aligned}$$

These result in

$$\cos \alpha = -\cos(\theta_E + \theta_w) \sin \delta_N.$$

The Earth’s equator is in a plane which has \hat{N} as its normal. Thus, the latitude at which the Sun’s rays are normal to the surface (λ_S) is given by $\alpha + \lambda_S = \pi/2$, so,

$$\sin \lambda_S = -\cos(\theta_E + \theta_w) \sin \delta_N \quad (8.2)$$

and by definition the solar declination is $\delta = \lambda_S$. As required, near the winter solstice $\theta_E + \theta_w = 0$ and $\lambda_s = -\delta_N$, i.e., in the Southern Hemisphere, and near the summer solstice, $\lambda_s = \delta$. If N is the day of the year in question ($N = 0$ is January 1), noting that the winter solstice is 10 days earlier than the $N = 0$, if desired, $\theta_E + \theta_w \approx (N + 10) \times 360/365$ in degrees.

The motion of the Earth is influenced by the motion of the Moon, which is not accounted for in our derivation. Through Fourier analysis, the declination (in degrees) can be approximated by

$$\begin{aligned}\delta &= 0.396^\circ - 22.868^\circ \cos \gamma + 4.025^\circ \sin \gamma - 0.387^\circ \cos 2\gamma \\ &\quad + 0.052^\circ \sin 2\gamma - 0.155^\circ \cos 3\gamma + 0.0848^\circ \sin 3\gamma,\end{aligned} \quad (8.3)$$

where $\gamma = 2\pi N/365$. The declination of the Sun as a function of time during the year is provided in Figure 8.4.

8.2.2 Position of the Sun in the Sky

The geometry for determination of the position of the sun in the sky is provided in Figure 8.5. In the figure, the observer is at the origin. The bold vector \hat{S} is a unit vector from the position of the observer pointed toward the sun. The bold vector \hat{n} is the normal to the Earth's surface at the observer's position, i.e., it points from the observer directly toward the observer's zenith. The z -axis of the coordinate system is parallel to the rotational axis of the Earth, and points north. The x -axis is fixed with respect to the Earth and falls along the Greenwich meridian (dash-dot-dash curve from the z -axis to the x -axis). The polar and azimuth angles of \hat{S} are $\pi/2 - \delta$ and Lon_s , respectively. The value of Lon_s increases with time during the day. The angle δ is the declination of the Sun. The polar angles of the observer's normal are θ_o and Lon_o , respectively. The observer's latitude is $\lambda_o = \pi/2 - \theta_o$. The solar zenith angle is θ_s . It is the angular distance between \hat{S} and \hat{n} in the plane formed by these two vectors. It can be found from $\cos \theta_s = \hat{S} \bullet \hat{n}$. The solar azimuth angle ϕ_s is defined as the angle between the plane formed by \hat{n} and the z -axis, and the plane formed by \hat{S} and \hat{n} measured toward the East from the former to the latter. It can be found from the small spherical triangle formed by \hat{S} , \hat{n} , and the z -axis. From the law of sines for spherical triangles,

$$\frac{\sin \phi_s}{\sin(\pi/2 - \delta)} = \frac{\sin(Lon_o - Lon_s)}{\sin \theta_s}.$$

The Sun's hour angle at the observer is defined to be $S_h \triangleq Lon_s - Lon_o$, i.e., as shown in the figure, the hour angle is negative.

$$\sin \phi_s = -\frac{\sin(S_h)}{\sin \theta_s} \cos \delta. \quad (8.4)$$

To derive the solar zenith angle it is easiest to write the vectors \hat{S} and \hat{n} in cartesian coordinates:

$$\begin{aligned} \hat{S} &= \hat{e}_x \cos(Lon_s) \sin(\pi/2 - \delta) + \hat{e}_y \sin(Lon_s) \sin(\pi/2 - \delta) + \hat{e}_z \cos(\pi/2 - \delta) \\ &= \hat{e}_x \cos(Lon_s) \cos \delta + \hat{e}_y \sin(Lon_s) \cos \delta + \hat{e}_z \sin \delta; \\ \hat{n} &= \hat{e}_x \cos(Lon_o) \sin \theta_o + \hat{e}_y \sin(Lon_o) \sin \theta_o + \hat{e}_z \cos \theta_o \\ &= \hat{e}_x \cos(Lon_o) \cos \lambda_o + \hat{e}_y \sin(Lon_o) \cos \lambda_o + \hat{e}_z \sin \lambda_o, \end{aligned}$$

so

$$\begin{aligned} \cos \theta_s &= \hat{S} \bullet \hat{n} = \cos(Lon_o) \cos \lambda_o \cos(Lon_s) \cos \delta \\ &\quad + \sin(Lon_o) \cos \lambda_o \sin(Lon_s) \cos \delta + \sin \lambda_o \sin \delta \\ &= \cos(Lon_s - Lon_o) \cos \lambda_o \cos \delta + \sin \lambda_o \sin \delta \\ &= \cos(S_h) \cos \lambda_o \cos \delta + \sin \lambda_o \sin \delta. \end{aligned} \quad (8.5)$$

Thus, to determine the position of the Sun in the sky given the time of day using Eqs. (8.4) and (8.5) we need only know the longitude of the sun at the given time, the observer's position on the Earth (latitude and longitude) and the declination of the sun on the given day. The declination is given by Eq. (8.3).³ Finding the Lon_s requires more work.

The Earth rotates 360° on its axis (\hat{N}) in $23^{\text{hr}}56^{\text{m}}4.1^{\text{s}} = 23.934470$ hr. Assume for the moment that the Earth's orbit is circular ($\varepsilon = 0$) and the $\delta_N = 0 = \delta$. Then, because of the Earth's orbital motion, for the Sun to return the same position in the sky at the same time each day, the Earth has to rotate a little more than 360° in one day. How much more? In the Appendix we show that in a year the Earth must make one full rotation more than number of days per year, so the length of the day is $23.934470 \times (366.2422/365.2422)$ or almost exactly 24 hr. This defines what is called the *mean solar day* ("mean" because of the assumptions $\varepsilon = 0$ and $\delta_N = 0$). Thus, the mean motion of the Sun as seen from Earth is $360^\circ/24\text{hr} = 15^\circ\text{hr}^{-1}$, i.e., 15° longitude per hour. (Note that under these assumptions, Eq. (8.5) gives $\cos \theta_s = \cos S_h \cos \lambda_o$, so the maximum value of θ_s is λ_o .) As the mean solar day is exactly 24 hr, it is in synchrony with our mechanical clocks. It is customary to define the zero of time each day so that the fictive mean Sun is at its highest point in the sky over Greenwich, UK (where the longitude is *defined* to be 0°) at 12 noon. This is called Greenwich Mean Time (GMT). Since the Sun moves 15° longitude per hour toward the West, the longitude of the mean Sun is given by

$$\langle Lon_s \rangle = (GMT - 12) \times 15^\circ,$$

where GMT is the Greenwich Mean Time in hours (the angle brackets on Lon_s indicate that this is the *mean* Lon_s). However, we are not finished yet: we need the actual not the mean longitude of the Sun.

The actual longitude of the Sun (Lon_s) differs from the mean longitude ($\langle Lon_s \rangle$) for two reasons: (1) the orbit of the Earth is elliptical not circular; and (2) $\delta_N \neq 0$. In the Appendix (Section 8.5) we show how these two facts determine $Lon_s - \langle Lon_s \rangle$. The final result is called the equation of time (EOT):

$$EOT = Lon_s - \langle Lon_s \rangle$$

Through Fourier analysis of solar data, the EQT can be approximated by

$$EQT = \frac{180^\circ}{\pi} \times (0.0000075 + 0.001868 \cos \gamma - 0.032077 \sin \gamma - 0.014615 \cos 2\gamma - 0.040849 \sin 2\gamma) \quad (8.6)$$

where again, $\gamma = 2\pi N/365$. The variation of EQT throughout the year is provided in Figure 8.6.

³The declination may be assumed to be constant throughout the day. The maximum change in δ according to Eq. (8.2) is 0.4° in 24 hr (0.2° in 12 hr) near $\theta_E + \theta_w = 90^\circ$.

Using the equation of time, we have

$$\begin{aligned}
 S_h &= Lon_s - Lon_o \\
 &= \langle Lon_s \rangle - Lon_o + EQT \\
 &= (GMT - 12) \times 15^\circ - Lon_o + EQT,
 \end{aligned} \tag{8.7}$$

providing the last ingredient needed to compute the solar azimuth and zenith angles using Eqs. (8.4) and (8.5).⁴

It should be noted, that these formulas for computing the position of the Sun in the sky ignore refraction by the atmosphere. For $\theta_s < 80^\circ$ the refractive effect is small: the error in θ_s when it is near 60° is $\sim 0.08^\circ$. When the Sun is near the horizon, the error is about one solar diameter, i.e., when the sun appears at the horizon its actual zenith angle is about 90.5° rather than 90° . This effect can be ignored in remote sensing applications where solar zenith angles are almost always $< 70 - 80^\circ$.

8.2.3 Extraterrestrial Solar Irradiance

We wish to derive an expression for the solar irradiance falling on a plane at the position of the Earth and oriented normal to the propagation direction of radiation from the Sun. Consider an area dA_{Si} at position i on the solar disk. The radiance leaving this area is L_{Si} . This is also the radiance falling on an area dA_E at the top of the atmosphere. From the definition of radiance, the power falling on this area is

$$\frac{d^2 P_{Ei}}{dA_E d\Omega_{Si}} = L_{Si},$$

where $d\Omega_{Si}$ is the solid angle subtended by dA_{Si} at the position of dA_E , i.e., $d\Omega_{Si} = dA_{Si}/r^2$, where r is the Earth-Sun distance. Thus, the irradiance on dA_E due to element at position i on the Sun is

$$E_{Ei} = \frac{d^2 P_{Ei}}{dA_E} = \frac{L_{Si} dA_{Si}}{r^2},$$

⁴The “equation of time” derived its name from the fact that it represented the difference between solar time and the time provided by mechanical clocks. If one is at the center of a time zone, local solar noon is defined by the instant at which the sun is highest in the sky, i.e., the time when a person’s shadow is shortest. However, the mean solar noon is defined as the time the *mean* Sun is highest in the sky, which is the same time indicated by a mechanical clock. Noting that the Earth rotates through 15° of longitude per hour, the time difference between a mechanical clock and a Sun clock (i.e., sundial) is $t_{\text{Sun}} - t_{\text{Mechanical}} = EQT/15^\circ$, where the time difference is in hours. Figure 8.6 shows that the maximum correction to a Sun clock is about 16 min, and that the Sun clock is running at its slowest compared to a mechanical clock near February 12 and at its fastest near November 4.

and the total irradiance from all portions of the Sun is

$$E_E = \frac{dP_E}{dA_E} = \frac{\int L_{Si} dA_{Si}}{r^2}. \quad (8.8)$$

The integral is a constant (C) characteristic of the Sun. E_E is usually denoted by F_0 , the extraterrestrial solar irradiance. Thus, $F_0 = C/r^2$, and the mean F_0 averaged over the course of a year is

$$\langle F_0 \rangle = C \left\langle \frac{1}{r^2} \right\rangle,$$

so the instantaneous solar irradiance is

$$F_0 = \frac{\langle F_0 \rangle}{r^2 \langle 1/r^2 \rangle},$$

and it varies inversely as the square of the Earth-Sun distance. The mean extraterrestrial spectral solar irradiance $\langle F_0(\lambda) \rangle$ is provided in Figure 8.7.⁵ The irradiance spectrum of the Sun is close to that of a black body at an absolute temperature of approximately 5700 K. The maximum in $\langle F_0(\lambda) \rangle$ is near 500 nm, when the units are power divided by the product of area and wavelength interval, e.g., mW/cm²μm as in the figure.⁶ The sharp features of decreased irradiance are absorption lines (called Fraunhofer lines) caused by elements in the solar atmosphere. For example, the absorption feature at 656 nm is the greatest wavelength in the *Balmer series* of atomic Hydrogen.

The total irradiance from the Sun, i.e.,

$$S \triangleq \int_0^\infty \langle F_0(\lambda) \rangle d\lambda,$$

⁵In Chapter 7 we defined the extra terrestrial solar irradiance, at an Earth-Sun separation of 1 AU, to be \bar{F}_0 . By definition, the semi-major axis to the Earth's orbit, a , is 1 AU. Thus, $r^2 F_0 = a^2 \bar{F}_0$, so

$$\bar{F}_0 = \frac{\langle F_0 \rangle}{a^2 \langle 1/r^2 \rangle} = \langle F_0 \rangle \frac{(1 - \varepsilon^2)^2}{1 + \varepsilon^2/2} \approx \langle F_0 \rangle \left[1 - \frac{5}{2} \varepsilon^2 \right],$$

and therefore $\langle F_0 \rangle$ and \bar{F}_0 differ by less than 0.07%.

⁶Recall that the spectral irradiance is the power falling on a plane sensor ($\Delta \mathcal{P}_\lambda(\lambda)$) within a wavelength interval $\Delta \lambda$ divided by $\Delta \lambda$ and by the area of the surface ΔA , i.e., $\Delta \mathcal{P}_\lambda(\lambda)/(\Delta \lambda \Delta A)$. If we described the separation of the spectrum according to frequency, $\nu = c_\ell/\lambda$, rather than wavelength λ then the irradiance would be $\Delta \mathcal{P}_\nu(\nu)/(\Delta \nu \Delta A)$. But, if we are examining a given spectrum (with a spectrometer at a given setting — wavelength or frequency), the radiant power measured by the instrument is independent of how it will be described, i.e., $\Delta \mathcal{P}_\nu(\nu) = \Delta \mathcal{P}_\lambda(\lambda)$. Therefore, the irradiance spectrum in one representation is related to that in the other representation by

$$\frac{\Delta \mathcal{P}_\nu}{\Delta \nu \Delta A} = \frac{\Delta \mathcal{P}_\lambda}{\Delta \lambda \Delta A} \times \frac{\Delta \lambda}{\Delta \nu} = \frac{\Delta \mathcal{P}_\lambda}{\Delta \lambda \Delta A} \times \frac{\lambda^2}{c_\ell},$$

and the maximum of $\Delta \mathcal{P}_\nu(\nu)/(\Delta \nu \Delta A)$ for the solar spectrum occurs at a wavelength near 880 nm.

where λ is wavelength, is called the *Solar Constant*. Its value is $S \approx 1368 \text{ W/m}^2$ and is stable to within about 0.1%. spectrally, this stability is extant throughout the visible and the infrared; however, in the ultraviolet ($\lambda < 400 \text{ nm}$) the variability increases and can reach 1% between 300 and 400 nm and 5% between 200 and 300 nm. In remote sensing applications, the variability of the mean extraterrestrial solar irradiance is below the sensitivity of most instrumentation.

In order to compute the instantaneous extraterrestrial solar irradiance, we need $\langle 1/r^2 \rangle$. This is given by

$$\left\langle \frac{1}{r^2} \right\rangle = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{r^2} = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta (1 + \varepsilon \cos \theta)^2}{a^2 (1 - \varepsilon^2)^2} = \frac{1 + \varepsilon^2/2}{a^2 (1 - \varepsilon^2)^2}.$$

Then

$$\begin{aligned} F_0 &= \frac{\langle F_0 \rangle}{r^2 \langle 1/r^2 \rangle} = \langle F_0 \rangle \frac{(1 + \varepsilon \cos \theta)^2}{1 + \varepsilon^2/2} \\ &= \frac{1}{1 + \varepsilon^2/2} (1 + 2\varepsilon \cos \theta + \varepsilon^2 \cos^2 \theta) \\ &= \frac{1}{1 + \varepsilon^2/2} \left(1 + \frac{\varepsilon^2}{2} + 2\varepsilon \cos \theta + \frac{\varepsilon^2}{2} \cos 2\theta \right). \end{aligned}$$

Neglecting terms of order ε^3 and higher,

$$F_0 = \langle F_0 \rangle \left(1 + 2\varepsilon \cos \theta + \frac{\varepsilon^2}{2} \cos 2\theta \right). \quad (8.9)$$

The perihelion is on January 3 (actually between January 2 and 5), so if we write this in terms of the day of the year, we must replace θ by $\gamma - \gamma_0$, where $\gamma_0 = 2 \times 2\pi/365$ and $\gamma = 2\pi N/365$, with $N = 0$ on January 1. This introduces sine terms in the above equation. When the value of ε is inserted, an accurate expression for this factor is

$$\begin{aligned} \frac{F_0}{\langle F_0 \rangle} &= 1 + 2\varepsilon \cos \gamma_0 \cos \gamma + 2\varepsilon \sin \gamma_0 \sin \gamma \\ &\quad + \frac{\varepsilon^2}{2} \cos 2\gamma_0 \cos 2\gamma + \frac{\varepsilon^2}{2} \sin 2\gamma_0 \sin 2\gamma. \end{aligned} \quad (8.10)$$

In these equations, the dominate variability in F_0 comes from the terms that are linear in ε . Equation (8.9) shows that the extremes of the deviation of F_0 from the mean are approximately $\pm 2\varepsilon$ or $\sim \pm 3.3\%$.

The Sun subtends an angle of approximately $1/2^\circ$ in the sky.⁷ This means that the radiation from the Sun is not strictly a collimated beam, but is slightly diverging. If

⁷Interestingly, the Moon also subtends the same angle in the sky. This is the reason that total eclipses of the Sun and the Moon exist.

the radiance from the solar disk were uniform across the disk, i.e., L_{Si} in Eq. (8.8) the maximum divergence of the Sun's radiation would be $1/2^\circ$. However, the solar disk is not of uniform radiance. In the visible, the solar radiance is considerably smaller at the edges than at the center of the disk. In fact, at 550 nm, L_{Si} near the edge is roughly 40% of its value at the center. This phenomenon is known as *limb darkening*. It decreases in magnitude into the infrared, but is still significant in spectral regions of interest in remote sensing. Because of solar limb darkening, the effective divergence of the solar beam is actually less than $1/2^\circ$.

8.3 The Moon

All sensors show a radiometric sensitivity loss with time in-orbit. It is critical that this sensitivity decay be accurately assessed. The moon is of interest in ocean remote sensing because it provides a stable source of radiance that can be used to assess the long-term radiometric stability of in-orbit sensors.

8.3.1 Motion of the Moon

The moon moves in an elliptical orbit about the center of mass of the Earth-Moon system (1710 km below the Earth's surface). The mean distance from the center of the Earth at perihelion is about 362,500 km and a aphelion 405,500 km. The time interval required for the Moon to return to the same phase (synodic period) is about 29.53 days, while the time interval required for the Moon to return to the same position relative to the fixed stars is about 27.32 days and is identical to the orbital period. The Moon's rotational period about its own axis and orbital period around the Earth are identical. This is the result of tidal forcing by the Earth. If the Moon's orbit were perfectly circular and the axis of the Moon's rotation were perpendicular to the plane of its orbit, the Moon would always present the same face to the Earth as viewed from the center of the Earth, and the same 50% of the lunar surface would be visible, and 50% invisible. However, as the Moon's orbit is elliptical, conservation of angular momentum requires the moon to move more slowly near its aphelion, more rapidly near the perihelion. Thus near the perihelion, the Moon's rotation is not rapid enough to keep its face exactly pointed toward the Earth. Vice versa, near the aphelion the Moon rotates too rapidly to keep the face perfectly aligned toward Earth. Likewise, because of the tilt of the Moon's axis⁸ it is sometimes possible to observe a portion of the back side at both poles. These effects enable about 59% of the Moon's

⁸The Moon's rotational axis is tilted at an angle 5.145° to the normal to the ecliptic (plane of Earth's orbit around the Sun).

surface to be visible from Earth over the course of time rather than the 50% mentioned above. This of course means that two successive images of the Moon, e.g., at full phase, will view different areas on the Moon near the limbs.

8.3.2 Radiometry of the Moon

The radiance reflected by the lunar surface is specified by its bi-directional reflectance distribution function (BRDF) through

$$L(\hat{\xi}, \lambda) = BRDF(\hat{\xi}_0 \rightarrow \hat{\xi}; \lambda) |\hat{n} \bullet \hat{\xi}_0| F_M(\lambda), \quad (8.11)$$

where $\hat{\xi}_0$ is the propagation direction of the solar irradiance, $\hat{\xi}$ is the propagation direction of the reflected radiance, \hat{n} is the normal to the lunar surface, and $F_M(\lambda)$ is the solar irradiance incident on the lunar surface. If the lunar surface were a lambertian reflector, then $BRDF(\hat{\xi}_0 \rightarrow \hat{\xi}; \lambda) = \mathcal{A}(\lambda)/\pi$, where \mathcal{A} is the surface albedo (reflected irradiance \div incident irradiance). However, were that the case, $L(\hat{\xi}, \lambda) \propto |\hat{n} \bullet \hat{\xi}_0|$ and, as discussed in Chapter 2, the full Moon would become darker near the limbs and would not resemble a disk. This means that $BRDF(\hat{\xi}_0 \rightarrow \hat{\xi}; \lambda)$ must be peaked near $\hat{\xi} = -\hat{\xi}_0$. The angle between $\hat{\xi}$ and $-\hat{\xi}_0$ is called the phase angle: the Full Moon has a phase angle near 0° , and the New Moon a phase angle of 180° . The reflectance of the moon is best described by providing the disk-equivalent albedo of the Full Moon. Imagine the Moon to be a disk oriented with its normal directed toward the Earth and the Sun. The reflected radiance is

$$L(\hat{\xi}, \lambda) = BRDF(-\hat{\xi} \rightarrow \hat{\xi}; \lambda) F_M(\lambda). \quad (8.12)$$

Furthermore, if the surface of the disk is assumed to be lambertian,

$$L(\hat{\xi}, \lambda) = \frac{\mathcal{A}(\lambda)}{\pi} F_M(\lambda). \quad (8.13)$$

The quantity $\mathcal{A}(\lambda)$ is the disk-equivalent albedo. At the position of the Earth, a radiometer viewing any portion of the lunar disk it would measure the radiance $L(\hat{\xi}, \lambda)$. As the disk is not uniform, the radiance is a function of position on the disk, i.e., a function of the selenographic latitude and longitude on the disk. If the disk was viewed from the position of the Earth with a radiometer having a viewing solid angle Ω_r dividing the disk into N equal-area sections, then the total radiance of the Moon defined to be

$$L_{\text{Total}}(\lambda) \triangleq \sum_{i=1}^N L_i(\lambda) = \sum_{i=1}^N \frac{P_i(\lambda)}{A_r \Omega_r} = \frac{E_r(\lambda)}{\Omega_r},$$

where P_i is the power falling on the radiometer sensor, which has an area A_r , and E_r is the irradiance from the whole Moon at the position of the radiometer. Since $L_i(\lambda) =$

$\mathcal{A}_i(\lambda)F_M(\lambda)/\pi$, we have

$$E_r(\lambda) = \frac{\Omega_r F_M(\lambda)}{\pi} \sum_{i=1}^N \mathcal{A}_i(\lambda).$$

But, if Ω_M is the solid angle subtended by the entire Moon at the radiometer, then $\Omega_r = \Omega_M/N$, so

$$E_r(\lambda) = \frac{\Omega_M F_M(\lambda)}{\pi} \frac{\sum_{i=1}^N \mathcal{A}_i(\lambda)}{N} = \frac{\Omega_M F_M(\lambda)}{\pi} \langle \mathcal{A}(\lambda) \rangle, \quad (8.14)$$

where $\langle \mathcal{A}(\lambda) \rangle$ is the spatially averaged disk-equivalent albedo of the Moon. Likewise, if the radiometer viewed the entire disk, the lunar irradiance at the radiometer would be $E_r(\lambda) = \langle L(\hat{\xi}, \lambda) \rangle \Omega_M$, with $\langle L(\hat{\xi}, \lambda) \rangle = \langle \mathcal{A}(\lambda) \rangle F_M(\lambda)/\pi$.

The albedo $\mathcal{A}(\lambda)$ of the lunar Terrae or Highlands is compared to a typical reflectance of the ocean-atmosphere system in Figure 8.8. The Terrae are the areas of the moon that appear to be of about average brightness to the naked eye. The darker areas are called Maria, as they were first thought to be oceans or seas. The brightest area has an albedo of about 0.17 near the center of the visible spectrum. Figure 8.9 provides the probability distribution of lunar albedo. Figure 8.8 shows that the Earth (oceans) and Moon have reflectances that are of the same order of magnitude. A dramatic demonstration of this is provided in Figure 8.10. This, and the fact that the lunar surface is very stable, makes the Moon an ideal source for monitoring the long-term stability of an ocean color sensor. However, such monitoring is complicated by the fact that the radiance of the Moon is a strong function of the phase angle. Figure 8.11 provides the relative radiance of the moon as a function of the phase angle, with the right panel showing that near Full Moon the radiance decreases exponentially with increasing phase angle. The decrease of radiance is due to two effects: (1) as the phase angle increases, less of the lunar surface is illuminated by the Sun, so E_r decreases; and (2) the BRDF decreases as the phase angle increases. Thus, it is essential that the Moon always be viewed by the sensor at the same (or nearly the same) phase angle. Monitoring is also complicated by the motions of the moon. In addition, the reflectance spectrum of the Moon varies with the phase angle, with the blue reflectance decreasing relative to the red as the magnitude of the phase angle increases. This is another reason for viewing the Moon at a constant (or nearly constant) phase angle.

In viewing the Moon, other than the effect of the phase on reflectance, the lunar radiance varies because of variations in the distance from the Moon to the Sun. The solar irradiance falling on the Moon is the same as that falling on the Earth, modified by the distance from the Moon to the Sun. If D_{MS} is the distance from the Moon to the Sun, and D_{ES} the distance from the Earth to the Sun, then the solar irradiance at the surface of the Moon on a plane normal to the solar beam (propagating in the direction $\hat{\xi}_0$), F_M , is given by $F_M = F_0(D_{ES}/D_{MS})^2$, where F_0 is the extraterrestrial solar irradiance received at the Earth. A sensor with a field of view sufficiently small to image a small area on the Moon

and aimed to receive radiation from $\hat{\xi}$ would measure the radiance given by Eq. (8.13).⁹ The albedo is of course a function of position on the lunar surface.

There are two ways that the Moon could be employed to monitor the stability of an orbiting sensor. The first is to simply examine individual pixels in the lunar image. In this case, the radiance observed by the sensor is just that given by Eq. (8.13), where the albedo used in the equation is that of the particular location on the lunar surface. Utilizing this of course requires periodically imaging the same area at exactly the same phase. Having a detailed map of the $\mathcal{A}(\lambda)$ as a function of selenographic latitude and longitude would be required in this application. The second method consists of adding the radiances from all of the pixels, i.e., forming $L_{\text{Total}} = \sum_{i=1}^N L_i$ as above. As the number of pixels imaged (N) is the solid angle subtended by the Moon at the sensor, Ω_M , divided by the field-of-view solid angle of the sensor, Ω_s , which is constant, so $N \propto \Omega_M$. But $\Omega_M = A_M/D_{Ms}^2$, where $A_M = \pi R_M^2$ the projected area of the Moon and therefore the lunar disk, R_M is the lunar radius, and D_{Ms} is the distance from the sensor to the lunar disk. Combining these as in Eq. (8.14) results in

$$E_s = F_0(\lambda) \langle \mathcal{A}(\lambda) \rangle \left(\frac{R_M}{D_{Ms}} \right)^2 \left(\frac{D_{ES}}{D_{MS}} \right)^2$$

The D_{Ms} factor alone in this equation causes the resulting E_s to vary $\pm 10\%$ at a given phase as the Moon moves from aphelion to perihelion in its orbit around the Earth.

Finally, an important fact about the radiance reflected from the Moon is that near full phase, the polarization of the radiance is very small. The degree of polarization is 0 at full phase, $\sim 1\%$ at a phase angle of 10° and 0 near 20° . For larger phase angles, the degree of polarization increases, reaching a maximum of about 7.5% at a phase angle of 90° . Since it is important for sensor stability studies to view the Moon near full phase, it is reasonable to assume that it is a source of totally unpolarized radiance.

8.4 Summary

Here we have provided the basic information related to the Sun and the Moon. Of particular importance to ocean color remote sensing is the determination of the position of the Sun in the sky, Eq. (8.5)-(8.7), along with its mean extraterrestrial solar irradiance in Figure 8.7 and its variation over the year, Eq. (8.10). As the Moon has properties that make it

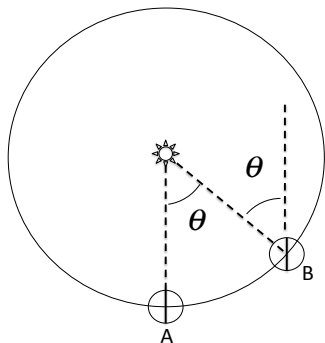
⁹For a typical ocean color sensor with a surface resolution of ~ 1 km from an altitude of ~ 700 km, the solid angle subtended by the sensor is $1 \text{ km}^2 / (700 \text{ km})^2 = 2 \times 10^{-6}$ Sr. When this sensor is aimed at the lunar disk, it will image an area of $\sim 550 \times 550 \text{ km}^2$. The radius of the Moon is about 1740 km, so the area imaged is about 3.2 % of the lunar area. Thus, for such a sensor the area of the Full Moon would be about 31 pixels.

useful to monitor the stability of sensors while they are in orbit, we have provided some limited basic information that should be useful as a starting point for further study of this important application.

8.5 Appendix: The Equation of Time

If one places a rod vertically into the ground (a crude sundial), solar noon occurs when the shadow of the rod is directed toward true North (not magnetic North). This is also the time when the shadow of the rod is shortest. However, solar noon is not in synchrony with mechanical (electrical or atomic) clocks, i.e., at the prime meridian, solar noon does not always occur at exactly the time a mechanical clock reads noon (assuming that they were synchronized at some earlier time).¹⁰ If the orbit of the Earth were circular, and the Earth's rotation axis perpendicular to the plane of the orbit (the ecliptic), then local noon would occur every day at the same time, according to a mechanical clock. So, the fact that the orbit is elliptical and the axis is tilted with respect to the ecliptic must be at the root of the lack of synchrony between solar noon and a mechanical clock.

The Earth rotates 360° on its axis in $23^{\text{hr}}56^{\text{m}}4.0916^{\text{s}} = 23.9344699$ hr. In astronomy this length of time is known as a *sidereal day*. This means that any particular star in the heavens reaches its highest position in the night sky from one night to the next after a this amount of time has passed. Notice that a solar day is defined to be 24 hr, i.e., slightly longer than a sidereal day. Why? The figure below shows the Earth in two positions in its orbit around the Sun. At “A” the observer (at the equator) sees the Sun directly overhead at noon. For the same observer n *sidereal* days later, when the Earth has moved



¹⁰The prime meridian is at longitude 0° . The same statement can be made for other meridians at the center of the various time zones, i.e., $\pm 15^\circ$, $\pm 30^\circ$, etc. If one is not located at one of these meridians, solar noon will occur at a different time, but that time will still not remain in synchrony with mechanical clocks.

through θ in its orbit, has rotated exactly n times, and is located at position “B,” the Sun is not directly overhead: the Earth has to rotate through an additional angle θ before the Sun is directly overhead. Thus, over the course of one year, the Earth must make one more rotation than the number of sidereal days that have passed, i.e., one year takes approximately 366 sidereal days, *not* 365. So, there are 365 solar days, but 366 sidereal days, in one year, and the solar day is $366.2422/365.2422$ longer than the sidereal day. This comes out to 24.00000030 hr for one solar day. The error in defining a solar day to be exactly 24 hr is about 0.4 s per year,¹¹ so that is what we will do. We define the *mean solar day* to be exactly 24 hr, so if the Earth’s orbit were exactly circular and the rotation axis perpendicular to the ecliptic, the Sun would pass over an observer at the prime meridian at intervals of exactly 24 hr. We call this the *mean Sun*. From the point of view of an observer on the Earth, the mean Sun moves in a circle around the Earth given by the same equation as for the actual motion of the Earth around the Sun (but in the opposite sense). Thus, the equation of the mean Sun’s orbit around the Earth is simply $\theta_m = \omega_m t = (360^\circ/365.2422)t = 0.9856473^\circ t$, where θ_m is the angular position of the Earth in its orbit, measured from some arbitrary point on the circular orbit and t is time in days.

The Earth’s orbit is actually elliptical as shown in Figure 8.1. Recalling Eq. (8.1) and the fact that conservation of angular momentum requires $r^2 d\theta/dt$ to be constant, we have

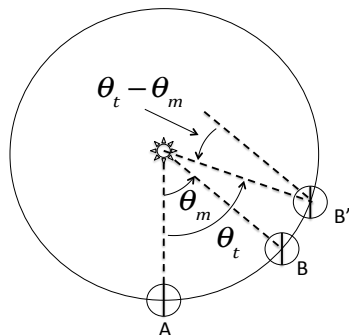
$$\frac{d\theta}{dt} = \frac{K}{r^2} = \frac{K(1 + \varepsilon \cos \theta)^2}{(1 - \varepsilon^2)^2},$$

where K is a constant, showing that the angular speed of the Earth in its orbit is not constant, i.e., the Earth moves faster when it is closer to the Sun and slower when it is farther away. This equation also gives the motion of the true Sun relative to Earth, so we indicate this by replacing θ by θ_t , where the subscript t stands for “true.” Because ε is small (0.0167), we can ignore ε^2 compared to ε itself, so

$$\frac{d\theta_t}{dt} \approx K(1 + 2\varepsilon \cos \theta_t) \quad \text{or} \quad \theta_t \approx K(1 + 2\varepsilon \sin \theta_t).$$

If $\varepsilon = 0$ (circular orbit of the mean Sun) this gives $\theta_m = K$, therefore $\theta_t = \theta_m(1 + 2\varepsilon \sin \theta_t)$. Thus, when the mean Sun moves through an angle θ_m , the true Sun moves through θ_t . The difference is $\theta_t - \theta_m = 2\varepsilon \sin \theta_t$; however, as the difference between θ_t and θ_m is of order ε , $\varepsilon \sin \theta_t$ can be replaced by $\varepsilon \sin \theta_m$ with an error that is of order ε^2 , so we shall make the replacement and take $\theta_t - \theta_m = 2\varepsilon \sin \theta_m$. The effect of this is shown schematically in the figure above which depicts the Earth moving from point “A,” which we take to be

¹¹Defining the solar day to be 24 hr means that as clocks get better the definition of the time unit “second” must change. Time was originally measured in terms of the motion of the Sun and stars in the sky. As astronomical measurements and clocks both got better the definition of the second changed, but it seems that the solar day remained 24 hr.



the position of the point where the Earth is nearest the Sun (perihelion) to the points “B” and “B’”, both of which are several sidereal days later. Position B is where the Earth would be were its orbit circular. In contrast, position B’ is the Earth’s actual position in its elliptical orbit, having progressed further because it is moving faster in this part of the orbit.¹² Because the same number of sidereal days have elapsed for both B and B’ the same number of rotations has taken place as indicated by the vertical line across the planet. Clearly, from the point of view of the Earth observer, the position of the true Sun is farther to the East than that of the mean Sun. Since the mean Sun is synchronized with mechanical clocks, the true Sun will reach its “noon” (highest point in the sky) *at a later time* (according to the mechanical clock) than the mean Sun. (Note that the angle $\theta_t - \theta_m$ is greatly exaggerated in the figure.) To put it another way, the true Sun will be at its highest point in the sky at a longitude Lon_t that is *East* of the mean Sun when the mechanical clock reads noon. If we indicate the longitude at which the true Sun is at its highest point at a given time as Lon_t and similarly the longitude at which the mean Sun is at its highest point at the same time as Lon_m , then clearly this longitude difference is just $\theta_t - \theta_m$, i.e.,

$$Lon_t - Lon_m = -(\theta_t - \theta_m) = -2\varepsilon \sin \theta_m,$$

where the negative sign is applied because the true Sun is farther East, i.e., at a *smaller* longitude. This longitude difference is in radians. To get the longitude difference in degrees, we multiply by $180^\circ/\pi$, and noting that $\varepsilon = 0.0167$,

$$Lon_t - Lon_m = -1.913^\circ \sin \theta_m.$$

Note that $\theta_m = \omega_m t$, where $\omega_m = 2\pi/365.2422 \text{ day}^{-1} = 2\pi/(365.2422 \times 24) \text{ hr}^{-1}$, etc., and t is the elapsed time since the perihelion. If we measure time in days, then $\theta_m = 0$ on

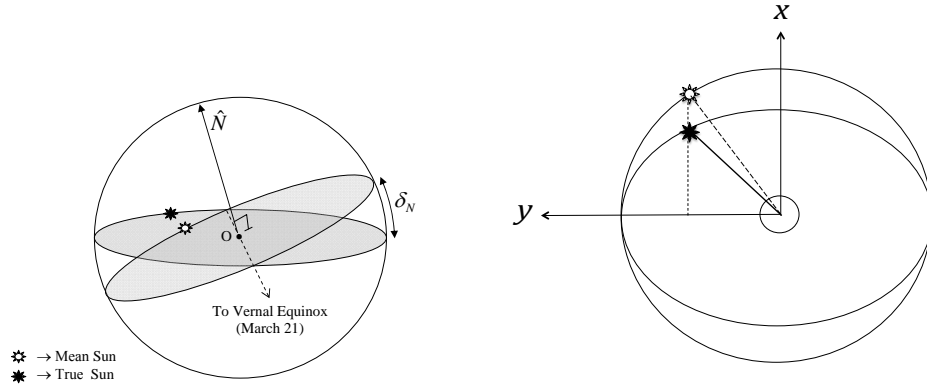
¹²The elliptical orbit and the circular orbit are actually very close together (and are shown here as identical). If one assumes that $\varepsilon \ll 1$, then a circular orbit centered at $x = c_x$ and $y = 0$ can be written in polar coordinates as $r \approx R + c_x \cos \theta$, where R is the radius of the circle. If we expand the equation for the ellipse under the same assumption, we find $r \approx a(1 - \varepsilon \cos \theta)$, so the ellipse at small ε is nearly a circle with radius a centered at $c_x = -\varepsilon a$, i.e., a circle centered slightly off the origin at a point directed away from the perihelion.

January 3, and returns to zero after 365.2422 days, thus

$$\theta_m = \frac{360^\circ}{365.2422}(N - 2) = 0.985647^\circ(N - 2),$$

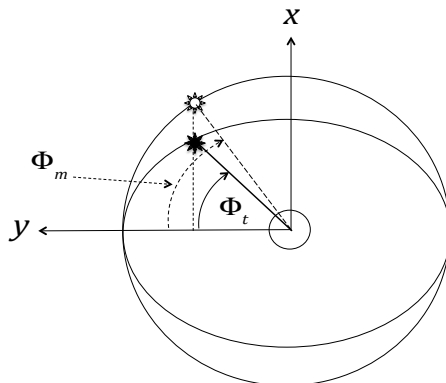
where N is the day of the year *minus* 1, i.e., $N = 0$ on January 1. Thus, for any clock time on any day of the year we can compute $\theta_m = \omega_m t$ and therefore $Lon_t - Lon_m$. The time gives us Lon_m , and therefore the true longitude of the Sun, Lon_t . How do we find Lon_m ? This is easy. By definition at longitude 0° the mean Sun is highest in the sky at noon, which is called 12 o'clock Greenwich Mean Time (GMT). Relative to the mean Sun, the Earth rotates 360° in 24 hr or 15° per hour. Thus, if one reckons time on a mechanical clock in GMT, then on any particular day $Lon_m = 15^\circ(\text{GMT} - 12)$. For example, at 11:30 AM (11.5 hr) GMT, the mean Sun is at $Lon_m = -7.5^\circ$. (Note, here time in GMT must be measured on a 24 hr clock.)

If the Earth's rotation axis were perpendicular to the plane of its orbit, we would be able to compute the longitude of the Sun at any given time from $Lon_t - Lon_m = -1.913^\circ \sin \theta_m$, but it is not. The tilt of the orbit with respect to the ecliptic also has an effect on the position of the Sun. To illustrate this it is useful to take the point of view of an observer on the Earth. Accordingly, the mean Sun is assumed to revolve around the Earth in a circular orbit once per year. The Earth rotates on its axis once per sidereal day. The true Sun similarly revolves around the Earth but in an orbit that is tilted with respect to the mean sun through an angle δ_N . This state of affairs is shown schematically in the drawing



above (left) of the celestial sphere, depicting the mean (tilted circle in the figure) orbit of the Sun around the Earth from the point of view of an observer “O” located on the Earth. The unit vector \hat{N} is normal to the the plane of the orbit of the mean Sun. The horizontal circle is the path of the actual Sun, which moves in the same plane as the Earth's orbit around the Sun (the ecliptic). The Earth's rotation axis (\hat{N}) is tilted with respect to the ecliptic through an angle δ_N . The summer solstice is on the left side of the drawing.

The drawing on the right is the view of the situation looking down toward the North pole of the Earth, where we see the mean Sun moving in a circular orbit and the true Sun moving in what we shall see is an elliptical orbit when projected on to the x - y plane. The equation of the orbit of the mean Sun is simply $x_m^2 + y_m^2 = R^2$, where R is the radius of the orbit. The equation of the true Sun's orbit is $x_t^2 + y_t^2 + z_t^2 = R^2$. Clearly, $z_t = x_t \tan \delta_N$, so $(1 + \tan^2 \delta_N)x_t^2 + y_t^2 = R^2$, the equation of an ellipse. The true and mean Suns are coincident at the vernal equinox ($x = 0$) and also at the autumnal equinox. They both move at the same uniform rate in their individual (circular) orbits of one revolution per year (365.2422 days) and, since the planes of the two orbits intersect along the y axis, both Suns must always have the same value of y as they go along their individual orbits, i.e., $y_t = y_m$, so $x_t = x_m \cos \delta_N$. During their traverse, the Earth rotates once every sidereal day in a counter clockwise direction. At some time between the beginning of Spring (vernal equinox) and Summer (summer solstice) the positions of the two Suns (mean and actual) are as shown on the right-hand figure. Clearly, in the position shown the true Sun is located at a more Easterly longitude than the mean Sun, i.e., when the mechanical clock reads noon (solar noon for the mean Sun), the true Sun will be to the East and will be highest in the sky at the observer's position some time after solar noon. In the drawing below,



we have labeled the Φ_t and Φ_m , the angles through which the true and mean Suns have progressed from the vernal equinox toward summer. Clearly, $Lon_t - Lon_m = \Phi_t - \Phi_m$, so to is necessary to relate these angles. We note that

$$\tan \Phi_m = \frac{x_m}{y_m} \quad \text{and} \quad \tan \Phi_t = \frac{x_t}{y_t},$$

But $y_t = y_m$ and $x_t = x_m \cos \delta_N$, so

$$\tan \Phi_t = \frac{x_m \cos \delta_N}{y_m} = \cos \delta_N \tan \Phi_m$$

This relationship, $\tan \Phi_t = \cos \delta_N \tan \Phi_m$, provides Φ_t as a function of Φ_m , which allows the longitude difference to be determined by Φ_m , i.e., by a mechanical clock. With an error

of less than 0.15° this can be fit to $\Phi_t - \Phi_m = -2.476^\circ \sin 2\Phi_m$, so

$$Lon_t - Lon_m = -2.476^\circ \sin 2\Phi_m$$

provides the longitude difference due to the tilt of the Earth's orbit to the ecliptic. Combining the two effects, we have

$$Lon_t - Lon_m = -1.913^\circ \sin \theta_m - 2.476^\circ \sin 2\Phi_m.$$

Recall that θ_m is the position of the Earth in its orbit measured from the perihelion (January 3), while Φ_m is measured from the vernal equinox (March 21). We need to write these in terms of the same reference time. We choose the reference to be January 1, and define the angle γ to be θ_m measured from January 1 rather than January 3, then $\theta_m = \gamma - 2 \times 0.9856^\circ$. Likewise since March 21 is the 80th day of the year, $\Phi_m = 80 \times 0.9856^\circ - \gamma$.¹³ Combining,

$$Lon_t - Lon_m = -1.912^\circ \sin \gamma + 0.065^\circ \cos \gamma - 2.291^\circ \sin 2\gamma - 0.939^\circ \cos 2\gamma.$$

From this equation we find that the maximum longitude difference is about $+4.15^\circ$ near $\gamma = 300^\circ$.

The equation in the text developed by Fourier analysis of actual observations of the Sun was (in the same units and using the notation here)

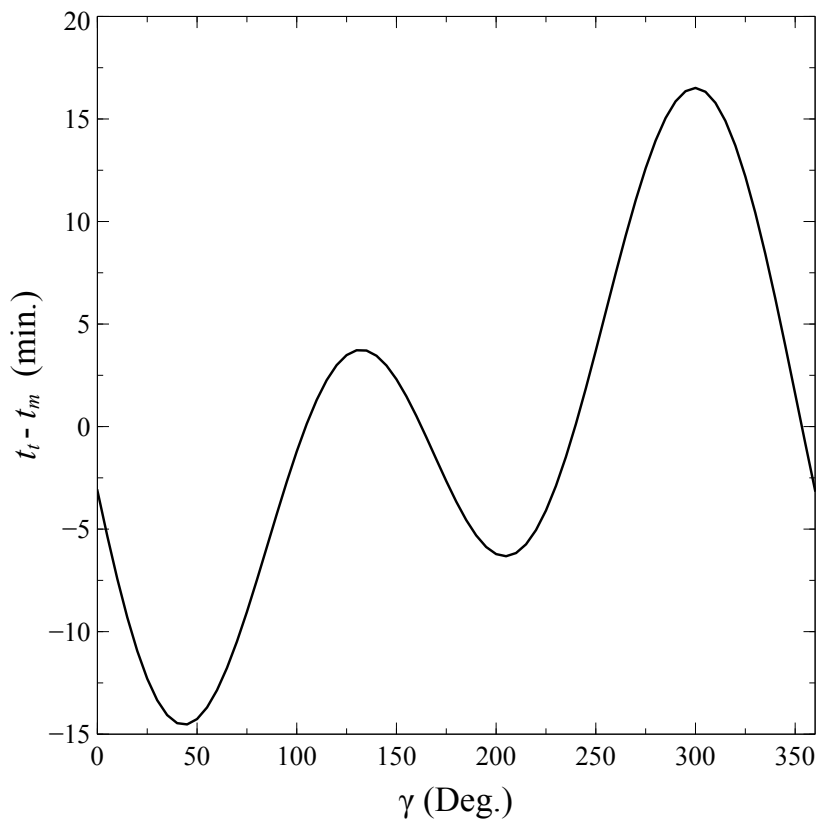
$$Lon_t - Lon_m = -1.840^\circ \sin \gamma + 0.107^\circ \cos \gamma - 2.340^\circ \sin 2\gamma - 0.837^\circ \cos 2\gamma + 0.00043,$$

which agrees well with our derived equation. It should be noted that our derivation is approximate. For example, the orbits of the mean and true Suns in the tilt contribution are really elliptical, and the effects of the motion of the Moon are ignored.

For a fixed observer, the longitude difference is manifest as a time difference, i.e., the difference between when the Sun is supposed to be at a given longitude, t_m (say the observer's meridian) and when it actually is at that longitude, t_t . Since the Earth rotates 15° per hr or 0.25° per min, dividing the longitude difference by 0.25° gives the time difference in minutes. From our derived result, we find for the time difference in minutes to be

$$t_t - t_m = -7.648 \sin \gamma + 0.260 \cos \gamma - 9.164 \sin 2\gamma - 3.751 \cos 2\gamma.$$

Note that, when $t_t - t_m$ is negative, it means that the mean Sun arrives at the given meridian of longitude *earlier* than the true Sun, i.e., the actual Sun arrives at a *later* time than the mean Sun. Thus, when the mechanical clock reads noon at the prime meridian (0° longitude) the Sun arrives there when the mechanical clock reads noon + $(t_m - t_t)$, i.e., a sundial at this location would be $t_m - t_t$ minutes *slow*. The figure below is a plot of the



variation of $t_t - t_m$ over the course of one year. The quantity $t_t - t_m$ is usually called the equation of time. It remains only to note that the angle γ is simply

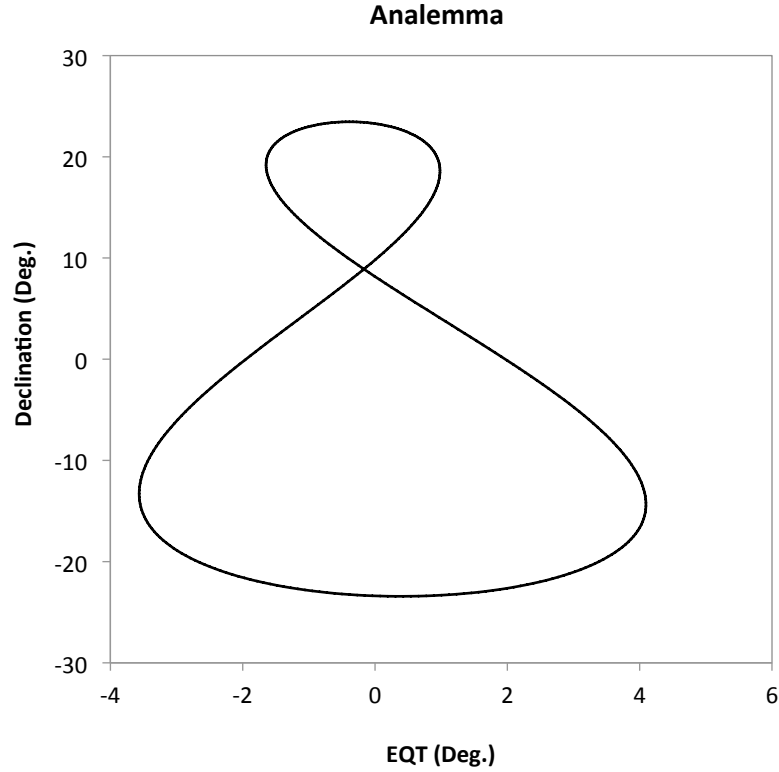
$$\gamma = \frac{2\pi N}{365.2422},$$

where N is the day of the year with $N = 0$ on January 1.

On older globes of the Earth one often sees a vertical “figure 8-like” object close to the Equator drawn in the Pacific Ocean. This object is called the *analemma*. It is simply a plot of the declination of the Sun (δ) as a function of the equation of time (*EQT*) as shown in the figure below. It represents the position of the Sun in the sky at a mean solar time of noon. If you were at the Equator in a room with a small hole in ceiling (to let in a beam of light from the Sun), and at noon (by a mechanical clock) each day, you marked the position of the Sun’s image on the floor, the resulting locus of positions would

¹³One might think that $\Phi_m = \gamma - 80 \times 0.9856^\circ$; however, note that the Sun moves in a direction opposite to the Earth in its orbit. Note 0.9856° is the angular distance the Earth moves in one day.

be similar to the curve in the figure. Thus, you would have discovered the effects of both declination and the equation of time. In addition, on the 366th day of this exercise you will have discovered the need for leap year.



8.6 Bibliographic Notes

8.2.1 Motion of the Earth relative to the Sun and Fixed Stars

The orbital motion of the Earth is developed in almost all textbooks on intermediate Newtonian mechanics. A good one is [Barger and Olssen \[1995\]](#). The derivation of Eq. (8.2) is our own, and the Fourier representation of δ is from [Spencer \[1971\]](#).

8.2.2 Position of the Sun in the Sky

Basic data on the Earth's rotation can be found in any textbook on astronomy. A

good derivation of the equation of time is provided by [Muller \[1995\]](#). The Fourier series representation of the equation of time and also the solar declination was developed by [Spencer \[1971\]](#).

8.2.3 Extraterrestrial Solar Irradiance

The Extraterrestrial Solar Irradiance is from [Wehrli \[1985\]](#). Measurements of the limb darkening of the Sun across the spectrum are found in [Pierce et al. \[1950\]](#).

8.3.1 Motion of the Moon

The material in the section can be found in almost any book on astronomy or astrophysics.

8.3.2 Radiometry of the Moon

The lunar spectral albedo (Figure 8.8) was developed from data provided in [Fessenkov \[1962\]](#). It agrees well with the measurements of [Lane and Irvine \[1973\]](#) for the full lunar disk.

The statistical distribution of the lunar albedo (Figure 8.9) was computed from data presented in [Minnaert \[1962\]](#) based on measurements made by [Sytinskaya \[1953\]](#). They were measured with the “V” (visual) optical filter with maximum transmittance at 550 nm and a full width at half maximum transmittance of ~ 89 nm. This filter approximates the spectral sensitivity of the human eye in full daylight.

The lunar phase curve (Figure 8.11) was measured by [Rougier \[1934\]](#) and the data presented were taken from [Fessenkov \[1962\]](#).

A detailed model of the spectral albedo of the Moon is described in [Kieffer and Stone \[2005\]](#). The use of the Moon as a stability monitor for ocean color systems appears to have been first proposed by [Kieffer and Wildey \[1985\]](#) and independently by [Gordon \[1987\]](#). The the initial application to the SeaWiFS sensor is provided in [Barnes et al. \[1999\]](#). A detailed comparison between SeaWiFS measurements of the Moon’s irradiance with the model of [Kieffer and Stone \[2005\]](#) is described in [Barnes et al. \[2004\]](#).

8.7 Figures

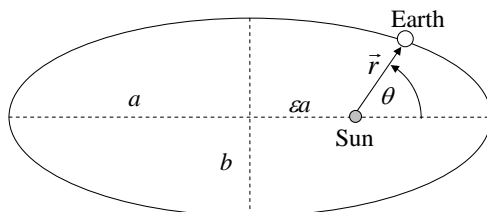


Figure 8.1: Geometry of the Earth's orbit around the Sun.

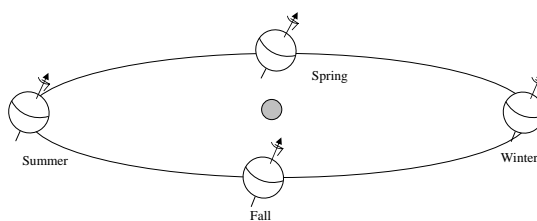


Figure 8.2: Schematic of the Earth's motion around the Sun including the Earth's rotation.

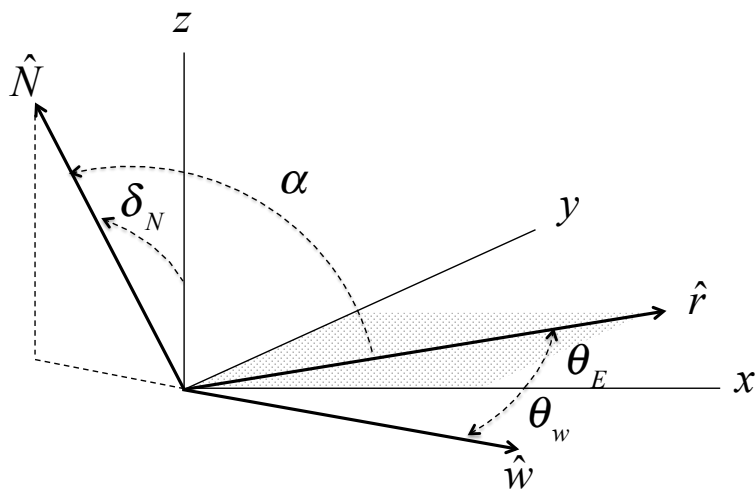


Figure 8.3: Geometry for determination of the declination of the Sun. (See text for detailed definitions of the symbols.)

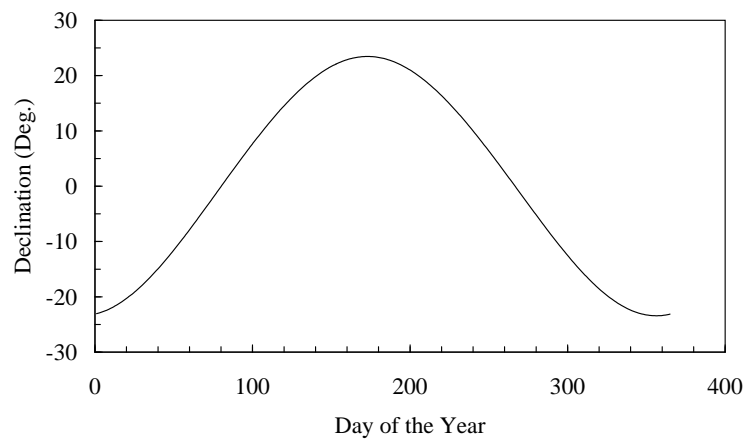


Figure 8.4: Declination of the Sun as a function of day of the year.

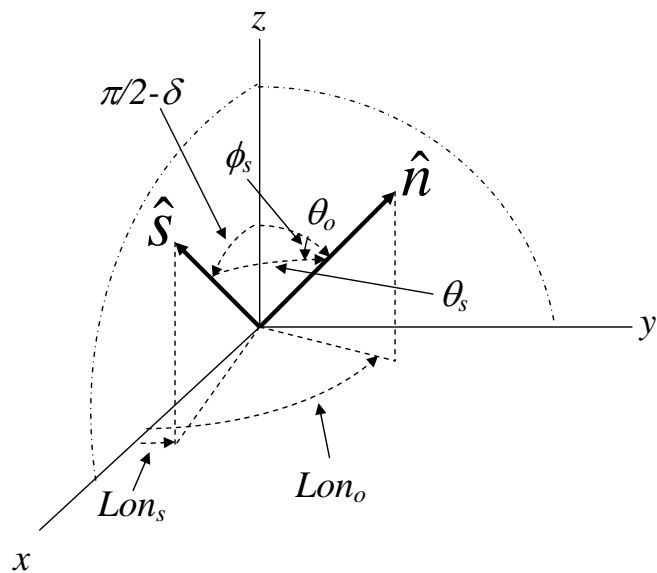


Figure 8.5: Geometry for determination of the solar zenith angle θ_s and solar azimuth angle ϕ_s . (See text for definitions of all the other angles.)

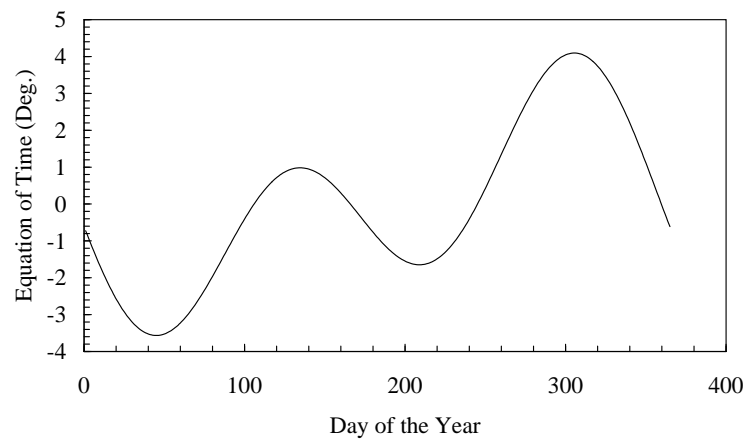


Figure 8.6: Value of $EQT = Lon_s - \langle Lon_s \rangle$ in degrees as a function of the day of the year.

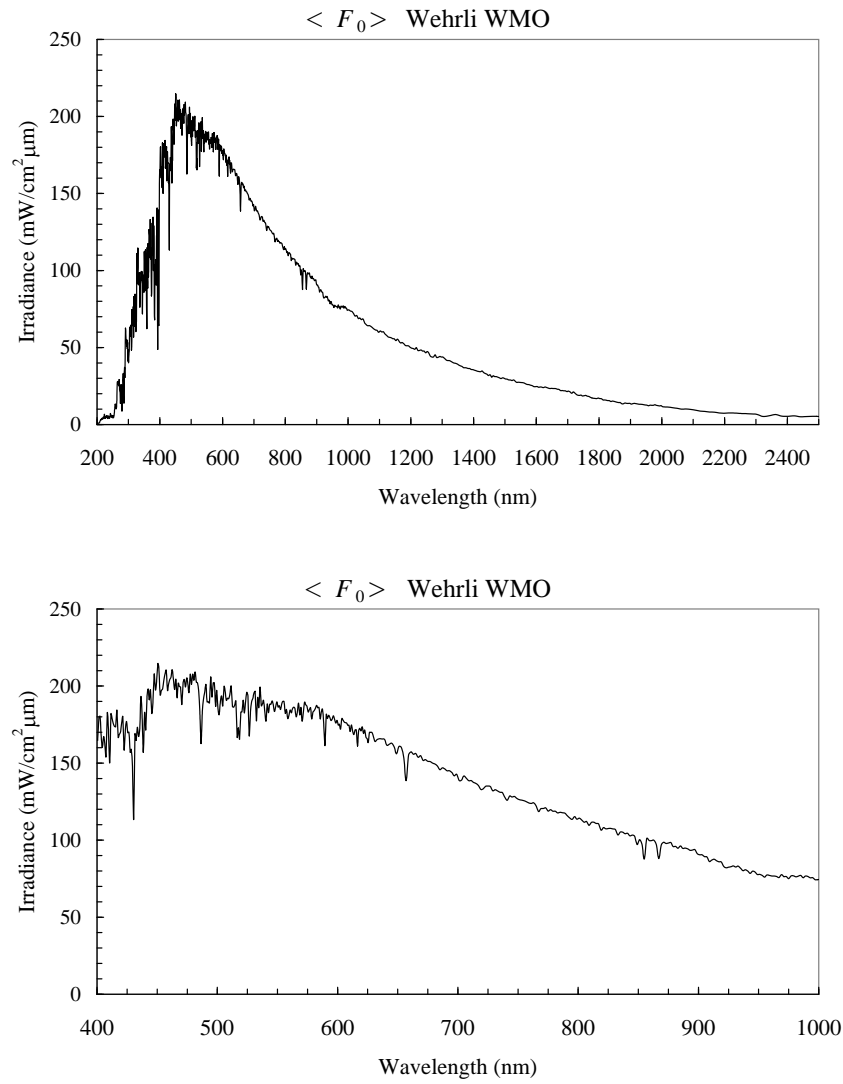


Figure 8.7: The spectrum of mean extraterrestrial solar irradiance $\langle F_0 \rangle$.

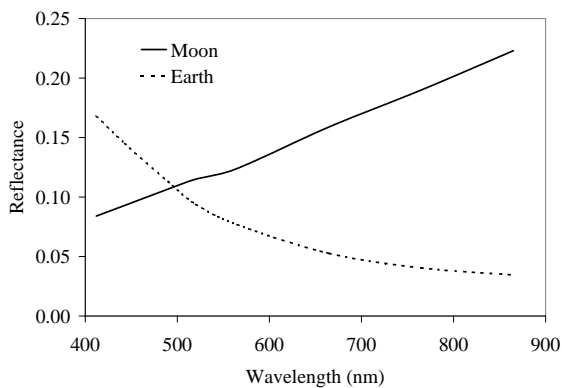


Figure 8.8: Reflectance (actually $\mathcal{A}(\lambda)$) of the bright areas (full phase) of the lunar surface (Terraes or Highlands) compared with the typical reflectance of the ocean-atmosphere system which is defined as $\pi L/F_0 \cos \theta_0$, where L is the reflected radiance, θ_0 the solar zenith angle, and F_0 the extraterrestrial solar irradiance. Here, the ocean-atmosphere system is viewed at nadir. If the ocean-atmosphere system were a lambertian reflector, the reflectance shown here would be its albedo.

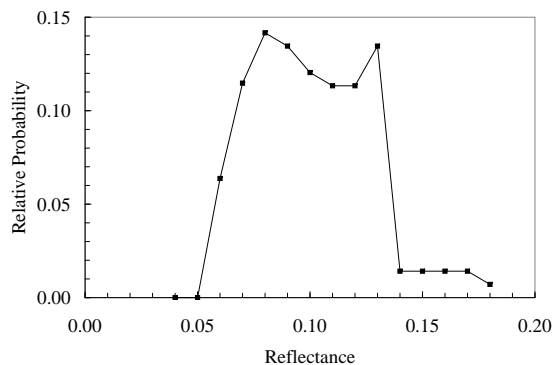


Figure 8.9: Probability of a given reflectance of the lunar surface near full phase. The wavelength corresponds to the “V” (visible) filter of astronomy, centered at 550 nm with a full width at half maximum of 89 nm.



Figure 8.10: A visible GOES image containing both the Earth and the Moon. Note that the non-cloudy areas of the Earth are close in brightness to the Moon.

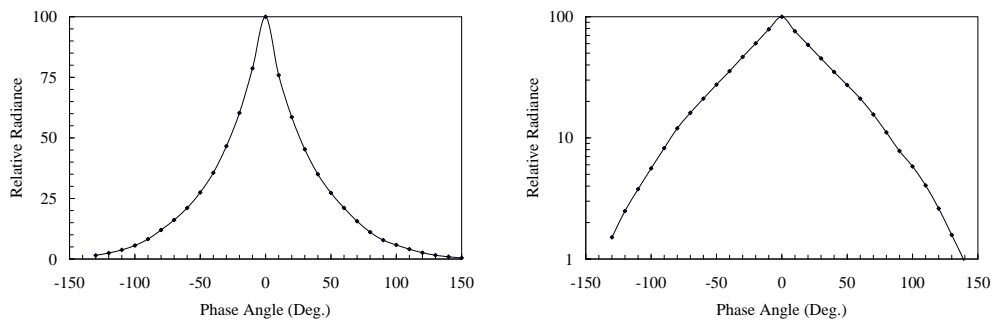


Figure 8.11: The radiance (relative) of the moon as a function of the phase angle in blue (440 nm) light. Note that the point corresponding to 0° is extrapolated from non-zero phase angles. The radiance of the Moon near 0° becomes much larger than shown here, a phenomenon known as the lunar surge or opposition effect.

Part III

Remote Sensing Ocean Color

Chapter 9

The Remote Sensing Problem

9.1 Introduction

In this chapter we examine the general problem of remote sensing water properties by measuring the upwelling radiance (in the visible) over a water body using airborne or space borne radiometers. One goal of this chapter is to provide the reader with an understanding of the various issues that must be faced to extract the water properties from the measured radiance. Another goal is to provide a preliminary estimate of the magnitude of the quantities involved. For simplicity, the single-scattering approximation¹ is used for radiative transfer in the atmosphere because it captures all of the essential physics of the problem and provides analytic formulas for most of the quantities involved.

We begin the discussion by describing the various components of the measured radiance through the application of single-scattering radiative transfer for a simple model of the atmosphere and the water surface. This will provide analytical formulas for the components and will be used later to suggest a path for retrieving the water properties as discussed in the next chapter. Next the range of validity of the single-scattering theory is examined. Finally, we show how to compute the *normalized* water-leaving radiance and reflectance — the quantities that are actually used in the bio-optical algorithms to retrieve constituents such as pigment concentration and Chlorophyll *a* concentration. In an appendix we examine

¹Following the practice we use throughout this work, the term “first-order solution” to the radiative transfer equation is reserved for the first-order term in the successive-order of scattering solution method. That term contains exponentials, that when expanded and the second and higher order terms are deleted (i.e., only *one* interaction with the medium is retained), results in what we call the “single-scattering approximation.”

the polarization properties of the radiance exiting the top of the atmosphere over the ocean. The polarization of the exiting radiance is important, as most sensors are somewhat sensitive to the polarization state of the radiance they measure, and corrections are required to account for this sensitivity.

9.2 The Remote Sensing Problem

Radiance in the visible portion of the spectrum (light) from the Sun and sky that enters a water body and is then backscattered *out* through the surface contains information regarding the water's constituents and their vertical distribution. This radiance exiting the water surface is called the “water-leaving radiance,” and is denoted by $L_w(\hat{\xi})$, $L_w(\theta, \phi)$, or $L_w(u, \phi)$. In general, $L_w(\hat{\xi})$ is nonzero for all $\hat{\xi}$ directed into the upward hemisphere. In remote sensing, this radiance contributes one component of the radiance measured by a sensor viewing the water from above. A radiometer at a height z above the sea surface aimed so as to detect radiance propagating in the direction $\hat{\xi}$ will measure a radiance $L_t(z, \hat{\xi})$ that is given by

$$L_t(z, \hat{\xi}) = L_{\text{Other}}(z, \hat{\xi}) + L_w(z, \hat{\xi}),$$

where $L_{\text{Other}}(z, \hat{\xi})$ is the totality of radiance at the sensor that was *not* backscattered *out* of the water. It results from photons that have been scattered in the atmosphere or reflected from the sea surface (before and/or after scattering in the atmosphere) into the radiometer. $L_w(z, \hat{\xi})$ is the contribution to $L_t(z, \hat{\xi})$ from photons that *were* backscattered *out* of the water. It consists of the photons that exit the water in the direction $\hat{\xi}$ (denoted by $L_w(\hat{\xi})$) and survive propagation from the surface to the radiometer, as well as photons that exit the ocean in a direction other than $\hat{\xi}$, but are redirected by scattering in the atmosphere into the radiometer. It is traditional to write

$$L_w(z, \hat{\xi}) = t(z, \hat{\xi})L_w(\hat{\xi}),$$

where $t(z, \hat{\xi})$ is called the *diffuse transmittance* of the atmosphere between z and the surface in the direction $\hat{\xi}$; however, we shall see that $t(z, \hat{\xi})$ is not a property of the atmosphere alone, as it depends on $L_w(\hat{\xi})$ as well.

9.3 Single-Scattering Analysis

To provide a firm foundation for understanding the various processes that influence $L_t(z, \hat{\xi})$ we will apply the single-scattering solution for radiative transfer to a simplified version of

the remote sensing problem. Consider a *homogeneous slab* atmosphere of thickness z_0 bounded on the bottom by a *flat*, horizontally homogeneous, water body. As in Appendix 1 to Chapter 2, we place the origin of coordinates at the top of the atmosphere with the z -axis directed toward the water. The top of the slab is illuminated by the solar beam propagating in the direction $\hat{\xi}_0$. The radiance associated with this beam is

$$F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0) = F_0 \delta(u - u_0) \delta(\phi - \phi_0)$$

where F_0 is the extraterrestrial solar irradiance. The remote radiometer is located at the top of the atmosphere ($z = 0$) measuring a radiance $L_t(0, \hat{\xi})$. The water-leaving radiance entering the bottom of the atmosphere is $L_w(z_0, \hat{\xi})$. For this problem

$$\begin{aligned} L_t(0, \hat{\xi}) &= L_{\text{Other}}(0, \hat{\xi}) + L_w(0, \hat{\xi}) \\ &= L_{\text{Other}}(0, \hat{\xi}) + t(z_0 \rightarrow 0, \hat{\xi}) L_w(z_0, \hat{\xi}), \end{aligned} \tag{9.1}$$

where $t(z_0 \rightarrow 0, \hat{\xi})$ is the diffuse transmittance from the surface to the top of the atmosphere. Thus, to understand how to extract the useful information, $L_w(z_0, \hat{\xi})$, from $L_t(0, \hat{\xi})$, we need to understand how to determine $L_{\text{Other}}(0, \hat{\xi})$ and $t(z_0 \rightarrow 0, \hat{\xi})$. Recall that $L_{\text{Other}}(0, \hat{\xi})$ consists of radiance from the Sun and sky scattered by the atmosphere (air molecules and aerosols) into the radiometer, radiance from the Sun and sky reflected by the water surface into the radiometer, and radiance from any whitecaps present on the sea surface, while $t(z_0 \rightarrow 0, \hat{\xi})$ accounts for the attenuation of $L_w(z_0, \hat{\xi})$ as it propagates to $z = 0$, as well as atmospheric scattering of $L_w(z_0, \hat{\xi}')$ (where $\hat{\xi}'$ is simply a direction different from $\hat{\xi}$) into $\hat{\xi}$ followed by propagation to the radiometer. Although the solar beam is the ultimate source for $L_w(z_0, \hat{\xi}')$, to carry out the single-scattering analysis, we will solve the radiative transfer equation for a slab illuminated by *two* sources: the direct solar beam illuminating the top of the atmosphere leading to $L_{\text{Other}}(0, \hat{\xi})$; and the water-leaving radiance $L_w(z_0, \hat{\xi})$ illuminating the bottom of the atmosphere leading to $L_w(0, \hat{\xi})$. These sources can be treated separately by virtue of the linearity of the radiative transfer equation.

9.3.1 Contribution of Atmospheric Scattering to $L_{\text{Other}}(0, \hat{\xi})$.

There are three processes involving a single interaction (scattering) within the atmosphere that can contribute to $L_{\text{Other}}(0, \hat{\xi})$: (1) backscattering of direct solar radiation into the radiometer; (2) forward scattering of solar radiation toward the water surface (sky radiance), followed by specular (Fresnel) reflection from the surface and subsequent propagation to the sensor; and (3) specular reflection of direct solar radiation from the sea surface followed by forward scattering into the direction $\hat{\xi}$ and propagation to the sensor. These processes are shown schematically in Figure 9.1.

For process number 1, the formulas developed in Appendix 1 to Chapter 2 are directly applicable. In particular, Eq. (2.102) states that this component (L_1) is given by

$$L_1(0, u, \phi) = -\frac{\omega_0 \tau_0 F_0}{4\pi u} P(u_0 \rightarrow u, \phi_0 \rightarrow \phi), \quad u < 0,$$

or

$$L_1(0, \hat{\xi}) = \frac{\omega_0 \tau_0 F_0}{4\pi |\hat{\xi} \bullet \hat{e}_z|} P(\hat{\xi}_0 \rightarrow \hat{\xi}), \quad (9.2)$$

where P is the scattering phase function, ω_0 is the single scattering albedo, and τ_0 is the optical depth of the atmosphere (cz_0).

To compute the contribution from process number 2, we note that Eq. (2.103) in the Appendix to Chapter 2 provides the radiance at the surface (the sky radiance), i.e.,

$$L(\tau_0, u', \phi) = F_0 \delta(u' - u_0) \delta(\phi' - \phi_0) \left[1 - \frac{\tau_0}{u'} \right] + \frac{\omega_0 F_0 \tau_0}{4\pi u'} P(u_0 \rightarrow u', \phi_0 \rightarrow \phi'), \quad u' > 0,$$

This is Fresnel-reflected from the flat surface with reflectance $r_f(u')$ and propagated unchanged to the top of the atmosphere.² If $\hat{\xi}' = \hat{\xi}_0$ the second term can be ignored because it is very small compared to the first (Sun glint, Chapter 7); however, because the first term is so large, in any real remote sensing situation the geometric setting in which $\hat{\xi}' = \hat{\xi}_0$ must be avoided. Thus,

$$L_2(0, u, \phi) = \frac{\omega_0 F_0 \tau_0}{4\pi u'} r_f(u') P(u_0 \rightarrow u', \phi_0 \rightarrow \phi') + r_f(u_0) F_0 \delta(u' - u_0) \delta(\phi' - \phi_0) \exp[-2\tau_0/u_0], \quad u' > 0$$

or

$$L_2(0, \hat{\xi}) = \frac{\omega_0 F_0 \tau_0}{4\pi |\hat{\xi} \bullet \hat{e}_z|} r_f(\hat{\xi}') P(\hat{\xi}_0 \rightarrow \hat{\xi}') + r_f(\hat{\xi}_0) F_0 \delta^{(2)}(\hat{\xi}' - \hat{\xi}_0) \exp[-2\tau_0/|\hat{\xi}_0 \bullet \hat{e}_z|], \quad (9.3)$$

where, $u' = |u|$, $\phi' = \phi$, $\hat{\xi}' = \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z)\hat{e}_z$, and we have replaced the single-scattering equation for the directly reflected solar beam by its exact counterpart.

Finally, for process 3, we recognize that this is equivalent to illuminating the atmosphere from the bottom with a radiance

$$r_f(\hat{\xi}_0) F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}'') = r_f(\hat{\xi}_0) F_0 \delta(u - u'') \delta(\phi - \phi''),$$

²Note that r_f depends only on the angle of incidence with the surface, $\cos^{-1} u'$. Thus, we use the notation $r_f(u')$, or equivalently $r_f(\hat{\xi}')$, where u' and $\hat{\xi}'$ are related by $u' = \hat{\xi}' \bullet \hat{e}_z$. In addition, since $u = -u'$ (the law of reflection), $r_f(u') = r_f(|u|)$.

where $\hat{\xi}'' = \hat{\xi}_0 - 2(\hat{\xi}_0 \bullet \hat{e}_z)\hat{e}_z$. A straightforward computation similar to that given in Appendix 1 to Chapter 2 yields,

$$\begin{aligned} L_3(0, u, \phi) &= -\frac{\omega_0 F_0 \tau_0}{4\pi u} r_f(u_0) P(u'' \rightarrow u, \phi'' \rightarrow \phi), \quad u < 0, \\ \text{or} \\ L_3(0, \hat{\xi}) &= \frac{\omega_0 F_0 \tau_0}{4\pi |\hat{\xi} \bullet \hat{e}_z|} r_f(\hat{\xi}_0) P(\hat{\xi}'' \rightarrow \hat{\xi}). \end{aligned} \quad (9.4)$$

Combining, the radiance due to these three processes exiting the top of the atmosphere is given by

$$L(0, \hat{\xi}) = \sum_{i=1}^3 L_i(0, \hat{\xi}).$$

The resulting formulas can be simplified in the case where the scattering phase function depends only on the angle between the incident and the scattered beams, i.e., the scattering angle Θ . We note that the scattering angles for processes 2 and 3 are given by, respectively

$$\begin{aligned} \cos \Theta_2 &= \hat{\xi}_0 \bullet \hat{\xi}' \\ &= \hat{\xi}_0 \bullet \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z)\hat{\xi}_0 \bullet \hat{e}_z, \end{aligned}$$

and

$$\begin{aligned} \cos \Theta_3 &= \hat{\xi}'' \bullet \hat{\xi} \\ &= \hat{\xi} \bullet \hat{\xi}_0 - 2(\hat{\xi}_0 \bullet \hat{e}_z)\hat{\xi} \bullet \hat{e}_z. \end{aligned}$$

Thus, we see that $\Theta_2 = \Theta_3$ and the scattering angles associated with processes 2 and 3 are identical. Following the notation of the earlier literature, we denote the scattering angle associated with process 1, i.e., $\cos^{-1}(\hat{\xi}_0 \bullet \hat{\xi})$, by Θ_- , and the angle associated with processes 2 and 3 by $\Theta_2 = \Theta_3 \triangleq \Theta_+$. Then these three components can be combined to yield L_{Other} exiting the top of the atmosphere, propagating in a direction $\hat{\xi}$:

$$\begin{aligned} L_{\text{Other}}(0, u, \phi) &= \frac{\omega_0 \tau_0 F_0}{4\pi |u|} \left[P(\Theta_-) + \left(r_f(u') + r_f(u_0) \right) P(\Theta_+) \right] \\ &\quad + r_f(u_0) F_0 \delta(u' - u_0) \delta(\phi' - \phi_0) \exp[-2\tau_0/u_0], \quad \text{or} \\ L_{\text{Other}}(0, \hat{\xi}) &= \frac{\omega_0 \tau_0 F_0}{4\pi |\hat{\xi} \bullet \hat{e}_z|} \left[P(\hat{\xi}_0 \rightarrow \hat{\xi}) + \left(r_f(\hat{\xi}') + r_f(\hat{\xi}_0) \right) P(\hat{\xi}_0 \rightarrow \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z)\hat{e}_z) \right] \\ &\quad + r_f(\hat{\xi}_0) F_0 \delta^{(2)}(\hat{\xi}' - \hat{\xi}_0) \exp[-2\tau_0/|\hat{\xi}_0 \bullet \hat{e}_z|], \end{aligned} \quad (9.5)$$

where $u' = |u|$, $\phi' = \phi$, $\hat{\xi}' = \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z)\hat{e}_z$, and we have ignored any contribution from whitecaps.³

In general the scattering in the atmosphere results from scattering by air molecules (Rayleigh scattering) and scattering by various species of aerosols. Consider a homogeneous slab with N species of scatterers. The attenuation coefficient c can be written

$$c = \sum_{i=1}^N c_i,$$

where c_i is the attenuation coefficient of the i^{th} species. Since $\tau_0 = cz_0$, we also have

$$\tau_0 = \sum_{i=1}^N \tau_{0i},$$

with $\tau_{0i} = c_i z_0$. Likewise,

$$\beta = \sum_{i=1}^N \beta_i \quad \text{and} \quad b = \sum_{i=1}^N b_i,$$

for the volume scattering function and the total scattering coefficient, respectively. The various components of L_{Other} (except for the term with the Dirac delta functions, which

³Equation (9.5) provides L_{Other} when the sensor is located at the top of the atmosphere, i.e., a space-borne sensor. In the interest of completeness, we provide the associated formulas for the case of a sensor viewing the water from *within* the atmosphere, e.g., an aircraft-borne sensor. Let the sensor be located at a depth z within the slab, where $0 < z < z_0$. Figure 9.1 shows that for process 1, only that portion of the atmosphere between τ and τ_0 can contribute. Since L_1 is a linear function of optical depth, the term analogous to L_1 is given by Eq. (9.2) with τ_0 replaced by the optical thickness of the atmosphere between the surface and the sensor, i.e., $\tau_0 - \tau$. A similar conclusion is reached for process 3, i.e., Eq. (9.4) still applies with τ_0 replaced by $\tau_0 - \tau$. In contrast, for process 2 the altitude of the sensor is irrelevant as the radiance is scattered by the entire atmosphere (once) *before* reflecting from the surface, and there are no additional interactions in the atmosphere (single scattering). Thus, Eq. (9.3) provides the correct radiance as it stands for process 2, except that the attenuation factor for the exact Fresnel-reflected solar beam, $\exp[-2\tau_0/|\hat{\xi}_0 \bullet \hat{e}_z|]$, must be replaced by $\exp[-(2\tau_0 - \tau)/|\hat{\xi}_0 \bullet \hat{e}_z|]$ to account for the shortened path to the sensor. Combining these observations provides L_{Other} propagating in an upward direction $\hat{\xi}$ at an optical depth τ within the atmosphere:

$$\begin{aligned} L_{\text{Other}}(\tau, u, \phi) &= \frac{\omega_0(\tau_0 - \tau)F_0}{4\pi|u|} \left[P(\Theta_-) + \left(\frac{\tau_0}{\tau_0 - \tau} r_f(u') + r_f(u_0) \right) P(\Theta_+) \right] \\ &\quad + r_f(u)F_0\delta(u' - u_0)\delta(\phi' - \phi_0)\exp[-(2\tau_0 - \tau)/u_0], \quad \text{or} \\ L_{\text{Other}}(\tau, \hat{\xi}) &= \frac{\omega_0(\tau_0 - \tau)F_0}{4\pi|\hat{\xi} \bullet \hat{e}_z|} \left[P(\hat{\xi}_0 \rightarrow \hat{\xi}) + \left(\frac{\tau_0}{\tau_0 - \tau} r_f(\hat{\xi}') + r_f(\hat{\xi}_0) \right) P(\hat{\xi}_0 \rightarrow \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z)\hat{e}_z) \right] \\ &\quad + r_f(\hat{\xi}_0)F_0\delta^{(2)}(\hat{\xi}' - \hat{\xi}_0)\exp[-(2\tau_0 - \tau)/|\hat{\xi}_0 \bullet \hat{e}_z|], \end{aligned} \tag{9.6}$$

where again $u' = |u|$, $\phi' = \phi$.

vanishes unless $u = -u_0$ and $\phi = \phi_0$) are all proportional to the product $\omega_0\tau_0P$. For a multi-component slab, this product can be written

$$\begin{aligned}\frac{\omega_0\tau_0P}{4\pi} &= \frac{b}{c}cz_0\frac{4\pi\beta}{4\pi} = z_0\beta \\ &= z_0\sum_{i=1}^N\beta_i = z_0\sum_{i=1}^N\frac{b_iP_i}{4\pi} \\ &= z_0\sum_{i=1}^N\frac{c_i\omega_{0i}P_i}{4\pi} = \sum_{i=1}^N\frac{\tau_{0i}\omega_{0i}P_i}{4\pi}.\end{aligned}$$

Thus, the product $\omega_0\tau_0P$ is summable over the scattering species in the atmosphere, and L_{Other} can be separated into

$$L_{\text{Other}}(0, u, \phi) = \sum_{i=1}^N \left(L_{\text{Other}}(0, u, \phi) \right)_i, \quad \text{for } u \neq -u_0,$$

where $\left(L_{\text{Other}}(0, u, \phi) \right)_i$ is $L_{\text{Other}}(0, u, \phi)$ when only species i is present in the atmospheric slab. In general, the various species of aerosol are rarely separated. Rather they are grouped together in a component simply called “aerosol.” Considering the aerosol as one species of scatterer and the Rayleigh scattering by air molecules as the other species, we have

$$\frac{\omega_0\tau_0P}{4\pi} = \frac{\tau_{0r}\omega_{0r}P_r}{4\pi} + \frac{\tau_{0a}\omega_{0a}P_a}{4\pi},$$

and

$$L_{\text{Other}}(0, u, \phi) = \left(L_{\text{Other}}(0, u, \phi) \right)_r + \left(L_{\text{Other}}(0, u, \phi) \right)_a, \quad \text{for } u \neq -u_0,$$

where the (additional) subscripts r and a refer to Rayleigh and aerosol, respectively. The notation above is cumbersome, so we simplify it to

$$L_{\text{Other}}(0, u, \phi) = L_r(0, u, \phi) + L_a(0, u, \phi), \quad \text{for } u \neq -u_0, \quad (9.7)$$

with

$$L_x(0, u, \phi) = \frac{\omega_{0x}\tau_{0x}F_0}{4\pi|u|} \left[P_x(\Theta_-) + \left(r_f(|u|) + r_f(u_0) \right) P_x(\Theta_+) \right], \quad (9.8)$$

where $x = r$ or a , and the scattering angles are given by

$$\cos \Theta_{\pm} = \mp uu_0 + \sqrt{1 - u^2}(1 - u_0^2) \cos(\phi - \phi_0).$$

The reader will recall that the atmospheric slab has been assumed to be homogeneous; however, in the single-scattering approximation (with $\tau_0 \ll 1$) in which photons are allowed to *interact* with the medium only *once*, the vertical structure of the atmosphere is

irrelevant. For example, contrast the situation with a thin aerosol layer at the top of the atmosphere with that of a thin aerosol layer near the bottom, with Rayleigh scattering comprising the rest of the atmosphere in both cases. If the aerosol is near the top, photons backscattered from it will not have suffered any attenuation due to Rayleigh scattering. When the aerosol is near the bottom, photons will be attenuated by two trips through the Rayleigh scattering layer – one prior to aerosol scattering and one after aerosol scattering. In the process just described the photon would have had to interact at least once (the process of attenuation) in *addition* to scattering from the aerosol layer. This cannot happen in the single scattering approximation. Thus, in the limit that $\tau_0 \ll 1$ the vertical structure of the atmosphere is unimportant. However, if $\tau_0 \not\ll 1$, so multiple scattering is important, one must expect that the vertical structure of the atmosphere will play a role in determining $L_{\text{Other}}(0, u, \phi)$.

9.3.2 Identification of Sun Glitter in $L_{\text{Other}}(0, \hat{\xi})$.

In Eqs. (9.5) and (9.6) the term containing the Dirac delta function provides the radiance due to specular reflection of the solar beam from the water surface. This term is the flat-surface equivalent of what is commonly called the “Sun glitter” and was studied in detail in Chapter 7. If the water surface is roughened by the wind, this radiance is no longer in a single direction, but exists in a narrow range of directions centered on (or close to) the direction $\hat{\xi}$ that is the solution to $\hat{\xi} = \hat{\xi}_0 - 2(\hat{\xi}_0 \bullet \hat{e}_z)\hat{e}_z$, or equivalently the direction specified by $u = -u_0$ and $\phi = \phi_0$. Also, the Dirac delta functions are replaced by a finite function that describes how the solar radiance reaching the surface is reflected. Then this term is usually written as $L_g(0, \hat{\xi})$ or $L_g(0, u, \phi)$, etc. We will retain this notation and write for the flat-surface case as well,

$$L_g(0, u, \phi) = r_f(u_0)F_0\delta(u + u_0)\delta(\phi - \phi_0)\exp[-2\tau_0/u_0],$$

where we have replaced $\delta(u' - u_0)$ in Eq. (9.3) by $\delta(-u - u_0) = \delta(u + u_0)$ and $\delta(\phi' - \phi_0)$ by $\delta(\phi - \phi_0)$ because $u + u_0 = 0$ and $\phi - \phi_0 = 0$ for this term to contribute. Furthermore, since $r_f(u_0)F_0\delta(u + u_0)\delta(\phi - \phi_0)\exp[-\tau_0/u_0]$ is the radiance at the water surface associated with the reflected solar beam, i.e.,

$$L_g(\tau_0, u, \phi) = r_f(u_0)F_0\delta(u + u_0)\delta(\phi - \phi_0)\exp[-\tau_0/u_0], \quad (9.9)$$

we can write

$$\begin{aligned} L_g(0, u, \phi) &= \exp[-\tau_0/u_0] \times L_g(\tau_0, u, \phi), \\ &= T(\tau_0 \rightarrow 0, u_0, \phi_0) \times L_g(\tau_0, u, \phi), \end{aligned} \quad (9.10)$$

where $T(\tau_0 \rightarrow 0, u_0, \phi_0) = T(\tau_0 \rightarrow 0, |u|, \phi) = \exp[-\tau_0/|u|]$ is the *direct* transmittance of the atmosphere along the path from the water surface to the sensor. Note that unlike the

diffuse transmittance, developed fully in the next section, T is a property of the atmosphere alone. It is simply the transmittance of a perfectly parallel beam of radiation through the entire atmosphere when the beam is inclined at an angle $\theta = \cos^{-1} |u|$ with respect to the normal to the water surface. When the surface is roughened by the wind the glitter is neither in the form of a parallel beam nor totally diffuse; therefore the appropriate transmittance will be somewhere between the direct transmittance (T) and the diffuse transmittance (t), which we study next.⁴

9.3.3 Contribution of Atmospheric Processes to $t(z_0 \rightarrow 0, \hat{\xi})L_w(z_0, \hat{\xi})$.

We now examine the second term in Eq. (9.1), i.e., the water-leaving radiance transmitted to the top of the atmosphere. In general, radiance exits the water and enters the atmosphere in all upward directions. This radiance is then modified by the atmosphere through absorption and scattering and exits the top of the atmosphere where it is measured by the sensor. Thus, the radiative transfer problem that must be solved consists of finding the radiance $L_w(0, \hat{\xi})$ exiting the top of the atmosphere measured by a radiometer (aimed toward $-\hat{\xi}$), given radiance $L_w(z_0, \hat{\xi}')$ entering the bottom of the atmosphere, where $\hat{\xi}'$ is a unit vector pointing in *any* upward direction (Figure 9.2).⁵ The diffuse transmittance is then

$$t(z_0 \rightarrow 0, \hat{\xi}) = \frac{L_w(0, \hat{\xi})}{L_w(z_0, \hat{\xi})}, \quad (9.11)$$

i.e., the radiance exiting the top of the atmosphere in the direction $\hat{\xi}$ divided by the radiance entering the bottom of the atmosphere in the *same* direction. Thus, we need to solve the radiative transfer equation subject to the boundary conditions,

$$\begin{aligned} L^{(0)}(0, u', \phi) &= 0 & \text{for } u' > 0, \\ L^{(0)}(\tau_0, u', \phi) &= L_w(\tau_0, u', \phi) & \text{for } u' < 0, \\ L^{(n)}(0, u', \phi) &= 0 & \text{for } u' > 0, \\ L^{(n)}(\tau_0, u', \phi) &= 0 & \text{for } u' < 0, \end{aligned}$$

where, as usual, $u' = \hat{\xi}' \cdot \hat{e}_z$, the unit vector $\hat{\xi}'$ is in an arbitrary upward direction, and $\tau_0 = cz_0$.

⁴If we try to incorporate the surface roughness, using the relationships developed in Chapter 7, into the single-scattering analysis, it is found that the terms for which sky radiance interact with the surface lead to divergent integrals. (The integrals do not diverge in the first-order solution.) Thus, no single scattering analysis is performed in the presence of surface roughness.

⁵The subscript “ w ” on $L_w(0, \hat{\xi})$ departs from our convention of using “ w ” to represent “water-leaving” quantities; however, we use the present notation here for clarity to remind the reader that $L_w(0, \hat{\xi})$ is the radiance exiting the top of the atmosphere when $L_w(\tau_0, \hat{\xi})$ is exiting the water surface, i.e., incident on the bottom of the atmosphere, in the *absence* of the solar beam.

The first-order solution can be developed in a manner analogous to that for diffuse light entering the top of the atmosphere, as presented in Appendix 1 to Chapter 2. The result for $u < 0$ is

$$\begin{aligned} L_w(0, u, \phi) &= L_w(\tau_0, u, \phi) \exp(\tau_0/u) \\ &+ \frac{\omega_0}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u, \phi' \rightarrow \phi)}{1 - u/u'} L_w(\tau_0, u' \phi') \\ &\times [\exp(\tau_0/u') - \exp(\tau_0/u)]. \end{aligned} \quad (9.12)$$

The interpretation of the two terms in this equation is provided in Figure 9.3. Noting that Eq. (9.12) provides accurate radiances only for $\tau_0 \ll 1$, we expand the exponentials and retain only the first-order term, yielding the single-scattering solution. This results in

$$\begin{aligned} L_w(0, u, \phi) &= L_w(\tau_0, u, \phi) \\ &\times \left[1 + \frac{\tau_0}{u} \left(1 - \frac{\omega_0}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) \frac{L_w(\tau_0, u' \phi')}{L_w(\tau_0, u, \phi)} \right) \right]. \end{aligned} \quad (9.13)$$

Thus, the diffuse transmittance can be written

$$\begin{aligned} t(\tau_0 \rightarrow 0, u, \phi) &= \frac{L_w(0, u, \phi)}{L_w(\tau_0, u, \phi)} \\ &= 1 + \frac{\tau_0}{u} \left(1 - \frac{\omega_0}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u, \phi' \rightarrow \phi) \frac{L_w(\tau_0, u' \phi')}{L_w(\tau_0, u, \phi)} \right), \end{aligned} \quad (9.14)$$

or, in coordinate-free form,

$$t(\tau_0 \rightarrow 0, \hat{\xi}) = 1 + \frac{\tau_0}{\hat{\xi} \bullet \hat{e}_z} \left(1 - \frac{\omega_0}{4\pi} \int_{\hat{\xi}' \bullet \hat{e}_z < 0} d\Omega(\hat{\xi}') P(\hat{\xi}' \rightarrow \hat{\xi}) \frac{L_w(\tau_0, \hat{\xi}')}{L_w(\tau_0, \hat{\xi})} \right), \quad (9.15)$$

where $d\Omega(\hat{\xi}')$ is a differential solid angle around $\hat{\xi}'$ and the integration is restricted to directions for which $\hat{\xi}' \bullet \hat{e}_z < 0$. Note that $\hat{\xi} \bullet \hat{e}_z < 0$ as well.

These single-scattering representations of the diffuse transmittance are incomplete because there is one single-scattering process that has been omitted. The omitted process is shown in Figure 9.4. Inclusion of the process in Figure 9.4 requires computing downward-scattered radiance falling on the water surface. In a computation similar to that leading to Eq. (9.12) this radiance is found to be

$$\begin{aligned} L(\tau_0, u'', \phi'') &= \frac{\omega_0}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{P(u' \rightarrow u'', \phi' \rightarrow \phi'')}{1 - u''/u'} L_w(\tau_0, u' \phi') \\ &\times \exp(\tau_0/u') [\exp(-\tau_0/u') - \exp(-\tau_0/u'')], \end{aligned} \quad (9.16)$$

or expanding the exponentials to first order,

$$L(\tau_0, u'', \phi'') = \frac{\omega_0 \tau_0}{4\pi u''} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P(u' \rightarrow u'', \phi' \rightarrow \phi'') L_w(\tau_0, u' \phi'). \quad (9.17)$$

This radiance is then Fresnel-reflected at the water surface, resulting in an upward radiance $L(\tau_0, u, \phi) = r_f(u'')L(\tau_0, u'', \phi'')$. (Note, the law of reflection requires $u = -u''$ and $\phi = \phi''$, which for a given upward direction (u, ϕ) completely specifies the intermediate direction (u'', ϕ'') . Because one scattering event has already taken place in this process, the reflected radiance is then transmitted to the top of the atmosphere without further interaction. This contribution adds a term

$$-\frac{\omega_0 \tau_0 r_F(-u)}{4\pi u} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P(u' \rightarrow -u, \phi' \rightarrow \phi) \frac{L_w(\tau_0, u' \phi')}{L_w(\tau_0, u, \phi)}$$

to the right-hand-side of Eq. (9.14). Combining all of the contributions, the diffuse transmittance becomes

$$\begin{aligned} t(\tau_0 \rightarrow 0, u, \phi) &= \frac{L_w(0, u, \phi)}{L_w(\tau_0, u, \phi)} \\ &= 1 + \frac{\tau_0}{u} \left[1 - \frac{\omega_0}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{L_w(\tau_0, u' \phi')}{L_w(\tau_0, u, \phi)} \right. \\ &\quad \left. \times \left(P(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \right], \end{aligned} \quad (9.18)$$

or in the coordinate-free representation,

$$\begin{aligned} t(\tau_0 \rightarrow 0, \hat{\xi}) &= 1 + \frac{\tau_0}{\hat{\xi} \bullet \hat{e}_z} \left[1 - \frac{\omega_0}{4\pi} \int_{\hat{\xi}' \bullet \hat{e}_z < 0} d\Omega(\hat{\xi}') \frac{L_w(\tau_0, \hat{\xi}')}{L_w(\tau_0, \hat{\xi})} \right. \\ &\quad \left. \times \left(P(\hat{\xi}' \rightarrow \hat{\xi}) + r_f(\hat{\xi}'') P(\hat{\xi}' \rightarrow \hat{\xi}'') \right) \right], \end{aligned} \quad (9.19)$$

where $\hat{\xi}'' = \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z)\hat{e}_z$.⁶ Note that u and $\hat{\xi} \bullet \hat{e}_z$ are both negative so the integrals make *positive* contributions to t . Also, the Fresnel reflectance can be taken outside the integration, as it is not dependent on the integration variables. Typically, the term containing the Fresnel reflectance is small compared to the term without it for two reasons:

⁶The associated formulas for the case where the sensor is within the atmosphere are easy to develop from the formulas already presented using Figures 9.3 and 9.4. If the sensor is at an optical depth τ , where $0 < \tau < \tau_0$ (or at z , where $0 < z < z_0$) then Figure 9.3 shows that only the atmosphere between τ and τ_0 can contribute to the various terms in Eq. (9.15). Thus, for the portions of t provided by Figure 9.3, Eq. (9.15) is applicable with τ_0 replaced by $\tau_0 - \tau$. In contrast, for the portion of t provided by Figure 9.4, the entire atmosphere can contribute to the downward radiance in the $\hat{\xi}''$ direction and, as there are no additional interactions between the surface and the sensor, this term will remain unchanged. The final

(1) the Fresnel reflectance is small ($< 10\%$) for useful viewing directions; and (2) the term containing the Fresnel reflectance is associated with backscattering directions for which the phase function (when aerosols are present) is smaller than for the forward scattering directions associated with the first term.

The single-scattering equations developed for the diffuse transmittance clearly show that $t(\tau_0 \rightarrow 0, \hat{\xi})$ is not a property of the atmosphere. Rather, it depends on $L_w(\tau_0, \hat{\xi})$, i.e., the angular distribution of radiance exiting the water surface (the *radiance distribution*). However, it does depend on the properties of the atmosphere through $P(\hat{\xi}' \rightarrow \hat{\xi})$, ω_0 , and τ_0 . Interestingly, in contrast to the direct transmittance, i.e., $\exp[\tau_0/(\hat{\xi} \bullet \hat{e}_z)]$, where $\hat{\xi} \bullet \hat{e}_z < 0$, there is no necessity that $t(\tau_0 \rightarrow 0, \hat{\xi}) \leq 1$ in the presence of scattering. In fact, in the *unlikely* event that $L_w(\tau_0, \hat{\xi}') = 0$ for some $\hat{\xi}'$, then unless $\omega_0 = 0$, $t(\tau_0 \rightarrow 0, \hat{\xi}') = +\infty$!

Equations (9.18) and (9.19) are easy to modify in the case of a multi-component atmosphere: simply replace $\omega_0 \tau_0 P$ with $\sum_{i=1}^N \omega_{0i} \tau_{0i} P_i$. Then, for example, Eq. (9.18) becomes

$$t(\tau_0 \rightarrow 0, u, \phi) = 1 + \frac{\tau_0}{u} - \frac{1}{4\pi u} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{L_w(\tau_0, u', \phi')}{L_w(\tau_0, u, \phi)} \\ \times \sum_{i=1}^N \omega_{0i} \tau_{0i} \left(P_i(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_i(u' \rightarrow -u, \phi' \rightarrow \phi) \right),$$

or in the case of Rayleigh and aerosol scattering,

$$t(\tau_0 \rightarrow 0, u, \phi) = 1 + \frac{\tau_{0r}}{u} + \frac{\tau_{0a}}{u} \\ - \frac{1}{4\pi u} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{L_w(\tau_0, u', \phi')}{L_w(\tau_0, u, \phi)} \\ \times \omega_{0r} \tau_{0r} \left(P_r(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_r(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \\ - \frac{1}{4\pi u} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{L_w(\tau_0, u', \phi')}{L_w(\tau_0, u, \phi)} \\ \times \omega_{0a} \tau_{0a} \left(P_a(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_a(u' \rightarrow -u, \phi' \rightarrow \phi) \right). \quad (9.21)$$

result is then

$$t(\tau_0 \rightarrow \tau, \hat{\xi}) = \frac{L_w(\tau, \hat{\xi})}{L_w(\tau_0, \hat{\xi})} \\ = 1 + \frac{(\tau_0 - \tau)}{\hat{\xi} \bullet \hat{e}_z} \left[1 - \frac{\omega_0}{4\pi} \int_{\hat{\xi}' \bullet \hat{e}_z < 0} d\Omega(\hat{\xi}') \frac{L_w(\tau_0, \hat{\xi}')}{L_w(\tau_0, \hat{\xi})} \right. \\ \left. \times \left(P(\hat{\xi}' \rightarrow \hat{\xi}) + \frac{\tau_0}{(\tau_0 - \tau)} r_f(\hat{\xi}'') P(\hat{\xi}' \rightarrow \hat{\xi}'') \right) \right], \quad (9.20)$$

where $\hat{\xi}'' = \hat{\xi} - 2(\hat{\xi} \bullet \hat{e}_z) \hat{e}_z$.

This expression can be identified as the first-order expansion of the product of two polynomials, e.g., $1 + x + y \approx (1 + x)(1 + y)$ if both x and y are $\ll 1$. When such an identification is made, the diffuse transmittance is approximated as

$$t(\tau_0 \rightarrow 0, u, \phi) = t_r(\tau_{0r} \rightarrow 0, u, \phi)t_a(\tau_{0a} \rightarrow 0, u, \phi), \quad (9.22)$$

where

$$\begin{aligned} t_x(\tau_0 \rightarrow 0, u, \phi) = 1 - \frac{\tau_{0x}}{|u|} \\ + \frac{\omega_{0x}\tau_{0x}}{4\pi|u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{L_w(\tau_0, u'\phi')}{L_w(\tau_0, u, \phi)} \\ \times \left(P_x(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u)P_x(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \end{aligned} \quad (9.23)$$

$x = r$ or a , and, as $u < 0$, we have replaced $+\tau_{0x}/u$ by $-\tau_{0x}/|u|$ to emphasize that fact that the second term results in a *decrease* in the transmittance, while the third results in an *increase*.

9.3.4 Summary of the Single-Scattering Analysis.

The results of this single-scattering analysis can be summarized in the equations below describing the various contributions to the radiance $L_t(0, u, \phi)$ exiting the top ($z = 0$) of a thin, homogeneous-slab, atmosphere of thickness z_0 above a flat, Fresnel-reflecting water body, that provides an upward radiance $L_w(z_0, u, \phi)$ into the bottom of the atmosphere when the top of the atmosphere is illuminated by the solar beam.

$$\begin{aligned} L_t(0, u, \phi) = L_r(0, u, \phi) + L_a(0, u, \phi) \\ + T(\tau_0 \rightarrow 0, u, \phi)L_g(\tau_0, u, \phi) + t(\tau_0 \rightarrow 0, u, \phi)L_w(\tau_0, u, \phi), \end{aligned} \quad (9.24)$$

where

$$L_x(0, u, \phi) = \frac{\omega_{0x}\tau_{0x}F_0}{4\pi|u|} \left[P_x(\Theta_-) + \left(r_f(|u|) + r_f(u_0) \right) P_x(\Theta_+) \right],$$

$$\cos \Theta_{\pm} = \mp uu_0 + \sqrt{1 - u^2}(1 - u_0^2) \cos(\phi - \phi_0),$$

$$T(\tau_0 \rightarrow 0, u, \phi) = \exp[-\tau_0/|u|],$$

$$L_g(\tau_0, u, \phi) = r_F(u_0)F_0\delta(u + u_0)\delta(\phi - \phi_0)\exp[-\tau_0/u_0],$$

and

$$t(\tau_0 \rightarrow 0, u, \phi) = t_r(\tau_{0r} \rightarrow 0, u, \phi)t_a(\tau_{0a} \rightarrow 0, u, \phi),$$

with

$$\begin{aligned}
 t_x(\tau_0 \rightarrow 0, u, \phi) &= 1 - \frac{\tau_{0x}}{|u|} \\
 &+ \frac{\omega_{0x}\tau_{0x}}{4\pi|u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \frac{L_w(\tau_0, u'\phi')}{L_w(\tau_0, u, \phi)} \\
 &\times \left(P_x(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u)P_x(u' \rightarrow -u, \phi' \rightarrow \phi) \right)
 \end{aligned}$$

In these equation, the subscript “ x ” can be either “ r ” or “ a .” The equations are appropriate when the optical depth of the atmosphere, $\tau_0 = cz_0 \ll 1$. Succinctly stated, the remote sensing problem is to retrieve $L_w(\tau_0, u, \phi)$ from measurement of $L_t(0, u, \phi)$. However, even in this single-scattering approximation, these equations indicate that the water-leaving component of Eq. (9.24), $t(\tau_0 \rightarrow 0, u, \phi)L_w(\tau_0, u, \phi)$ depends on $L_w(\tau_0, u', \phi')$, for all $u' < 0$ and all ϕ' , i.e.,

$$\begin{aligned}
 t_r(\tau_0 \rightarrow 0, u, \phi)t_a(\tau_0 \rightarrow 0, u, \phi)L_w(\tau_0, u, \phi) &= \left[1 - \frac{\tau_{0r} + \tau_{0a}}{|u|} \right] L_w(\tau_0, u, \phi) \\
 &+ \frac{\omega_{0r}\tau_{0r}}{4\pi|u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' L_w(\tau_0, u'\phi') \\
 &\times \left(P_r(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u)P_r(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \\
 &+ \frac{\omega_{0a}\tau_{0a}}{4\pi|u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' L_w(\tau_0, u'\phi') \\
 &\times \left(P_a(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u)P_a(u' \rightarrow -u, \phi' \rightarrow \phi) \right).
 \end{aligned}$$

The presence of the integrals containing $L_w(\tau_0, u', \phi')$ shows that it is not possible to retrieve $L_w(\tau_0, u, \phi)$ from the measurement of $L_t(0, u, \phi)$ propagating in the single direction (u, ϕ) , even if the atmospheric parameters are precisely known. Thus, further assumptions are required to make the remote sensing problem tractable. Note the *only* approximations made thus far are that the single-scattering approximation is valid, i.e., $\tau_0 \ll 1$, so contributions to the radiance of order $\propto \tau_0^n$, where $n > 1$ are negligible, and that the water surface is flat and free of whitecaps.

9.3.5 Assumption of a Uniform $L_w(\tau_0, u', \phi')$.

The remote sensing problem can at least be made tractable by assuming that $L_w(\tau_0, u', \phi')$ is independent of direction, i.e., independent of u' and ϕ' , so $L_w(\tau_0, u', \phi') \rightarrow L_w(\tau_0)$. Then,

the diffuse transmittances become independent of the water-leaving radiance, and

$$\begin{aligned}
 & t_r(\tau_0 \rightarrow 0, u, \phi) t_a(\tau_0 \rightarrow 0, u, \phi) L_w(\tau_0) \\
 &= \left[1 - \frac{\tau_0 r}{|u|} + \frac{\omega_{0r} \tau_0 r}{4\pi |u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \left(P_r(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_r(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \right] \\
 &\times \left[1 - \frac{\tau_0 a}{|u|} + \frac{\omega_{0a} \tau_0 a}{4\pi |u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \left(P_a(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_a(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \right] \\
 &\times L_w(\tau_0).
 \end{aligned}$$

To remind the reader that this assumption has been evoked, we denote the diffuse transmittance under these circumstances by $t_x^*(\tau_0 \rightarrow 0, u, \phi)$, i.e.,

$$\begin{aligned}
 & t_x^*(\tau_0 \rightarrow 0, u, \phi) \\
 &= \left[1 - \frac{\tau_{0x}}{|u|} \right. \\
 &\quad \left. + \frac{\omega_{0x} \tau_{0x}}{4\pi |u|} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \left(P_x(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_x(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \right],
 \end{aligned} \tag{9.25}$$

and

$$\begin{aligned}
 L_t(0, u, \phi) &= L_r(0, u, \phi) + L_a(0, u, \phi) \\
 &\quad + T(\tau_0 \rightarrow 0, u, \phi) L_g(\tau_0, u, \phi) + t^*(\tau_0 \rightarrow 0, u, \phi) L_w(\tau_0, u, \phi).
 \end{aligned} \tag{9.26}$$

Clearly, in contrast to t , the transmittance t^* is a property of the atmosphere *alone*. Equation (9.25) is often considered to represent the first two terms in an exponential expansion, i.e., $1 + x \approx \exp(x)$ when $x \ll 1$. When this identification is made,

$$\begin{aligned}
 & t_x^*(\tau_0 \rightarrow 0, u, \phi) \approx \exp \left[-\frac{\tau_{0x}}{|u|} \left(1 \right. \right. \\
 &\quad \left. \left. - \frac{\omega_{0x}}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' \left(P_x(u' \rightarrow u, \phi' \rightarrow \phi) + r_f(-u) P_x(u' \rightarrow -u, \phi' \rightarrow \phi) \right) \right) \right].
 \end{aligned} \tag{9.27}$$

It is useful to investigate the nature of the integrals. First, we note that r_F is independent of u' , so it can be taken outside the integral. Consider

$$\frac{1}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P_x(u' \rightarrow u, \phi' \rightarrow \phi) \triangleq p_x^{--}(|u|), \tag{9.28}$$

where $u < 0$ and $u' < 0$.⁷ This integral corresponds to the processes depicted in Figure 9.5. Recalling that $P_x(u' \rightarrow u, \phi' \rightarrow \phi)/4\pi$ is the probability density for scattering from (u', ϕ')

⁷Note that $p_x^{--}(|u|)$ does not depend on ϕ . The easiest way to verify this is to expand $P_x(u' \rightarrow u, \phi' \rightarrow \phi)$ in a Fourier series in $\phi' - \phi$ (assuming that P_x depend only on the angle between $\hat{\xi}'$ and $\hat{\xi}$). Upon integrating the series over ϕ' , only the term independent of $\phi' - \phi$ survives.

to (u, ϕ) , it is clear that $p_x^{--}(|u|)$ is the probability of scattering into a particular direction (u, ϕ) , where $u < 0$, from all directions (u', ϕ') with $u' < 0$. (The superscript “--” is used to indicate scattering from an upward directed $\hat{\xi}'$ to an upward directed $\hat{\xi}$, i.e., u' and u both < 0 .) Using the symmetry of the phase function,

$$P_x(u' \rightarrow u, \phi' \rightarrow \phi) = P_x(u \rightarrow u', \phi \rightarrow \phi'),$$

it is seen that $p_x^{--}(|u|)$ is also the probability of scattering into all directions (u', ϕ') , where $u' < 0$, from a particular direction (u, ϕ) with $u < 0$. In addition, it is easy to see that $p_x^{--}(|u|) = p_x^{++}(|u|)$. Likewise

$$\frac{1}{4\pi} \int_{-1}^0 du' \int_0^{2\pi} d\phi' P_x(u' \rightarrow -u, \phi' \rightarrow \phi) \triangleq p_x^{-+}(|u|) \quad (9.29)$$

corresponds to the process in Figure 9.6. Here $p_x^{-+}(|u|)$ is the probability of scattering into a particular direction $(-u, \phi)$, where $u < 0$, from all directions (u', ϕ') with $u' < 0$, or equivalently, the probability of scattering into all directions (u', ϕ') , where $u' < 0$, from a particular direction $(-u, \phi)$ with $u < 0$.⁸ Clearly, $p_x^{--}(|u|) = p_x^{+-}(|u|)$ and $p_x^{--}(|u|) + p_x^{-+}(|u|) = 1$. Also, when $|u| = 1$ ($\hat{\xi}$ directed vertically), $p_x^{--}(1)$ is the probability of scattering in the *forward* hemisphere ($\Theta \leq 90^\circ$), while $p_x^{-+}(1)$ is the probability of scattering in the *backward* hemisphere ($\Theta \geq 90^\circ$). With these identifications,

$$\begin{aligned} t_x^*(\tau_0 \rightarrow 0, u, \phi) &\approx \exp \left[-\frac{\tau_{0x}}{|u|} \left(1 - \omega_{0x} [p_x^{--}(|u|) + r_f(|u|) p_x^{-+}(|u|)] \right) \right] \\ &= \exp \left[-\frac{\tau_{0x}}{|u|} \left(1 - \omega_{0x} [p_x^{--}(|u|)(1 - r_f(|u|)) + r_f(|u|)] \right) \right]. \end{aligned} \quad (9.30)$$

As analytic expressions of P_x are usually not available, the integrals for p_x^{--} and p_x^{-+} must be carried out numerically; however, for Rayleigh scattering an analytical formula is available and it is straightforward to carry out the integrals in Eqs. (9.28) and (9.29).⁹ The result is $p_r^{--}(|u|) = p_r^{-+}(|u|) = \frac{1}{2}$, so

$$t_r^*(\tau_0 \rightarrow 0, u, \phi) \approx \exp \left[-\frac{\tau_{0r}}{|u|} \left(1 - \frac{\omega_{0r}}{2} [1 + r_f(|u|)] \right) \right]. \quad (9.31)$$

⁸The quantities $p_x^{--}(|u|)$ and $p_x^{-+}(|u|)$ are identical to the quantities P_{rd} and P_{ru} , respectively, in Chapter 6, when the full phase function is used in the integrals, rather than the portion that remains after removal of the Dirac delta function.

⁹Insert the Fourier expansion of the Rayleigh phase function from Chapter 2 (Eq. (2.66)) into Eqs. (9.28) and (9.29) and carry out the integrals.

9.4 Accuracy of the Single-Scattering Solution.

The purpose of employing the single-scattering solution of the remote sensing problem was to develop approximate, but analytical, formulas for the various quantities in the remote sensing equation (Eq. (9.1)), $L_{\text{Other}}(0, \hat{\xi})$ (or $L_r(0, \hat{\xi}) + L_a(0, \hat{\xi})$) and $t(\tau_0 \rightarrow 0, \hat{\xi})$ (or $t_r(\tau_0 \rightarrow 0, \hat{\xi}) \times t_a(\tau_0 \rightarrow 0, \hat{\xi})$). The resulting single-scattering formulas are valid for $\tau_{0r} \ll 1$ and $\tau_{0a} \ll 1$; however, we have seen in Chapter 4 that at 412 nm (the minimal wavelength for spectral bands on many space-borne ocean color sensors) $\tau_{0r} \approx 0.32$. Also, the upper limit for τ_{0a} at 412 nm in a clear, totally maritime, atmosphere is ~ 0.1 to 0.2 . Are these optical depths sufficiently small for the single-scattering formulas to be useful? We now investigate this question by comparing the predictions of single-scattering theory with the results of “exact” solutions of the *scalar* radiative transfer equation (for the same physical setting) at a wavelength of 412 nm. As usual, by “exact” solutions we mean numerical solutions that can be made as accurate as desired.

9.4.1 Accuracy of the Single-Scattering Solution for L_r .

We begin by examining the accuracy of the approximation when $\tau_{0a} = 0$, i.e., in the absence of the aerosol. Consider a pure Rayleigh scattering slab above a Fresnel-reflecting water body for which $L_w(\tau_{0r}, u, \phi) = 0$ for all $u < 0$ and ϕ , i.e., for which no radiance *exits* the water. Assume that the remote sensor is viewing radiance exiting the top of the atmosphere making a polar angle $\theta_{ex} = \pi - \theta = \cos^{-1}(|u|)$ with the *upward* normal (see insert in Figure 9.7) and an azimuth angle $\phi = \phi_0 \pm \frac{\pi}{2}$. A sensor scanning through various values of θ_{ex} , but keeping the azimuth constant at this value is said to be scanning in the *perpendicular* plane. This approximates the sun-viewing geometry used in many space-borne sensors.

It is convenient at this point to switch from radiance L to *reflectance* ρ . The reflectance ρ corresponding to a radiance L is defined to be

$$\rho = \frac{\pi L}{F_0 \cos \theta_0} = \frac{\pi L}{F_0 u_0},$$

where F_0 is the extraterrestrial solar irradiance leading to L , and θ_0 is the solar zenith angle at the time of the measurement. (Note, if the radiance exiting the top of the atmosphere is independent of the direction of propagation, the atmosphere is said to be *lambertian* and ρ is the *albedo*.) With this change, Eq. (9.8) for L_x becomes

$$\rho_x(0, u, \phi) = \frac{\omega_{0x} \tau_{0x}}{4|u|u_0} \left[P_x(\Theta_-) + (r_F(|u|) + r_F(u_0)) P_x(\Theta_+) \right], \quad (9.32)$$

where $x = r$ or a , and

$$\cos \Theta_{\pm} = \mp uu_0 + \sqrt{1 - u^2}(1 - u_0^2) \cos(\phi - \phi_0).$$

In Figure 9.7 we compare the reflectance ρ_r at 412 nm ($\tau_{0r} = 0.32$) computed using Eq. (9.32) with those computed from a full solution of the scalar RTE for three separate solar zenith angles: $\theta_0 = 20^\circ$, 40° , and 60° . As the spectral bands associated with remote sensing systems are chosen to be in atmospheric “windows,” i.e., spectral regions of little or no absorption, the computations are for $\omega_{0r} = 1$ (no atmospheric absorption). The figure shows excellent agreement between the two computations. This is emphasized in Figure 9.8 which provides the % error in ρ_r when computed using the single-scattering approximation. The error in the figure is computed as $100\% \times (\rho_r(\text{Single-scattering}) - \rho_r(\text{Exact})) / \rho_r(\text{Exact})$, so a positive error means that the single-scattering solution is too large. Since most remote sensors operate with $\theta_{ex} < 50^\circ$ to 60° , the results suggest that the single-scattering approximation to ρ_r is in error by $\sim \pm 3\%$ or an absolute error in ρ_r of $\sim \pm 0.004$.

9.4.2 Accuracy of the Single-Scattering Solution for $L_r + L_a$.

To examine the accuracy of the single-scattering solution when aerosols are present requires optical properties for the aerosols. For this purpose, we use the M70 aerosol model of Shettle and Fenn which, as we have seen in Chapter 4, approximates the properties of pure maritime aerosols. A value for the aerosol optical thickness of 0.2 is assumed because it is close to the maximum for “pure” maritime atmospheres (actually, it is somewhat above the largest values observed). In addition, because such an aerosol is known to be confined to the marine boundary layer, we assume that the aerosols are confined to a thin layer near the surface. For evaluation of the single-scattering solution, we consider the “exact” values of $\rho_r + \rho_a$ to be those computed using full multiple scattering methods for a two-layer atmosphere with aerosols confined to the lower layer and Rayleigh scattering by the air confined to the upper layer. As in the case of a pure Rayleigh-scattering atmosphere, the water body is bounded by a Fresnel-reflecting surface and is totally absorbing ($L_w = 0$). Figure 9.9 provides the comparison between the exact and the single-scattering computation of $\rho_r + \rho_a$, and Figure 9.10 provides the error in the single-scattering solution. That the addition of aerosols increases the error in single-scattering solution is to be expected, as the total optical thickness is increased to 0.52. However, even at this value of τ_0 the error is *not* excessive, except for small θ_{ex} and θ_0 . The increased error in this regime is due to small-angle scattering by the aerosol followed by Fresnel reflection from the surface. The linearized first-order solution simply does not properly account for this process, which is much more important when aerosols are present because their phase function strongly favors scattering through small angles. In contrast to the pure molecular atmosphere,

and except for this regime, with the addition of aerosols the error in the single-scattering estimate has the tendency to be negative over most of the range of θ_{ex} and θ_0 . The magnitude is somewhat higher than that for ρ_r alone, and the error in the single-scattering $\rho_r + \rho_a$ is usually $\lesssim \pm 0.008 - 0.010$.

9.4.3 Accuracy of the Single-Scattering Solution for L_a .

Figure 9.11 provides the single-scattering approximation to ρ_a . Note that ρ_a is $\approx \frac{1}{10}$ to $\frac{1}{5}$ of ρ_r (Figure 9.7) even though the optical thickness is 60% of τ_{0r} . This disparity is due to the much smaller backscattering associated with the aerosol phase function compared to that for Rayleigh scattering. The accuracy with which the single-scattering approximation yields ρ_a would at first seem to be irrelevant because τ_{0r} is never zero, i.e., the aerosol never scatters in the absence of Rayleigh scattering. However, there is a multiple scattering analog to ρ_a : the analog is $[(\rho_r + \rho_a) - \rho_r]$. Thus, we ask how well the single-scattering formulas reproduce the exact values of $[(\rho_r + \rho_a) - \rho_r]$. Figures 9.9 and 9.7 provide exact values of $\rho_r + \rho_a$ and ρ_r respectively. Subtracting these yields the multiple-scattering or exact value of our analog to ρ_a . Comparing these with the single-scattering approximation to ρ_a provides the error in the single-scattering approximation as shown in Figure 9.12. With the exception of the regime mentioned above, where both θ_{ex} and θ_0 are small, the single-scattering formulas provide the quantity $[(\rho_r + \rho_a) - \rho_r]$ with an error ~ -20 to -30% . Figure 9.13 shows that the absolute error in this quantity is -0.008 to $+0.010$ even including the small-angle regime.

9.4.4 Accuracy of the Single-Scattering Solution for t_r^* .

We cannot compute t_r without knowing $L_w(\tau_0, \hat{\xi})$; however, in the case when $L_w(\tau_0, \hat{\xi})$ is independent of $\hat{\xi}$, we can calculate t_t^* . Assuming as above, that $\omega_{0r} = 1$, Equation (9.31) for t_r^* becomes

$$t_r^*(\tau_{0r} \rightarrow 0, u, \phi) \approx \exp \left[-\frac{\tau_{0r}}{2|u|} \left(1 - r_F(|u|) \right) \right]. \quad (9.33)$$

The exact values for t_r^* can be computed by solving the RTE for a homogeneous, Rayleigh-scattering slab, bounded on the bottom by a Fresnel-reflecting surface and illuminated

from below by a uniform $L_w(\tau_{0r}, \hat{\xi})$, i.e., L_w independent of $\hat{\xi}$. Then,¹⁰

$$t_r^*(\tau_{0r} \rightarrow 0, \hat{\xi}) = \frac{L_w(0, \hat{\xi})}{L_w(\tau_{0r}, \hat{\xi})}.$$

Figure 9.14 compares the single-scattering t_r^* computed using Eq. (9.33) with the exact value. Clearly, the single-scattering formula provides a t_r^* that is too small, with the error ranging from 1.5% near $\theta_{ex} = 0$ to 7% near $\theta_{ex} = 75^\circ$. This error can be rendered almost negligible for small θ_{ex} ($|u|$ near 1) by replacing $2|u|$ in Eq. (9.33) by $2.2|u|$. For comparison, the direct transmittance

$$T_r(\tau_{0r} \rightarrow 0, u, \phi) = \exp \left[\frac{\tau_{0r}}{|u|} \right],$$

is shown as the dotted line on Figure 9.14. Equation (9.31) shows that $t_r^* \rightarrow T_r$ as $\omega_{0r} \rightarrow 0$, so $t_r^* \geq T_r$.

9.4.5 Accuracy of the Single-Scattering Solution for t^* .

Computation of the single-scattering solution to t^* requires computation of p_a^{--} . This computation is provided in Figure 9.15 for the M70 aerosol model. That p_a^{--} is largest at $\theta_{ex} = 0$ and falls to 0.5 at $\theta_{ex} = 90^\circ$, owes to the strong forward scattering exhibited by aerosols. Combining the single-scattering solution for t_a^* with the single-scattering solution for t_r^* provides the single-scattering solution for t^* . This is compared with the exact t^* in Figure 9.16. The figure shows that the error in the single-scattering t^* is similar in magnitude to the error in the single-scattering t_r^* , suggesting that there is little error in the single-scattering t_a^* . Because Rayleigh scattering is never absent in the atmosphere, an exact t_a^* has no meaning; however, a meaningful counterpart to an exact t_a^* is t^*/t_r^* . Figure 9.17 shows excellent agreement between the exact t^*/t_r^* and the single-scattering t_a^* for the M70 aerosol model with $\tau_{0a} = 0.2$. The error in t_a^* is $\leq 0.5\%$ for $\theta_{ex} \leq 55^\circ$, suggesting that the first order formula for t_a^* can provide useful information regarding t^* . In particular, Figure 9.18 shows that t^* is expected to be only a weak function of τ_{0a} .

9.4.6 Summary

Examples presented in this section suggest that the single-scattering solutions for ρ_r , $\rho_r + \rho_a$, and $[(\rho_r + \rho_a) - \rho_r]$ at 412 nm are accurate to within ± 0.004 , ± 0.008 to ± 0.100 , and

¹⁰Actually, this is not the method used to compute t_r^* . In Chapter 10 we use the reciprocity principle to show that $t_r^*(\tau_0 \rightarrow 0, \hat{\xi})$ can be computed by taking $L_w = 0$, illuminating the top of the atmosphere with the solar beam directed so that $\hat{\xi}_0 = -\hat{\xi}$ and computing the total (sun *plus* sky) irradiance ($E_d(-\hat{\xi}, \tau_0)$) incident just *above* the water surface. Then $t_r^*(\tau_0 \rightarrow 0, \hat{\xi}) = E_d(-\hat{\xi}, \tau_0)/|\hat{\xi} \bullet \hat{e}_z|F_0$.

-0.008 to $+0.010$, respectively. These errors range from $\sim 3\%$ for ρ_r to -20 to -30% for $[(\rho_r + \rho_a) - \rho_r]$. Since $\tau_{0a} = 0.2$ is the upper limit for a clean maritime atmosphere, the errors in $\rho_r + \rho_a$ and $[(\rho_r + \rho_a) - \rho_r]$ will generally be smaller than shown in the figures. Thus, given the optical properties of the aerosol, ρ_{Other} in a typical remote sensing setting can be computed using single-scattering formulas with an error $\lesssim 0.010$. However, it should be noted that this error represents a significant fraction of the contribution from the aerosol (Figure 9.11). As we shall see in the next section, it also represents a considerable fraction of the water-leaving reflectance at 412 nm for Case 1 waters, for which $0 \lesssim \rho_w \lesssim 0.04$.

At 412 nm, over the range $0 \leq \theta_{ex} \leq 55^\circ$, the diffuse transmittances t_r^* and t^* computed with the single-scattering formulas are in error by -1.5 to -3.5% and -1.25 to -4.5% , respectively. In contrast, the first order t_a^* agrees with the exact values of t^*/t_r^* to within 0.5% over the same angular range. Given a precise value of ρ_{Other} , this shows that use of the single-scattering t^* would result in an error $\lesssim +1.25$ to $+3.5\%$ in ρ_w . Given the excellent accuracy of t_r^* and t^* computed in single-scattering, we expect that the t_r and t computed with single-scattering formulas (Eq. (9.25)), with the addition of the exponential approximation (i.e., $1 - x$ replaced by $\exp(-x)$ as in t^*), would provide about the same accuracy.¹¹

9.5 The Water-Leaving Reflectance

Now that we have examples of the various reflectances, e.g., ρ_r and ρ_a , and their accuracies using the single-scattering scattering theory, we can compare these in magnitude to the water-leaving reflectance. The water-leaving reflectance ρ_w is related to the water-leaving radiance L_w through

$$\rho_w(\hat{\xi}_v, \hat{\xi}_0) = \frac{\pi L_w(\hat{\xi}_v, \hat{\xi}_0)}{F_0 \cos \theta_0},$$

where $\hat{\xi}_v$ is the propagation direction of the viewed water-leaving radiance, and $\hat{\xi}_0$ is the propagation direction the solar beam.¹² In Chapter 7 we defined the normalized water-leaving radiance $[L_w]_N$ and related it to L_w through (Eq. (7.37))

$$[L_w(\hat{\xi}_v, \hat{\xi}_0)]_N = L_w(\hat{\xi}_v, \hat{\xi}_0) \frac{a_\oplus^2}{\cos \theta_0 t_E(\hat{\xi}_0)},$$

¹¹In reality, it is more accurate to assume that the upwelling radiance just beneath the water surface is uniform (Chapter 6). In fact, this is the assumption used in processing ocean color imagery from many sensors, e.g., SeaWiFS and MODIS. In this case, t^* is computed by illuminating the atmosphere (from the top) by the solar beam at a zenith angle θ_0 and computing the downward irradiance ($E_d^{(-)}$) just beneath the interface (for a totally absorbing ocean). The diffuse transmittance is then $t^*(\theta_0) = E_d^{(-)} / (t_f(\theta_0) \cos \theta_0 F_0)$.

¹²In Chapter 7 we used $\hat{\xi}_s$ to denote the propagation direction of the solar beam.

where $t_E(\hat{\xi}_0)$ is the transmittance of *irradiance* from the top of the atmosphere to the water surface, a_\oplus is the Earth-Sun distance in astronomical units, and \bar{F}_0 is the solar irradiance for $a_\oplus = 1$ AU. If we multiply this by π/\bar{F}_0

$$\frac{\pi[L_w(\hat{\xi}_v, \hat{\xi}_0)]_N}{\bar{F}_0} = \frac{\pi L_w(\hat{\xi}_v, \hat{\xi}_0) a_\oplus^2}{\bar{F}_0 \cos \theta_0 t_E(\hat{\xi}_0)},$$

and recognize \bar{F}_0/a_\oplus^2 as F_0 , we see that

$$\frac{\pi[L_w(\hat{\xi}_v, \hat{\xi}_0)]_N}{\bar{F}_0} = \frac{\rho_w(\hat{\xi}_v, \hat{\xi}_0)}{t_E(\hat{\xi}_0)}.$$

If we define the left-hand-side as the normalized water-leaving reflectance $[\rho_w(\hat{\xi}_v, \hat{\xi}_0)]_N$, i.e., as

$$[\rho_w(\hat{\xi}_v, \hat{\xi}_0)]_N \triangleq \frac{\pi[L_w(\hat{\xi}_v, \hat{\xi}_0)]_N}{\bar{F}_0}, \quad (9.34)$$

then

$$[\rho_w(\hat{\xi}_v, \hat{\xi}_0)]_N = \frac{\rho_w(\hat{\xi}_v, \hat{\xi}_0)}{t_E(\hat{\xi}_0)} \quad \text{or} \quad \rho_w(\hat{\xi}_v, \hat{\xi}_0) = [\rho_w(\hat{\xi}_v, \hat{\xi}_0)]_N t_E(\hat{\xi}_0). \quad (9.35)$$

Most measurements of $L_w(\hat{\xi}_v, \hat{\xi}_0)$ (actually the measured quantity is $L_u(\hat{\xi}_v, \hat{\xi}_0)$) are made with $\hat{\xi}_v = -\hat{e}_z$, i.e., of upwelling radiance propagating toward the zenith. From these, one can estimate $[\rho_w(-\hat{e}_z, \hat{\xi}_0)]_N$ given $t_E(\hat{\xi}_0)$, which we now show how to compute.

There are three processes that contribute to the irradiance on the water surface. The first (1) is the direct transmission of the solar beam through the atmosphere, the second (2) is the scattering from the solar beam in the downward direction, and the third (3) is upward Fresnel reflection of the solar beam followed by downward scattering toward the water surface. The irradiance transmittance in single scattering is the sum of the irradiance due to each of these three processes divided by $u_0 F_0$. We note that the radiance of the solar beam at the top of the atmosphere is $F_0 \delta(u - u_0) \delta(\phi - \phi_0)$. We have already computed the radiance at the surface for processes (1) and (2) in Chapter 2 (Eq. (2.103)):

$$\begin{aligned} L_1(\tau_0, u, \phi) + L_2(\tau_0, u, \phi) &= F_0 \delta(u - u_0) \delta(\phi - \phi_0) \left[1 - \frac{\tau_0}{u} \right] \\ &\quad + \frac{\omega_0 F_0 \tau_0}{4\pi u} P(u_0 \rightarrow u, \phi_0 \rightarrow \phi), \end{aligned}$$

where $u > 0$. We have actually almost completed the computation of the radiance for process (3) as well. Consider Eq. (9.17). For a radiance $L_w(\tau_0, u', \phi')$ incident into the bottom

of the atmosphere, this equation gives the radiance backscattered toward the surface. All we need to do is replace $L_w(\tau_0, u', \phi')$ with $r(u_0)F_0\delta(u - u_0)\delta(\phi - \phi_0)(1 - \tau_0/u_0)$. Then

$$L_3(\tau_0, u, \phi) = r(u_0)\frac{\omega_0\tau_0 F_0}{4\pi u}P(-u_0 \rightarrow u, \phi_0 \rightarrow \phi),$$

where the term in τ_0^2 is dropped as it would indicate more than one interaction with the medium. The downward irradiance at the surface is just

$$\begin{aligned} E_d(\tau_0, u_0, \phi_0) &= \int_0^{2\pi} d\phi \int_0^1 u [L_1(\tau_0, u, \phi) + L_2(\tau_0, u, \phi) + L_3(\tau_0, u, \phi)] du \\ &= u_0 F_0 \left(1 - \frac{\tau_0}{u_0}\right) \\ &\quad + \frac{\omega_0 F_0 \tau_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du \\ &\quad + r(u_0) \frac{\omega_0 F_0 \tau_0}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P(-u_0 \rightarrow u, \phi_0 \rightarrow \phi) du. \end{aligned} \tag{9.36}$$

Figure 9.19 provides a schematic of the first integral E_d :

$$\frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^1 P(u_0 \rightarrow u, \phi_0 \rightarrow \phi) du,$$

in which light from $\hat{\xi}_0$ is (single) scattered into various directions $\hat{\xi}$ and integrated over $d\Omega(\hat{\xi})$ constrained to downward directions ($\hat{\xi} \bullet \hat{e}_z > 0$). Reversing the direction the rays and using the symmetry of the phase function shows that this integral is identical to $p^{--}(|u_0|)$ in Eq. (9.28). Likewise, the second integral is identical to $p^{-+}(|u_0|)$. Thus,

$$\begin{aligned} t_E(\hat{\xi}_0) &= \frac{E_d(\tau_0, u_0, \phi_0)}{u_0 F_0} = 1 - \frac{\tau_0}{u_0} \left(1 - \omega_0 [p^{--}(|u_0|) - r(u_0)p^{-+}(|u_0|)]\right) \\ &\approx \exp \left[-\frac{\tau_0}{u_0} \left(1 - \omega_0 [p^{--}(|u_0|) - r(u_0)p^{-+}(|u_0|)]\right) \right]. \end{aligned} \tag{9.37}$$

The reader should note that this is identical to Eq. (9.30). This is no accident, it is demanded by the reciprocity principle. In fact for the exact as well as the single-scattering solution, $t_E(u_0) = t^*(u_0)$, and Figures 9.14-9.18 provide estimates of t_E as well as t^* , and show that most of the deviation of t_E from unity is due to Rayleigh scattering in the atmosphere. Since $p^{--} = p^{-+} = 1/2$ for Rayleigh scattering (for which $\omega_0 = 1$),

$$t_{Er}(\hat{\xi}_0) \approx \exp \left[-\frac{\tau_r}{2u_0} (1 + r(u_0)) \right].$$

On the basis of this discussion of t_E , and Figures 9.14-9.18 we can safely assume for the present discussion that $t_E \approx 1$ as a basis of computing $[\rho_w]_N$ from ρ_w or L_w .

Figure 9.20 provides $[\rho_w]_N$ at 443, 520, 550, and 670 nm as a function of the pigment concentration, i.e., the concentration Chlorophyll *a* plus its degradation product Phaeophytin *a*, in Case 1 waters (Chapter 5).¹³ The magnitude of $[\rho_w]_N$ should be compared to the magnitude of $\rho_r + \rho_a$ and ρ_a itself. Figures 9.7-9.13 are all for 412 nm: the shortest wavelength band on all existing ocean color sensors. This wavelength was chosen as it is the wavelength of maximum Rayleigh scattering and therefore maximum atmospheric optical depth, and provides the most severe test if the single scattering approximation. Note that $\rho_r \propto \lambda^{-4}$, so it decreases rapidly with wavelength and the single-scattering theory becomes increasingly more accurate as λ increases. The aerosol component ρ_a usually varies with wavelength proportional to λ^{-n} , where $n \lesssim 1$, so it is a much weaker function of wavelength. In Chapter 6 we found that L_u at 412 nm is similar in magnitude to that at 443 nm, so we take, for discussion purposes $[\rho_w(412)]_N \approx [\rho_w(443)]_N$. Thus, we see that $[\rho_w(412)]_N$ is at most only 20-25% of $\rho_r + \rho_a$ and the same order of magnitude as ρ_a , i.e., in the measured reflectance $\rho_t = \rho_r + \rho_a + t\rho_w$, the desired quantity ρ_w at low pigment concentrations is *at most* $\sim 20 - 25\%$ of the measured reflectance. At higher pigment concentration, the water “signal” is an even smaller fraction of the total. The required retrieval of ρ_w from ρ_t is what is usually called *atmospheric correction*, and is discussed in detail in Chapter 10.

9.6 Concluding Remarks

Here, we have defined the remote sensing problem — the retrieval of the water-leaving reflectance ρ_w from the top-of-atmosphere exiting radiance ρ_t , or equivalently, the estimation of ρ_{Other} and t . We then used single-scattering theory to separate ρ_{Other} into its component parts and to estimate the magnitudes of the various quantities and to compare them to one another, e.g., the magnitude of ρ_w compared to ρ_a . We examined the accuracy of the single-scattering approximation compared to exact solutions of the radiative transfer equation and concluded that while single-scattering theory is sufficiently accurate to provide a tool for understanding the important processes involved in ocean color remote sensing, its accuracy is marginal for the purpose of actually retrieving the water-leaving reflectance.¹⁴ Thus, accurate retrieval of ρ_w from ρ_t requires a full solution to the radiative

¹³Examination of Figure 9.20 for 520 and 550 nm suggests that at these wavelengths $[\rho_w]_N$ is approximately constant ($[\rho_w(520)]_N \approx 0.0075$ and $[\rho_w(550)]_N \approx 0.005$) for $C_P \lesssim 0.25$ mg/m³. The values of $[L_w]_N$ associated with these reflectances are called “clear-water radiances.” This concept — approximately constant radiances in the green for low C_P — is often referred to as the “clear-water radiance concept.”

¹⁴It is important to note that the single-scattering theory implies that ρ_{Other} is independent of the vertical distribution of the aerosol. The level of agreement with the exact computations for the M70 model and a *two-layer* atmosphere suggests that the full multiple-scattering solution is also independent of the aerosol’s vertical distribution. This, however, is only true for weakly absorbing aerosols ($\omega_{0a} \approx 1$). For strongly absorbing aerosols ($\omega_{0a} \lesssim 0.8$), their vertical distribution is critical, e.g., contrast ρ_{Other} computed for a

transfer equation, and is the subject of the next chapter. In the Appendix we showed that the single-scattering analysis of the polarization properties of ρ_{Other} agree reasonably well with their exact counterparts and can be used when less than exact computations are required.

9.7 Appendix: Single-Scattering Approximation to L_{Other} , Including Polarization

It is important to understand the polarization of the radiance exiting the top of the atmosphere because some ocean color sensors are sensitive to the polarization state of this radiance. Here we provide the single-scattering approximation to the atmospheric portion of this radiance $L_{\text{Other}} = L_r + L_a$ when polarization is included in the analysis.

Figure 9.1 shows the three single-scattering processes that contribute to the top-of-atmosphere radiance: (1) direct backscattering of solar radiance to the sensor without interaction with the sea surface; (2) forward scattering of solar radiation by the atmosphere followed by reflection from the sea surface and propagation to the sensor; and (3) reflection of solar radiation from the sea surface followed by forward scattering by the atmosphere and propagation to the sensor. These processes contribute radiance with different states of polarization.

Following the notation in Chapters 2 and 3, the scattering phase matrix (referred to the scattering plane) for spherical particles or for Rayleigh scattering can be written in the form

$$\mathbf{P}(\Theta) = \begin{pmatrix} P_{11}(\Theta) & P_{12}(\Theta) & 0 & 0 \\ P_{12}(\Theta) & P_{11}(\Theta) & 0 & 0 \\ 0 & 0 & P_{33}(\Theta) & -P_{34}(\Theta) \\ 0 & 0 & P_{34}(\Theta) & P_{33}(\Theta) \end{pmatrix},$$

where Θ is the scattering angle. Recalling that in the description of polarized radiance the Stokes vector is referenced to a particular coordinate system, a *natural* system for describing the polarization state of the radiance in the atmosphere. In this system we take the direction of the unit vector \hat{e}_ℓ (or $\hat{\ell}$) to be parallel to the plane formed by the vertical \hat{e}_z and the direction of propagation $\hat{\xi}$. Thus, to quantify the polarization of radiance scattered from one direction to another requires transforming the incident Stokes vector to the scattering-plane reference, where the phase matrix takes the simple form above, computing the scattered Stokes vector in this plane, and then transforming the scattered

totally absorbing aerosol ($\omega_{0a} = 0$) located in a thin layer at the sea surface with that computed when the aerosol is all in a thin layer at the top of the slab.

Stokes vector back to the natural system. These transformations are effected by the rotation matrix $\mathbf{R}(\psi)$ described in Chapter 2. The Stokes vector for radiance reflected from the sea surface is related to the incident Stokes vector by the Mueller matrix for reflection,

$$\mathbf{I}_{\text{Reflected}} = \mathbf{R}_f \mathbf{I}_{\text{Incident}},$$

where from Chapter 1,

$$\mathbf{R}_f = \begin{pmatrix} \rho_+ & \rho_- & 0 & 0 \\ \rho_- & \rho_+ & 0 & 0 \\ 0 & 0 & \rho_{33} & 0 \\ 0 & 0 & 0 & \rho_{33} \end{pmatrix},$$

with

$$\rho_{\pm}(\theta_i) = \frac{1}{2} \left[\frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} \pm \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} \right],$$

and

$$\rho_{33}(\theta_i) = -\frac{\tan(\theta_i - \theta_t) \sin(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t) \sin(\theta_i + \theta_t)}.$$

Here, the radiation is incident at an angle θ_i with respect to surface normal, and $\sin \theta_t = m_w^{-1} \sin \theta_i$. If the sea surface is flat, the surface normal is parallel to \hat{e}_z and the reference system for \mathbf{R}_f is the natural system.

To compute the Stokes vector for the three processes that contribute in first order, we need to determine the scattering angles and the rotation angles involved in the required $\mathbf{R}(\psi)$'s. The scattering angles are the same as those developed in the scalar case, i.e., Θ_{\pm} , so we really need only determine the various rotational transformations. Determination of the required transformations is most easily accomplished by displaying the vectors $\hat{\xi}_0$ and $\hat{\xi}$ on the unit sphere centered on our standard coordinate system for radiative transfer (z -axis directed *into* the atmosphere from the top). This is shown in Figure 9.21, in which the incident radiance from the sun is propagating in the direction $\hat{\xi}_0$ (point “A” on the unit sphere) and radiance formed by processes 1, 2, and 3 is exiting the top of the atmosphere in the direction $\hat{\xi}$ (point “B” on the unit sphere). The direction $\hat{\xi}_0$ is specified by the angles θ_0 and ϕ_0 (or equivalently u_0 and ϕ_0). Likewise, $\hat{\xi}$ is specified by the angles θ and ϕ (or equivalently u and ϕ). Note that the angle $\pi - \theta$ (rather than θ) is provided on the diagram. Also, note that the angle $\Delta\phi \triangleq \phi - \phi_0$ is *negative* for the directions $\hat{\xi}_0$ and $\hat{\xi}$ as drawn in the figure.

Now, we examine Figure 9.21, looking along the plane formed by \hat{e}_z and $\hat{\xi}_0$ (the incident plane in the natural system). This view is provided in Figure 9.22, where the line ZDAN is the great circle ZAN in Figure 9.21, and the arc ZBCN is the great circle ZBN: the labeled points are in the corresponding positions on the two figures. Focussing on Figure 9.22, the incident solar beam (striking the sphere at point A) can be redirected to point B

in three ways (corresponding to the three processes in Figure 9.1. First it can be directly backscattered in the atmosphere through an angle Θ_- toward point B (process 1, $A \rightarrow B$). Secondly, it can be scattered toward the sea surface through an angle Θ_+ , reaching point C on the sphere, and then reflected from the surface toward point B (process 2, $A \rightarrow C \rightarrow B$). Finally, it can be reflected from the sea surface to point D and then forward scattered through an angle Θ_+ to reach point B (process 3, $A \rightarrow D \rightarrow B$).

Determination of the various rotations required is straightforward given the figures. Radiance arriving at point A on the figure is unpolarized (from the sun), so there is no rotation required to place its reference in the scattering plane. Thus, for processes 1 and 2, the only transformation required is the rotation of the scattering plane reference through angles i_1 and i_2 , respectively, into the natural reference system. The figure shows that these rotations are both in the same sense: the scattering plane reference frame must be rotated *counter clockwise* and as such the rotation matrices are $\mathbf{R}(+i_1)$ and $\mathbf{R}(+i_2)$ for processes 1 and 2 respectively. No rotation is required upon Fresnel reflection from the sea surface as \mathbf{R}_f is already referred to the natural system. In the case of process 3, the radiance arriving at point D is polarized by reflection from the sea surface. Thus, at D the natural reference frame must be rotated into the scattering plane through an angle i_0 to compute the scattered Stokes vector, which is then referenced to the natural coordinate system by a rotation through an angle i_3 . Both of these rotations are *clockwise*, so the corresponding rotation matrices are $\mathbf{R}(-i_0)$ and $\mathbf{R}(-i_3)$.

The various rotation angles can be determined from the laws of spherical trigonometry. Applying the laws of cosine and sine to the spherical triangle shown in Figure 9.23, i.e.,

$$\cos I = \cos \theta_{S1} \cos \theta_{S2} + \sin \theta_{S1} \sin \theta_{S2} \cos i \quad (9.38)$$

and

$$\frac{\sin i}{\sin I} = \frac{\sin i_{S1}}{\sin \theta_{S1}} = \frac{\sin i_{S2}}{\sin \theta_{S2}}, \quad (9.39)$$

we have

$$\cos i = \frac{\cos I - \cos \theta_{S1} \cos \theta_{S2}}{\sin \theta_{S1} \sin \theta_{S2}},$$

and the signs of the individual angles i , i_{S1} , and i_{S2} , are the same as the signs of the opposite arcs I , θ_{S1} , and θ_{S2} , respectively. Using this, the rotation angles i_0 through i_3 are

seen to be given by

$$\begin{aligned}\cos i_0 &= \frac{\cos(\pi - \theta) - \cos \Theta_+ \cos \theta_0}{\sin \Theta_+ \sin \theta_0} \\ \cos i_1 &= \frac{\cos \theta_0 - \cos \Theta_- \cos \theta}{\sin \Theta_- \sin \theta} \\ \cos i_3 &= \frac{\cos(\theta_0) - \cos \Theta_+ \cos(\pi - \theta)}{\sin \Theta_+ \sin(\pi - \theta)} \\ i_2 &= i_3\end{aligned}$$

with

$$\text{Sign}(i_0) = \text{Sign}(i_1) = \text{Sign}(i_2) = \text{Sign}(i_3) = \text{Sign}(\Delta\phi).$$

Following the procedures in Appendix 1 to Chapter 2, the single-scattered Stokes vector can be written

$$\begin{aligned}\mathbf{I}_x(0, u, \phi) &= -\frac{\omega_x \tau_x}{4\pi u} \left[\mathbf{R}(+i_1) \mathbf{P}_x(\Theta_-) \right. \\ &\quad + \mathbf{R}_f(\theta_0) \mathbf{R}(+i_2) \mathbf{P}_x(\Theta_+) \\ &\quad \left. + \mathbf{R}(-i_3) \mathbf{P}_x(\Theta_+) \mathbf{R}(-i_0) \mathbf{R}_f(\pi - \theta) \right] \mathbf{I}_0,\end{aligned}$$

where

$$\mathbf{I}_0 = \begin{pmatrix} F_0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and $x = r$ or a .

Carrying out the matrix multiplication, we find that the Stokes vector components of the radiance exiting the top of the atmosphere:

$$\begin{aligned}
I(u, \phi) &= -\frac{\omega_x \tau_x F_0}{4\pi u} \left[P_{11}(\Theta_-) \right. \\
&\quad + \left(\rho_+(\theta) + \rho_+(\theta_0) \right) P_{11}(\Theta_+) \\
&\quad + \left(\cos(2i_2) \rho_-(\theta) + \cos(2i_0) \rho_-(\theta_0) \right) P_{12}(\Theta_+) \Big], \\
Q(u, \phi) &= -\frac{\omega_x \tau_x F_0}{4\pi u} \left[\cos(2i_1) P_{12}(\Theta_-) \right. \\
&\quad + \left(\rho_-(\theta) + \cos(2i_3) \cos(2i_0) \rho_-(\theta_0) \right) P_{11}(\Theta_+) \\
&\quad + \left(\cos(2i_2) \rho_+(\theta) + \cos(2i_3) \rho_+(\theta_0) \right) P_{12}(\Theta_+) \\
&\quad \left. - \sin(2i_3) \sin(2i_0) \rho_-(\theta_0) P_{33}(\Theta_+) \right], \\
U(u, \phi) &= -\frac{\omega_x \tau_x F_0}{4\pi u} \left[\sin(2i_1) P_{12}(\Theta_-) \right. \\
&\quad - \sin(2i_3) \cos(2i_0) \rho_-(\theta_0) P_{11}(\Theta_+) \\
&\quad + \left(\sin(2i_2) \rho_3(\theta) - \sin(2i_3) \rho_+(\theta_0) \right) P_{12}(\Theta_+) \\
&\quad \left. - \cos(2i_3) \sin(2i_0) \rho_-(\theta_0) P_{33}(\Theta_+) \right], \\
V(u, \phi) &= -\frac{\omega_x \tau_x F_0}{4\pi u} \left[-\sin(2i_0) \rho_-(\theta_0) P_{34}(\Theta_+) \right],
\end{aligned} \tag{9.40}$$

where,

$$\cos(\Theta_{\pm}) = \mp \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos \Delta \phi,$$

and the “ x ” on the elements of the phase matrix has been suppressed. Note that if polarization is neglected, the associated equation for the radiance is

$$I(u, \phi) = -\frac{\omega_x \tau_x F_0}{4\pi u} \left[P_{11}(\Theta_-) + \left(\rho_+(\theta) + \rho_+(\theta_0) \right) P_{11}(\Theta_+) \right],$$

so the neglect of polarization leads to an error in the radiance even in single scattering when a Fresnel-reflecting lower boundary is present.

Recall from Chapter 1 that the elements of the Stokes vector can be written in the form

$$\begin{aligned}
Q &= I_P \cos 2\chi \\
U &= I_P \sin 2\chi \cos \Delta \\
V &= I_P \sin 2\chi \sin \Delta,
\end{aligned}$$

where Δ is the phase difference between E_ℓ and E_r , $\tan \chi = |E_r|/|E_\ell|$, and $I_P = \sqrt{Q^2 + U^2 + V^2}$, the radiance of the *polarized* component. Thus, we can provide the state of polarization of

the reflected radiance by specifying χ , Δ , and the degree of polarization

$$P = \sqrt{\frac{Q^2 + U^2 + V^2}{I^2}},$$

where I is the sum of the radiances of the *polarized and unpolarized* components $I = I_P + I_U$.

Now we investigate the efficacy of the single-scattering solution for providing the polarization properties of the reflected radiation by comparing these quantities (P , χ , and Δ) with exact computations for a number of realistic atmospheres. The atmospheres will exhibit Rayleigh scattering and aerosol scattering, with the aerosols described by either the M70 or the M99 aerosol models. Figure 9.24 provides $P_{12}(\Theta)/P_{11}(\Theta)$ for these three components. This ratio is the degree of polarization P of light scattered from a sample of the indicated component.¹⁵ Note the large polarization indicated for the M99 model near the *rainbow angle* ($\Theta \approx 138^\circ$), where the aerosol scattering produces larger polarization than molecular scattering.

First, consider a purely Rayleigh-scattering, i.e., an aerosol-free, atmosphere. Figure 9.25 provides the comparison of P and χ computed using single-scattering theory with exact values for τ_r characteristic of earth's atmosphere at 412 and 865 nm, the lower and upper limits of the spectral region employed by many ocean color sensors. Clearly the single-scattering theory provides an excellent estimate of χ over the entire range of possible viewing angles and a reasonably accurate estimate of P for viewing angles up to $\sim 60^\circ$. The degree of polarization is always higher in the single-scattering computation – an indication that multiple scattering tends to *depolarize* the polarization induced at each scattering. The angle Δ is not provided because it is identically zero for both the single-scattering and the exact solutions, i.e., the radiance reflected from a pure Rayleigh-scattering atmosphere is *linearly* polarized.¹⁶

In contrast to a pure Rayleigh-scattering atmosphere, we now consider a purely aerosol-scattering atmosphere. Figure 9.26 provides examples of P and χ computed for the M70 and M99 aerosol models. Again, we see that single-scattering theory compares well with exact computations as long as the viewing angle is less than about 60° . As before, multiple scattering tends to depolarize the first order radiance.

Figures 9.27 and 9.28 provide similar comparisons for more realistic atmospheres with both Rayleigh and aerosol scattering present. Note that the Stokes vectors for the Rayleigh

¹⁵To verify this, compute P using Eqs. (9.40) with $\rho_+ = \rho_- = 0$.

¹⁶It is easy to verify that when $P_{34}(\Theta) = 0$, as in Rayleigh scattering, the V component of the Stokes vector satisfies a radiative transfer equation that is independent of the other components. Since the incident radiance from the Sun is unpolarized, no V component can ever be *generated* in a pure Rayleigh-scattering atmosphere.

and aerosol components are linearly additive in the single-scattering theory. These comparisons are similar to those above even though the total optical thickness is as high as 0.718 in the right panels of Figure 9.28.

Computations of Δ have not been provided for Figures 9.26 through 9.28 because its value is always close to zero, i.e., the polarization of radiation in the atmosphere is almost linear, even in the presence of aerosols. For example, Figure 9.29 provides Δ for the M70 aerosol model in the absence of Rayleigh scattering. The single-scattering computation of Δ is poor near its maximum, but reasonably good elsewhere. The main message of this figure is that $\Delta \approx 0$ is a good approximation to the exact computations, i.e., in this case the maximum value of Δ was $\sim 7^\circ$, so at the maximum $\cos \Delta \approx 0.9926$ and the U component can specified by χ alone (with $\Delta = 0$) to within 1%.

9.8 Figures

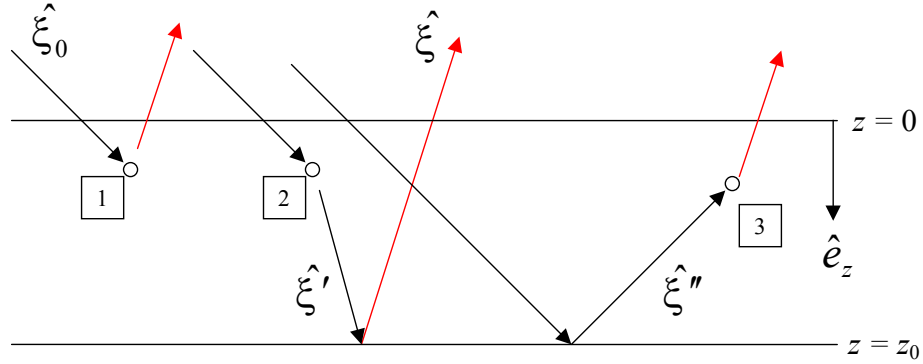


Figure 9.1: Schematic of the three processes involving a single interaction in the atmosphere that can contribute to $L_{\text{Other}}(0, \hat{\xi})$. The solar radiation enters from the top ($z = 0$) and is propagating in a direction specified by $\hat{\xi}_0$. The radiance exiting the top is propagating in the $\hat{\xi}$ direction. The color of the lines on the drawing change from black to red after the interaction. The boxed labels refer to the three processes described in the text. Recalling that in Fresnel reflection the surface normal bisects the angle between the incident and reflected propagation vectors, the intermediate vectors $\hat{\xi}'$ and $\hat{\xi}''$ are given by $\hat{\xi}' = \hat{\xi} - 2(\hat{\xi} \cdot \hat{e}_z)\hat{e}_z$ and $\hat{\xi}'' = \hat{\xi}_0 - 2(\hat{\xi}_0 \cdot \hat{e}_z)\hat{e}_z$, when the surface is flat. In addition, $u = -u'$ in process 2 and $u'' = -u_0$ in process 3.

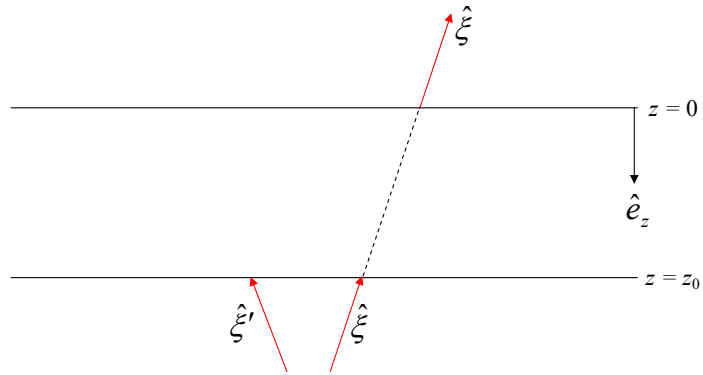


Figure 9.2: Radiance from the water enters the bottom of the atmosphere at $z = z_0$ from an arbitrary direction $\hat{\xi}'$. A sensor detects the radiance leaving the top of the atmosphere ($z = 0$) propagating in a particular direction $\hat{\xi}$.

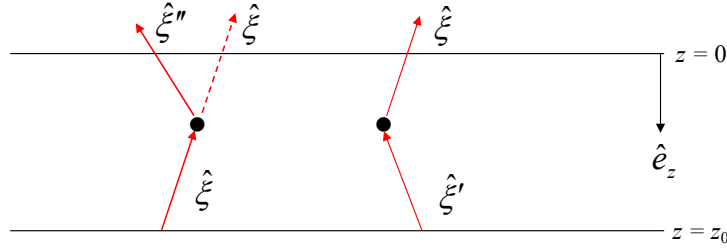


Figure 9.3: Schematic of the processes contributing to the two terms in Eq. (9.12). In the process on the left, radiance leaves the water propagating toward the sensor (direction $\hat{\xi}$). In the atmosphere it is scattered into an arbitrary direction $\hat{\xi}'' \neq \hat{\xi}$ resulting in a loss of radiance along the path from the surface to the sensor. This process represents the first term. On the right of the figure radiance exits the water in an arbitrary direction $\hat{\xi}' \neq \hat{\xi}$ and is scattered into $\hat{\xi}$ resulting in a gain in radiance at the sensor. This process is represented by the second term.

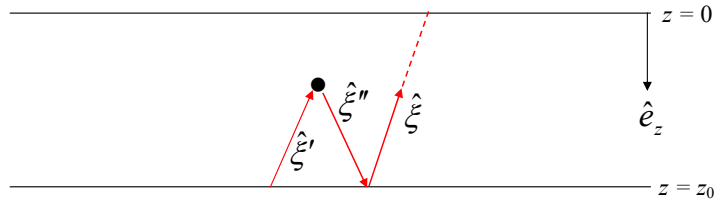


Figure 9.4: The first order process omitted from Eqs. (9.14) and (9.15). Radiance exits the water propagating in the direction $\hat{\xi}'$. It is then backscattered in a direction $\hat{\xi}''$ toward the surface, Fresnel-reflected into the direction $\hat{\xi}$, and subsequently propagates to the top of the atmosphere. When the water surface is flat, the law of reflection requires that $\hat{\xi}'' = \hat{\xi} - 2(\hat{\xi} \cdot \hat{e}_z)\hat{e}_z$.

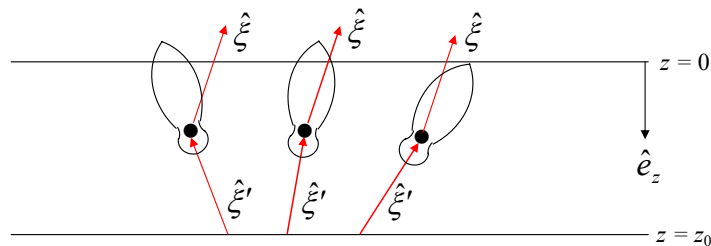


Figure 9.5: Processes involved in the integral in Eqs. (9.28). Radiance propagating in various directions $\hat{\xi}'$ is scattered into a particular direction $\hat{\xi}$. The curve on the figure represents a polar plot of a scattering phase function, as might for example be characteristic of aerosols, centered on the scattering location (the black dot). As $\hat{\xi}'$ sweeps over the upward hemisphere ($u' < 0$), the integral covers a large part of the “forward lobe” of the phase function.

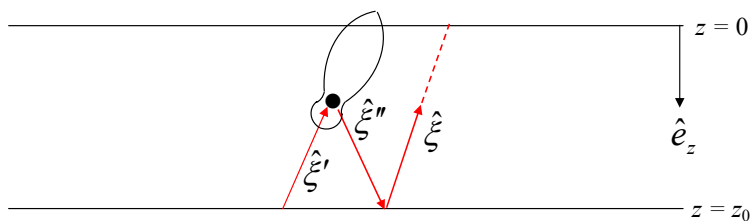


Figure 9.6: Processes involved in the integral in Eqs. (9.29). Radiance propagating in various directions $\hat{\xi}'$ is scattered into direction $\hat{\xi}''$ that is appropriate for Fresnel-reflection from the surface into a particular direction $\hat{\xi}$. (Note, $\hat{\xi}$ and $\hat{\xi}''$ are *fixed* directions for a *flat* surface.) The curve on the figure represents a polar plot of a scattering phase function. As $\hat{\xi}'$ sweeps over the upward hemisphere ($u' < 0$), the integral covers a large part of the “backward lobe” of the phase function.

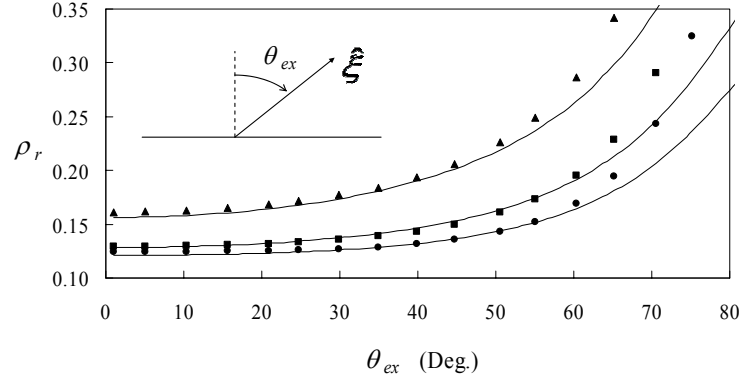


Figure 9.7: Comparison between ρ_r at 412 nm in the perpendicular plane computed in the single-scattering approximation (symbols) and the full scalar radiative transfer theory (curves). The angle the radiance exits the top of the atmosphere is θ_{ex} , where $\theta_{ex} = \pi - \theta = \cos^{-1}(|u|)$. The solar zenith angle $\theta_0 = 20^\circ$ (filled circles), 40° (filled squares), and 60° (filled triangles).

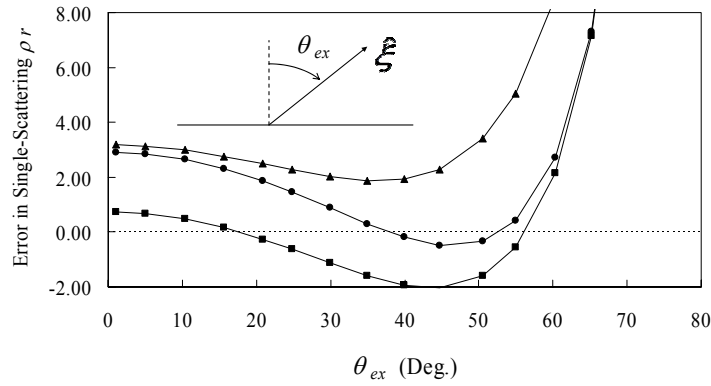


Figure 9.8: Error in ρ_r at 412 nm in the perpendicular plane (in %) computed using the single-scattering approximation. Positive error indicates that the single-scattering ρ_r is too high. Symbols have the same meaning as in Figure 9.7: $\theta_0 = 20^\circ$ (filled circles); 40° (filled squares); and 60° (filled triangles).

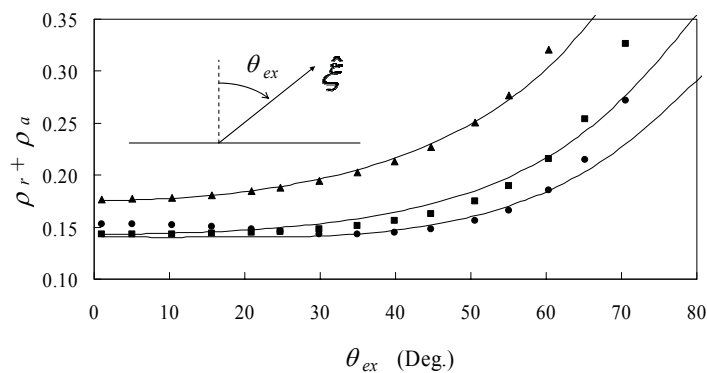


Figure 9.9: Comparison between $\rho_r + \rho_a$ at 412 nm in the perpendicular plane computed in the single-scattering approximation (symbols) and the full scalar radiative transfer theory (curves) for $\tau_{0a} = 0.2$. For this aerosol model (M70), ω_{0a} at 412 nm is 0.9853. Symbols have the same meaning as in Figure 9.7: $\theta_0 = 20^\circ$ (filled circles); 40° (filled squares); and 60° (filled triangles).

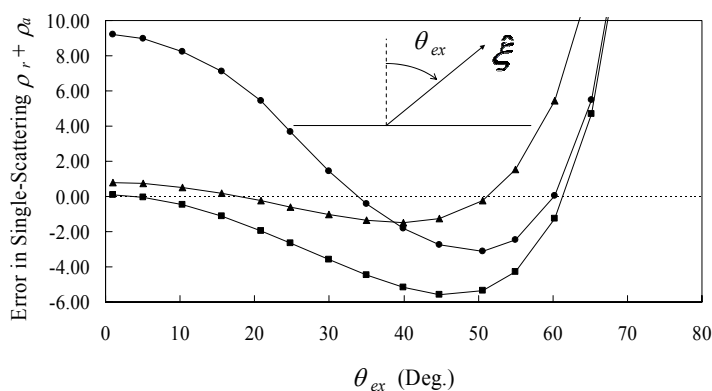


Figure 9.10: Error in $\rho_r + \rho_a$ (in %) at 412 nm in the perpendicular plane computed using the single-scattering approximation for $\tau_{0a} = 0.2$. Positive error indicates that the single-scattering $\rho_r + \rho_a$ is too high. Symbols have the same meaning as in Figure 9.7: $\theta_0 = 20^\circ$ (filled circles); 40° (filled squares); and 60° (filled triangles).

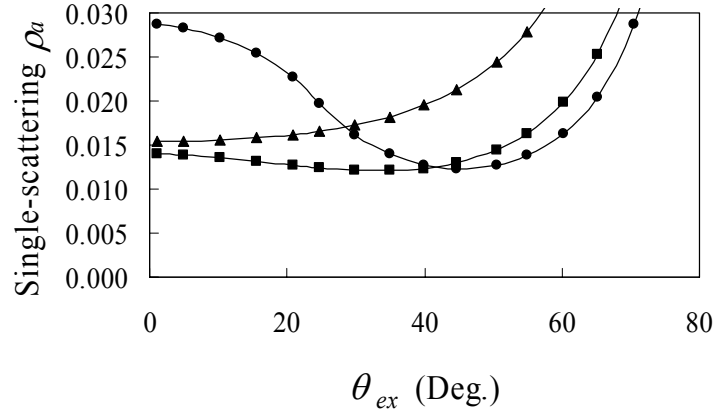


Figure 9.11: Computation of ρ_a at 412 nm in the perpendicular plane computed in the single-scattering approximation for $\tau_{0a} = 0.2$. Symbols have the same meaning as in Figure 9.7: $\theta_0 = 20^\circ$ (filled circles); 40° (filled squares); and 60° (filled triangles).

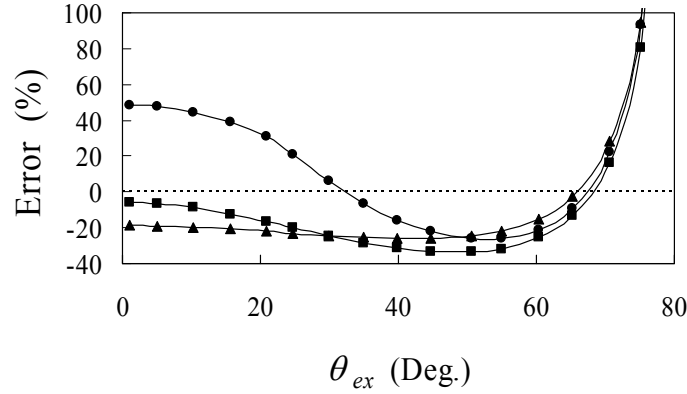


Figure 9.12: Error in the quantity $[(\rho_r + \rho_a) - \rho_r]$ at 412 nm in the perpendicular plane computed using the single-scattering approximation for $\tau_{0a} = 0.2$. Positive error indicates that the single-scattering $[(\rho_r + \rho_a) - \rho_r]$ is too high. Symbols have the same meaning as in Figure 9.7: $\theta_0 = 20^\circ$ (filled circles); 40° (filled squares); and 60° (filled triangles).

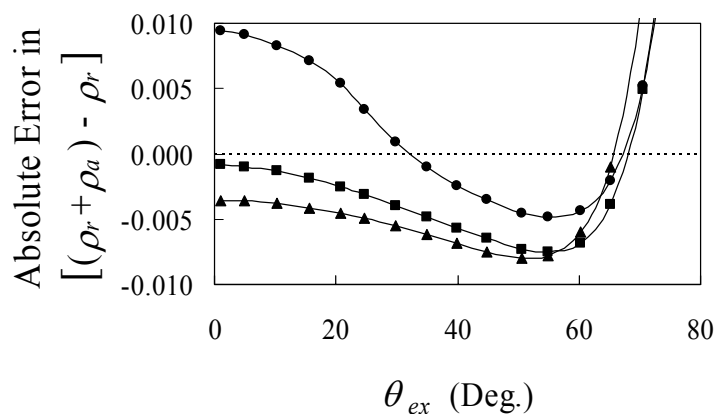


Figure 9.13: Absolute error in the quantity $[(\rho_r + \rho_a) - \rho_r]$ at 412 nm in the perpendicular plane computed using the single-scattering approximation for $\tau_{0a} = 0.2$. Positive error indicates that the single-scattering $[(\rho_r + \rho_a) - \rho_r]$ is too high. Symbols have the same meaning as in Figure 9.7: $\theta_0 = 20^\circ$ (filled circles); 40° (filled squares); and 60° (filled triangles).

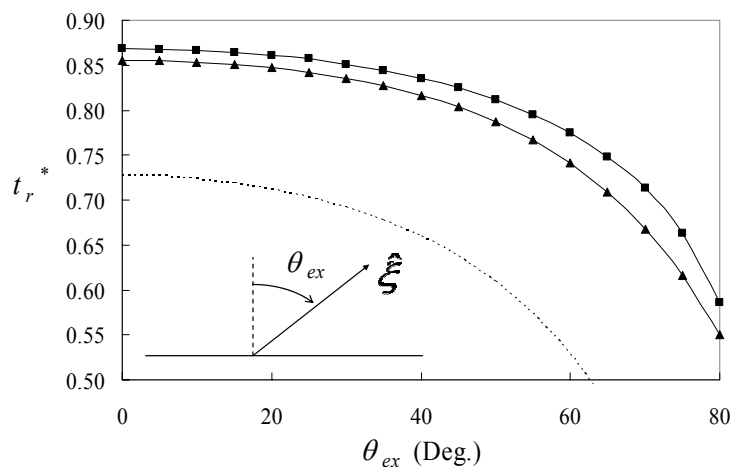


Figure 9.14: Comparison between the exact (filled squares) and the single-scattering (filled triangles) computations of $t_r^*(\tau_{0r} \rightarrow 0, \hat{\xi})$ at 412 nm. For comparison, the dotted line provides the direct transmittance T_r .

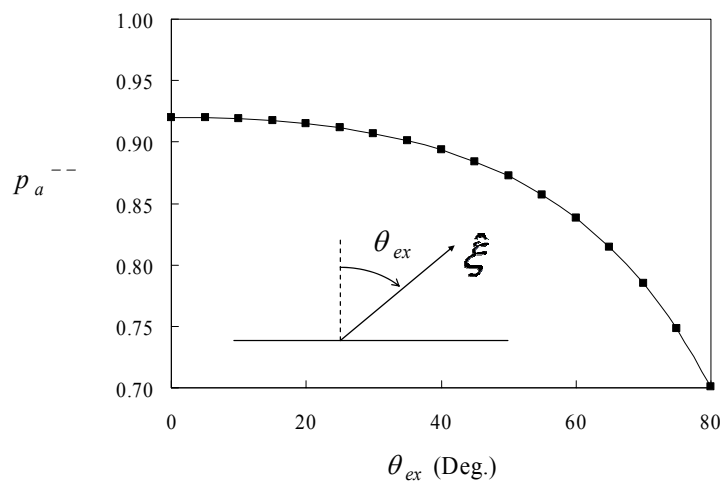


Figure 9.15: p_a^{--} as a function of the exiting angle for the M70 aerosol model.

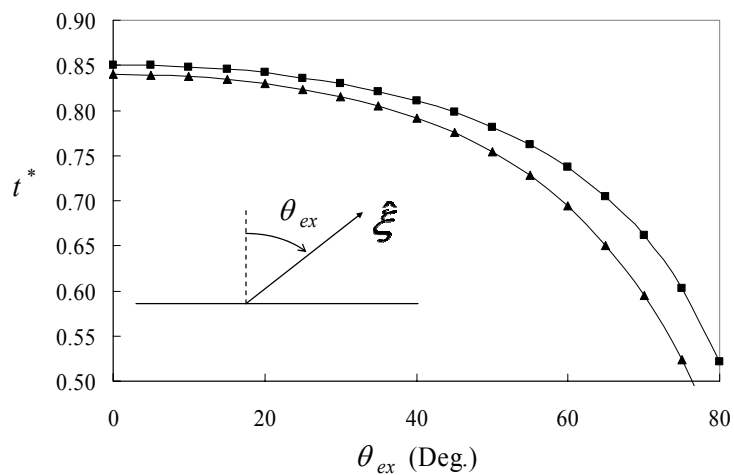


Figure 9.16: Comparison between the exact (filled squares) and the single-scattering (filled triangles) computations of $t^*(\tau_0 \rightarrow 0, \hat{\xi})$ at 412 nm with $\tau_{0a} = 0.2$.

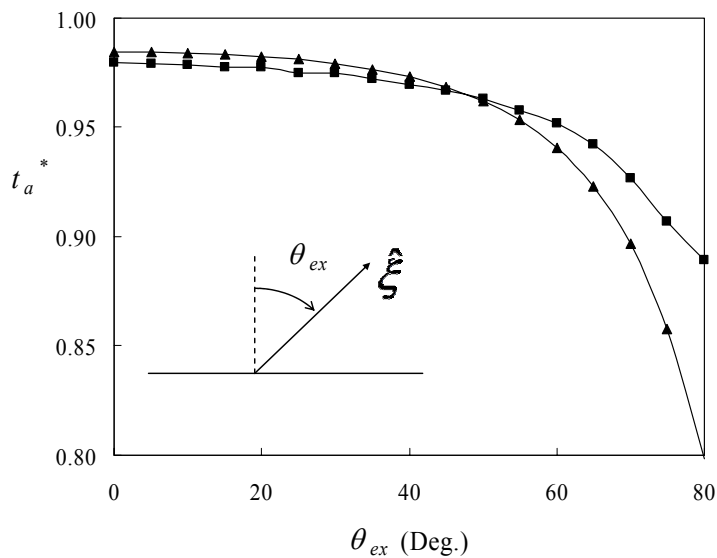


Figure 9.17: Comparison between t_a^* at 412 nm with $\tau_{0a} = 0.2$ computed using the single-scattering formulas (filled triangles) and its exact counterpart t^*/t_r^* (filled squares).

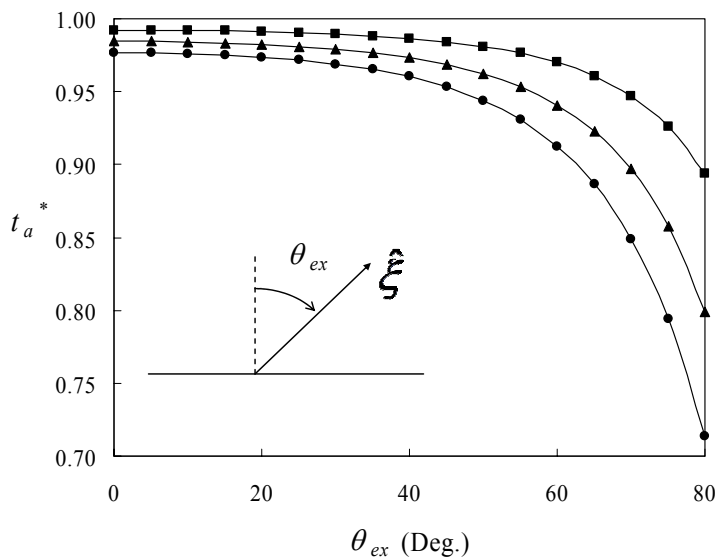


Figure 9.18: Influence of the aerosol optical thickness on the single-scattering computation of t_a^* . Filled squares — $\tau_{0a} = 0.1$, filled triangles — $\tau_{0a} = 0.2$, filled circles — $\tau_{0a} = 0.3$.

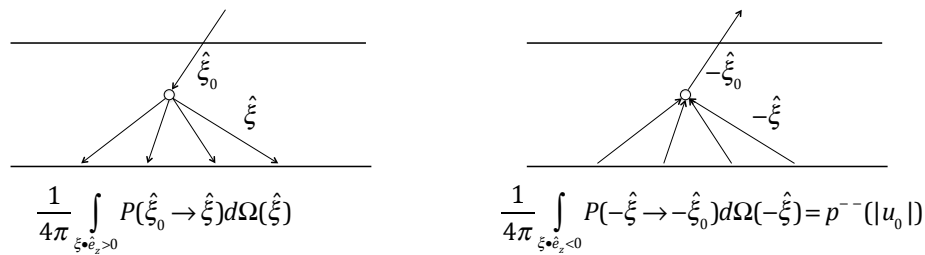


Figure 9.19: Left panel is a schematic showing the processes involved in the first integral in Eq. (9.36). Right panel shows the processes involved in the function $p^{--}(u_0)$ defined in Eq. (9.28). The two integrals are identical because $P(\hat{\xi}_0 \rightarrow \hat{\xi}) = P(-\hat{\xi} \rightarrow -\hat{\xi}_0)$.

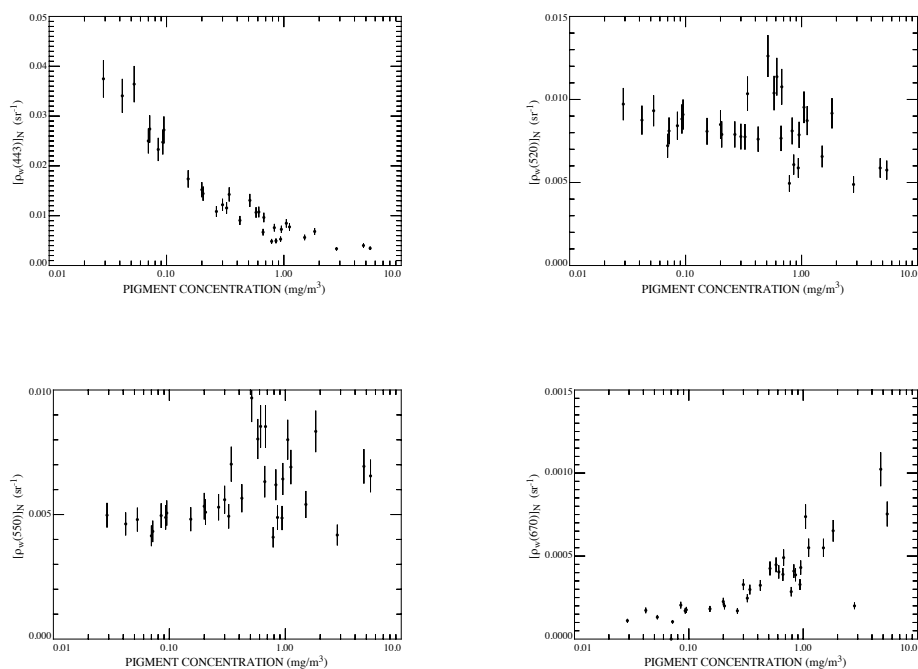


Figure 9.20: Normalized water-leaving reflectance at 443, 520, 550, and 670 nm, as a function of pigment concentration. Redrawn from figures in Gordon et al. [1988].

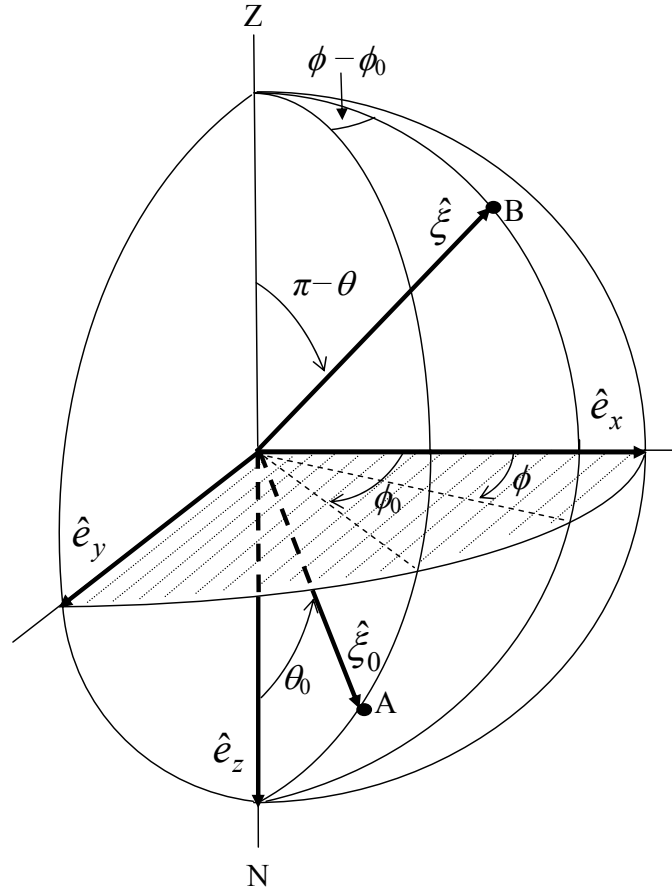


Figure 9.21: The directions $\hat{\xi}_0$ and $\hat{\xi}$ projected on the unit sphere. The relationships between the unit vectors and the angles are $\hat{\xi}_0 = \hat{e}_x \sin \theta_0 \cos \phi_0 + \hat{e}_y \sin \theta_0 \sin \phi_0 + \hat{e}_z \cos \theta_0$ and $\hat{\xi} = \hat{e}_x \sin \theta \cos \phi + \hat{e}_y \sin \theta \sin \phi + \hat{e}_z \cos \theta$.

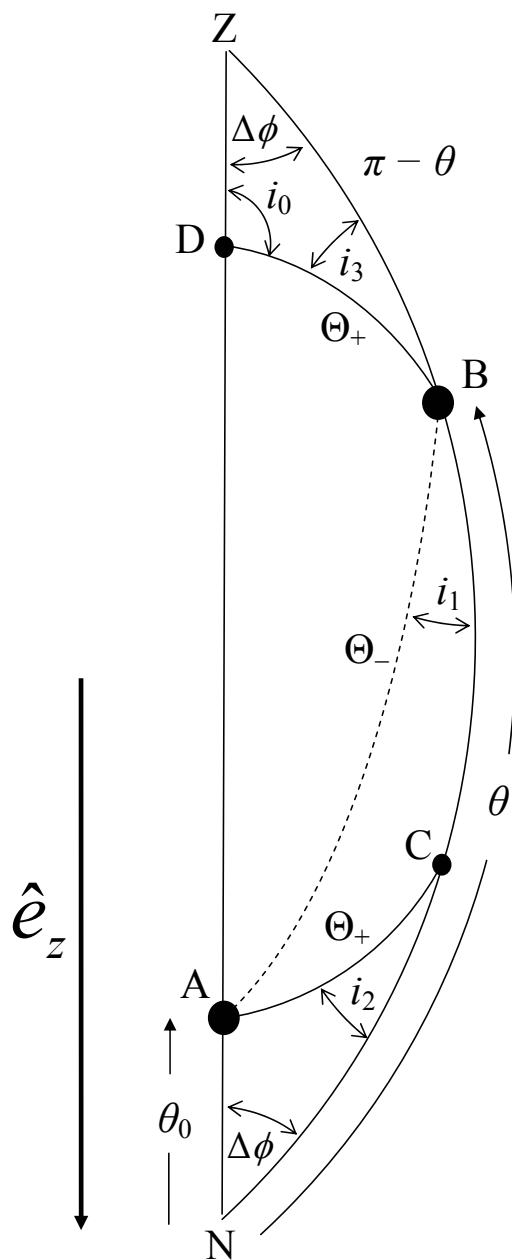


Figure 9.22: View of the unit sphere in Figure 9.21 looking toward the NAZ plane (NAZ plane is viewed “on edge”). This shows the scattering angles Θ_{\pm} and the various rotation angles (the i ’s) required for the three processes. The angle $\Delta\phi$ in this figure is the angle $\phi - \phi_0$ in Figure 9.21. Note that all the lines are arcs of great circles on the unit sphere.

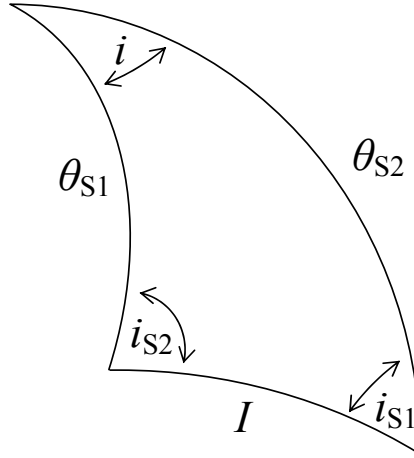


Figure 9.23: Spherical triangle used in the explanation of Eqs. (9.38) and (9.39).

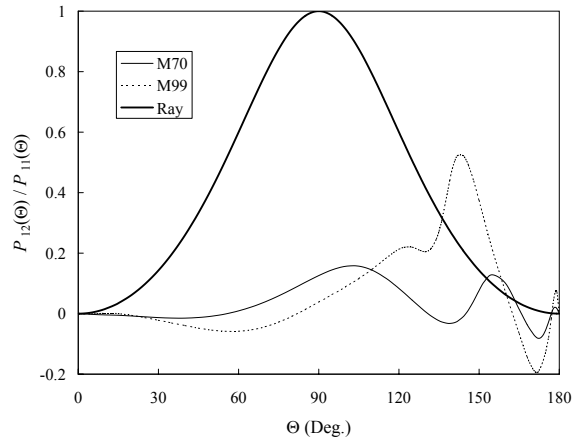


Figure 9.24: The degree of polarization of scattering of a beam of unpolarized light from aerosols described by the M70 and M99 aerosol models. Rayleigh scattering (multiplied by -1) is included for comparison.

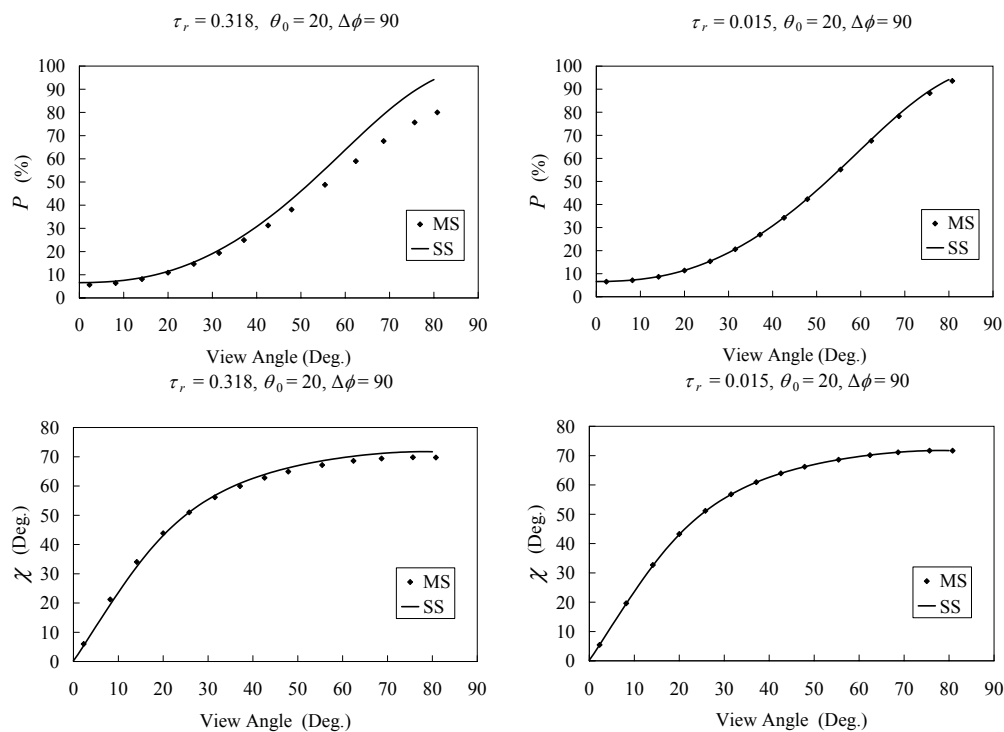


Figure 9.25: Comparison of P and χ as a function of the viewing angle $(\pi - \theta)$ of light scattered from a pure Rayleigh-scattering atmosphere using single-scattering theory (solid curves) with exact values (discrete points). The values of optical depth, solar zenith angle and relative azimuth are provided at the top of each panel. Values of τ_r of 0.318 and 0.015 correspond to wavelengths of 412 and 865 nm, respectively.

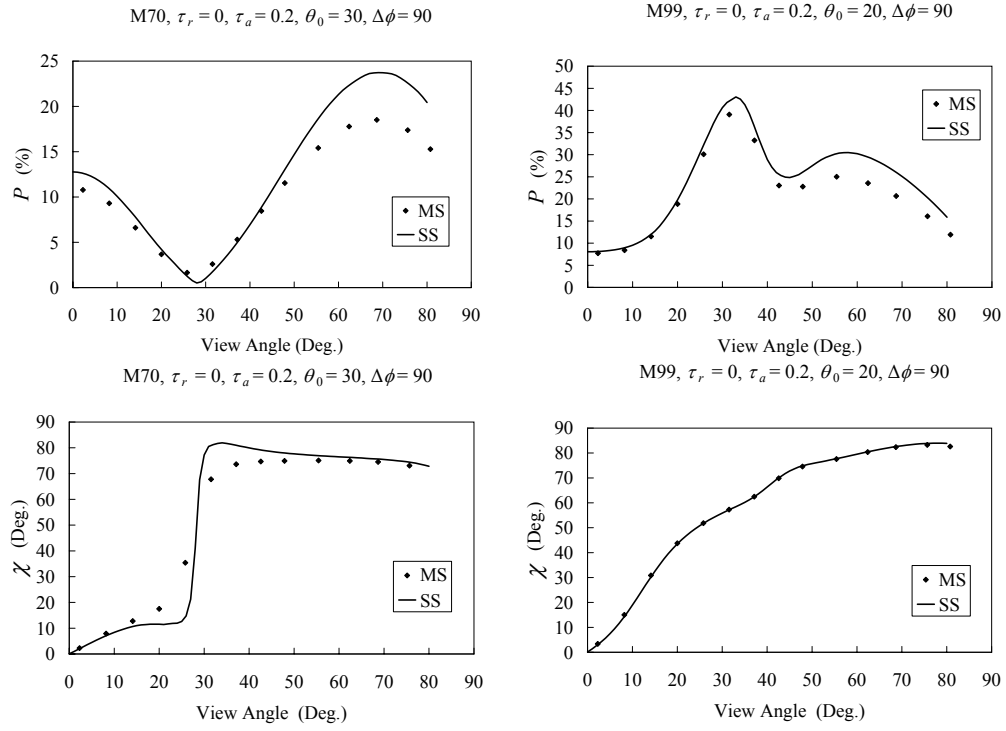


Figure 9.26: P and χ computed using single-scattering theory (solid curves) as a function of the viewing angle ($\pi - \theta$) for light scattered from an atmosphere composed purely of aerosols scattering according to the M70 (left panels) or M99 (right panels) aerosol models. Exact values (discrete points) are provided for comparison. The values of the aerosol optical depth, solar zenith angle and relative azimuth are provided at the top of each panel.

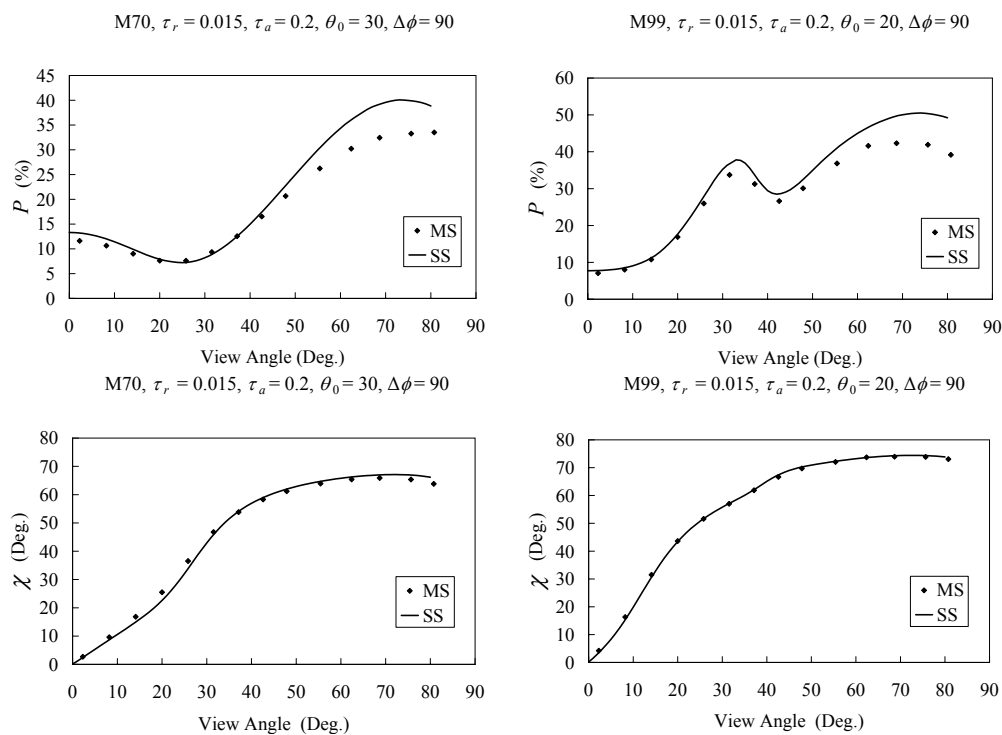


Figure 9.27: P and χ computed using single-scattering theory (solid curves) as a function of the viewing angle ($\pi - \theta$) for light scattered from earth's atmosphere at 865 nm if the aerosols scatter according to the M70 (left panels) or M99 (right panels) aerosol models. Exact values (discrete points) are provided for comparison. The values of the aerosol optical depth, solar zenith angle and relative azimuth are provided at the top of each panel.

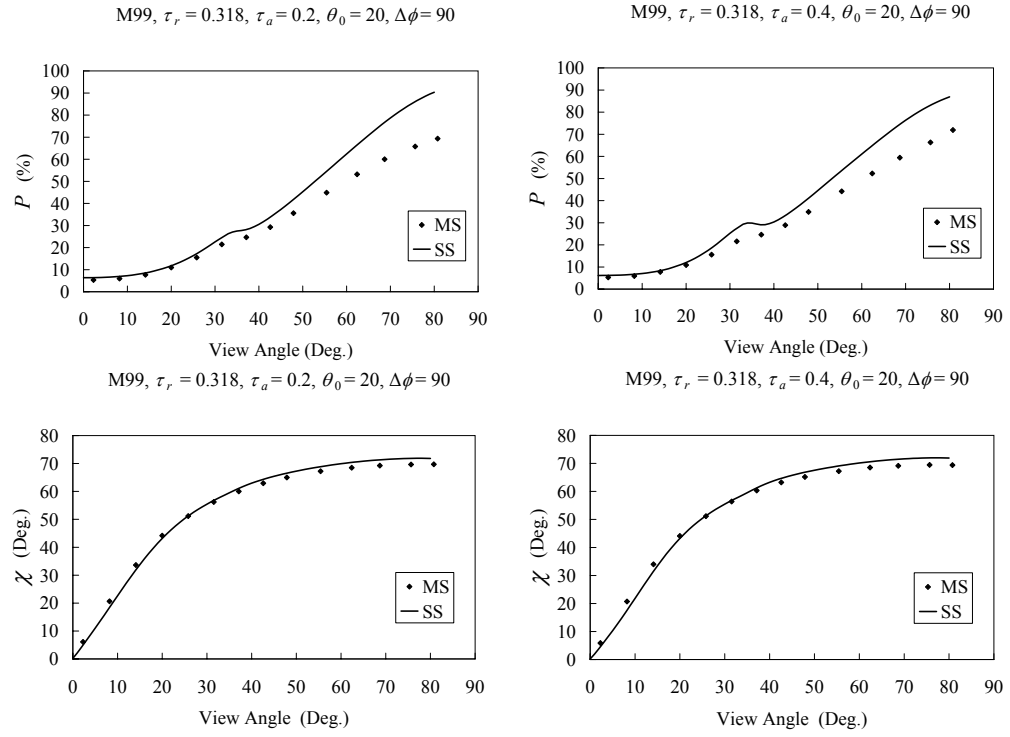


Figure 9.28: P and χ computed using single-scattering theory (solid curves) as a function of the viewing angle ($\pi - \theta$) for light scattered from earth's atmosphere at 412 nm if the aerosols scatter according to the M99 aerosol model computed using single-scattering theory (solid curves). Left panel is for $\tau_a = 0.2$ and right panel for $\tau_a = 0.4$. Exact values (discrete points) are provided for comparison. The values of the solar zenith angle and relative azimuth are provided at the top of each panel.

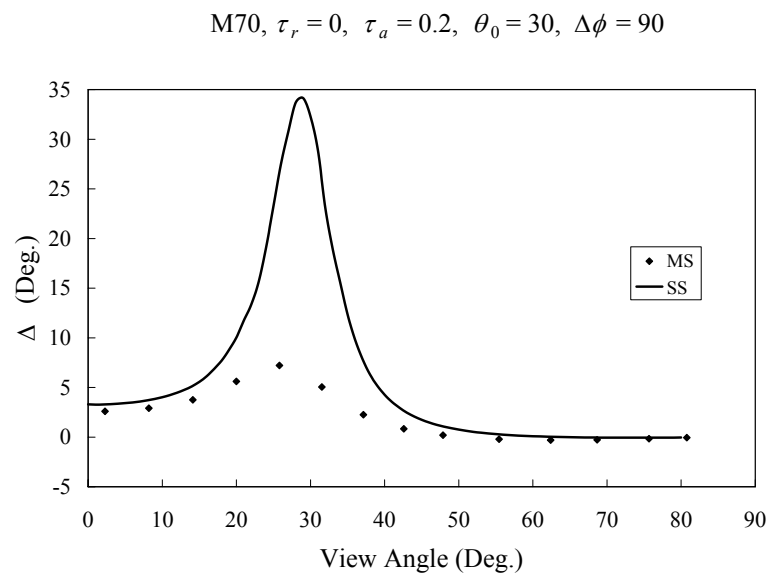


Figure 9.29: The angle Δ for the indicated aerosol model. If $\Delta = 0$, the radiation is linearly polarized, if $\Delta = 90^\circ$ and $\chi = 45^\circ$, the radiation is circularly polarized.

Chapter 10

Retrieval of Water-Leaving Radiance — Atmospheric Correction

10.1 Introduction

A principal objective of ocean color imagery is to study primary production, and its spatial and temporal variation, to better understand the ocean's role in the global carbon cycle. A required component in the estimation of primary productivity is the concentration of Chlorophyll *a*. Estimation of the concentration of Chlorophyll *a* from ocean color imagery requires the normalized water-leaving reflectance.¹ An example of how this is accomplished is provided by the CZCS. Figure 10.1 (left panel) displays a data set relating the pigment concentration (C_P) to $[\rho_w(443)]_N$ and $[\rho_w(550)]_N$. The data can be well represented by

$$\log_{10} C_P = -0.04 - 2.26 \log_{10} R + 3.03(\log_{10} R)^2 - 2.80(\log_{10} R)^3, \quad (10.1)$$

with $R = [\rho_w(443)]_N / [\rho_w(550)]_N$, i.e., the pigment concentration can be derived directly from the reflectance ratios.² Analysis suggests that, given error-free reflectances, the pigment concentration can be derived with an uncertainty of $\sim \pm 20\%$. Clearly, the normalized water-leaving reflectance plays a central role in the application of ocean color imagery.

¹Recall from Chapter 9 that the normalized water-leaving reflectance $[\rho_w]_N$ is related to the normalized water-leaving radiance $[L_w]_N$ by $[\rho_w]_N = \pi[L_w]_N / \bar{F}_0$, and related to the actual water-leaving reflectance ρ_w by $\rho_w = t_E[\rho_w]_N$, where t_E is the solar irradiance transmittance of the atmosphere.

²Other examples provided in Section 6.7.

In the preceding chapter, we learned that the water-leaving reflectance is at most 20% of the radiance received by a remote sensor in the blue part of the visible spectrum. Thus, we are faced with retrieval of this “signal” from that resulting from other processes such as scattering in the atmosphere by molecules and aerosols, absorption by atmospheric gases, and reflection of Sun and sky light from the water surface and from whitecaps. In principle such an effort is straightforward: use the known properties of the atmosphere and surface to estimate the effect of these processes and subtract them from the measured reflectance. Unfortunately, however, these properties are not constant in time: the molecular scattering in the atmosphere varies in proportion to the atmospheric pressure, causing variations in the Rayleigh scattering component of the signal; the aerosol concentration in the atmosphere undergoes significant variation in space and time; the absorption of atmospheric gases, e.g., Ozone, varies significantly; and of course the contribution of Sun glitter and whitecaps depends on the wind speed (and the wind direction). Modern remote sensing systems rely on the availability of data concerning surface atmospheric pressure, wind speed and direction, Ozone concentration, etc., for processing remotely-sensed reflectance. Such data are usually referred to as *ancillary data*, and are provided by sophisticated numerical models of atmospheric circulation and atmospheric properties. However, the most variable of these components is the aerosol concentration and optical properties, and they cannot be adequately estimated by modeling or by the sparse data sets provided by global aerosol measurement efforts such as AERONET. For example, a typical remotely sensed image may contain 10^6 individual pixels, each of which is of the order of $1 \text{ km} \times 1 \text{ km}$ in area. The aerosol contribution can vary significantly even over a single image of this size and such variation cannot be assessed by surface measurements. Thus, it is a significant challenge to assess the aerosol effect on the measured reflectance. Such an assessment *must* be made *utilizing the sensor measurements themselves*. This chapter concerns how this assessment can be carried out and, after correcting for whitecap contamination and gaseous absorption, the water-leaving radiance is retrieved.

We start by reviewing the remote sensing problem, then discuss an algorithm for determination of the aerosol effect based on the single-scattering approximation (SSA). As we learned in the last chapter the SSA provides a solution to the radiative transfer equation for the atmosphere, with an error of only a few %. The SSA algorithm is pedagogically simple and, as we shall see, actually works quite well, e.g., in the case of the CZCS. In addition, it is well suited to the spectral band set of more modern sensors: sensors with spectral bands in the near infrared (NIR), i.e., $\lambda > 700 \text{ nm}$, where the water-leaving radiance is negligible in most waters, so most of the radiance in the NIR is due to atmospheric effects. More importantly, it provides a framework for an algorithm utilizing full multiple-scattering solutions to the radiative transfer equation. Following discussion of the SSA algorithm, the extension to multiple scattering is developed. Finally, we consider in some detail the question of absorbing aerosols, and an algorithm capable of dealing with them.

10.2 Review of the Remote Sensing Problem

In Chapter 9 we divided the radiance $L_t(\lambda_i)$ in a spectral band centered at a wavelength λ_i , into two components: $L_{Other}(\lambda_i)$; and the water-leaving radiance transmitted from the surface to the top of the atmosphere (TOA), $tL_w(\lambda_i)$. We included only atmospheric scattering in $L_{Other}(\lambda_i)$, but now we include all of the processes that contribute to L_t :

$$L_t(\lambda_i) = L_{path}(\lambda_i) + t_g(\lambda_i)L_g(\lambda_i) + t_{wc}(\lambda_i)L_{wc}(\lambda_i) + t_w(\lambda_i)L_w(\lambda_i), \quad (10.2)$$

where $L_{path}(\lambda_i)$, is the radiance generated along the optical “path” by scattering in the atmosphere *and* by specular reflection of atmospherically-scattered light (skylight) from the sea surface (called $L_{Other}(\lambda_i)$ in Chapter 9); $L_g(\lambda_i)$ is the contribution arising from specular reflection of direct sunlight from the sea surface (Sun glitter); and L_{wc} and L_w (the desired quantity) are area-weighted averages of the radiance leaving whitecap-covered and whitecap-free areas of the surface, respectively. The quantities t_g , t_{wc} and t_w are the appropriate transmittances for glitter, whitecaps and water-leaving radiance respectively.³

Converting to reflectance, Eq. (10.2) becomes

$$\rho_t(\lambda_i) = \rho_{path}(\lambda_i) + t_g(\lambda_i)\rho_g(\lambda_i) + t_{wc}(\lambda_i)\rho_{wc}(\lambda_i) + t_w(\lambda_i)\rho_w(\lambda_i), \quad (10.3)$$

Thus, we must develop an algorithm that provides accurate estimates of $\rho_{path}(\lambda_i)$, $t_g(\lambda_i)\rho_g(\lambda_i)$, $t_{wc}(\lambda_i)\rho_{wc}(\lambda_i)$, and $t_w(\lambda_i)$, to retrieve $\rho_w(\lambda_i)$ from $\rho_t(\lambda_i)$. Near the Sun’s glitter pattern $t_g(\lambda_i)\rho_g(\lambda_i)$ is so large that the imagery is virtually useless and is usually discarded. Away from the glitter pattern $t_g(\lambda_i)\rho_g(\lambda_i)$ becomes negligibly small, and the largest, and most difficult to estimate, of the remaining terms is $\rho_{path}(\lambda_i)$. This difficulty is due to the aerosol component by virtue of its highly variable concentration and optical properties. So, we will concentrate on this term first, and consider the rest later.

Formally, ρ_{path} can be decomposed into several components:

$$\rho_{path} = \rho_r(\lambda) + \rho_a(\lambda) + \rho_{ra}(\lambda) \quad (10.4)$$

³For the water-leaving radiance, the appropriate t_w is diffuse transmittance t developed in Chapter 9, but refined here in Section 10.4.4. Assuming whitecaps are Lambertian reflectors, $t_{wc} = t^*$ of Chapter 9. In the case of the direct Sun glitter, when the wind speed is low the angular distribution of the radiance is approximately a Dirac delta function and t_g is close to the direct transmittance:

$$t_g \approx T = \exp \left[-\frac{\tau}{|u|} \right],$$

where $\tau = \tau_r + \tau_a + \dots$ is the total optical thickness of the atmosphere. In contrast, at high wind speed, L_g is more diffuse and the appropriate glitter transmittance is somewhere between T and t^* . Not knowing the appropriate transmittance for the glitter (although it could be estimated using the Cox-Munk model given the surface wind speed) is unimportant because the glitter radiance is so high that viewing directions with significant glitter must be discarded anyway.

where ρ_r is the reflectance resulting from *multiple scattering* by air molecules (Rayleigh scattering) in the absence of aerosols, and ρ_a is the reflectance resulting from *multiple scattering* by aerosols in the absence of the air. The term ρ_{ra} accounts for the interaction between Rayleigh and aerosol scattering, e.g., photons first scattered by the air then scattered by aerosols, or photons first scattered by aerosols then air, etc. This term is zero in the SSA, in which photons are only scattered once. It can be ignored as long as the amount of multiple scattering is small, i.e., at small Rayleigh and aerosol optical thicknesses. Although it may seem awkward when multiple scattering is important, the above separation is still useful because, given the surface atmospheric pressure (to determine the value of τ_r) and the surface wind speed (to define the roughness of the sea surface), ρ_r can be computed accurately, even accounting for polarization by scattering and by surface reflection.

To proceed we need an understanding of the basic relationship between the measured reflectance ρ_t and the desired ρ_w . In Chapter 7 we found that $\rho_w = t_E[\rho_w]_N$, where t_E is the irradiance transmittance of the atmosphere and $[\rho_w]_N$ is the normalized water-leaving reflectance. Table 10.1 provides typical values for ρ_t and $[\rho_w]_N$. It is developed for a most-

Table 10.1: Typical values of the reflectance (ρ_t) measured at the sensor and the normalized water-leaving reflectance $[\rho_w]_N$ for open ocean waters with low pigment concentrations. The values of ρ_t are for a solar zenith angle of 60° and viewing near the edge of the scan for a typical ocean color sensor.

λ (nm)	ρ_t (sr ⁻¹)	$[\rho_w]_N$ (sr ⁻¹)
412	0.34	0.040
443	0.29	0.038
488	0.23	0.024
531	0.19	0.0090
551	0.154	0.0040
670	0.105	0.0004
681	0.105	0.0003
748	0.081	—
869	0.069	—

challenging geometric viewing scenario: large solar zenith angle and large viewing angle ($\pi - \theta_v \approx 45^\circ$ and $\phi_v - \phi_0 = 90^\circ$).⁴ This represents viewing near the edge of the scan for a typical scanning radiometer. The geometry is most challenging because of the long path of both Sun light and backscattered light through the atmosphere. The wavelengths

⁴It is awkward to keep working with the viewing angle θ_v measured with respect to the *downward* normal. Thus, we will often use the supplement of this angle $\theta_v^{(s)} = \pi - \theta_v$, i.e., the angle of propagation of the viewed radiance measured with respect to the *upward* normal.

provided are those for MODIS. Clearly, the water “signal” is significant in the blue and negligible in the NIR. If one desires to retrieve information concerning the atmosphere by itself, the data in Table 10.1 suggest that this can be effected in the red and NIR. Thus, in our development of a ρ_w retrieval algorithm, we will assume that the sensor has two NIR spectral bands λ_s and λ_l with $\lambda_s < \lambda_l$ for which $\rho_w = 0$ (the subscripts “s” and “l” stand for short and long, respectively). These will provide the needed information concerning the state of the atmosphere. The problem we are required to solve can then be stated in a simple manner: in the absence of whitecaps and sun glitter, given the satellite measurement of the reflectance of the ocean-atmosphere system in the NIR, predict the reflectance ρ_t that would be observed in the visible if ρ_w were zero. The difference between the predicted and the measured reflectance of the ocean-atmosphere system is the water-leaving reflectance transmitted to the top of the atmosphere, $t_w\rho_w$.

Before continuing with the development of the algorithm for effecting the retrieval of ρ_w , it is important to assess the accuracy required for the procedure described in the previous paragraph to be successful. Lets assume that we will use the “blue-green” algorithm Eq. (10.1) from Figure 10.1 (left panel) to estimate the pigment concentration (C_P). More modern algorithms actually do not use such a single ratio as $R = [\rho_w(443)]_N / [\rho_w(550)]_N$ to estimate C_P throughout the entire range of pigment concentrations because, as we shall see below, $[\rho_w(443)]_N$ becomes too small at moderate values of C_P (Figure 10.1) to be retrieved with sufficient accuracy. Most algorithms now use 3 or 4 bands, e.g., 443, 490, 520, and 555 nm, wherein the ratio $[\rho_w(443)]_N / [\rho_w(555)]_N$ is used if $[\rho_w(443)]_N > [\rho_w(490)]_N$, which is true for low C_P . When C_P is large enough that $[\rho_w(443)]_N < [\rho_w(490)]_N$, then the ratio $[\rho_w(490)]_N / [\rho_w(555)]_N$ is used as long as $[\rho_w(490)]_N > [\rho_w(520)]_N$, etc. Nevertheless, to assess the required accuracy, Eq. (10.1) is sufficient. Using this we find the relative error in C_P is given by

$$\frac{\Delta C_P}{C_P} = \left[-2.26 + 2 \times 3.08 \log_{10} R - 3 \times 2.8 (\log_{10} R)^2 \right] \frac{\Delta R}{R},$$

with

$$\frac{\Delta R}{R} = \frac{\Delta[\rho_w(443)]_N}{[\rho_w(443)]_N} - \frac{\Delta[\rho_w(550)]_N}{[\rho_w(550)]_N}.$$

Here we have only one equation, but 2 unknowns: $\Delta[\rho_w(443)]_N$ and $\Delta[\rho_w(550)]_N$. That is, a given relative error in C_P depends on the relative errors at two wavelengths. Because of the extrapolation procedure from the NIR to the visible that will be used in the algorithm to estimate the aerosol contribution, we can expect the errors $\Delta[\rho_w(443)]_N$ and $\Delta[\rho_w(550)]_N$ to be correlated: in fact usually they will have the same sign. We will see in the next section that the errors are approximately related in the following manner:

$$\Delta[\rho_w(443)]_N \approx \gamma \Delta[\rho_w(550)]_N,$$

where $\gamma \approx 1/3$. This provides the needed second equation yielding

$$\frac{\Delta C_P}{C_P} \approx [-2.26 + 2 \times 3.08 \log_{10} R - 3 \times 2.8(\log_{10} R)^2] [1 - \gamma R] \times \frac{\Delta[\rho_w(443)]_N}{[\rho_w(443)]_N}.$$

From the data in Figure 10.1 (right panel), we can relate $[\rho_w(443)]_N$ directly to C_P through

$$[\rho_w(443)]_N \approx 0.0075 C_P^{-0.48},$$

so

$$\begin{aligned} \frac{\Delta C_P}{C_P} \approx & \left\{ \left(\frac{C_P^{+0.48}}{0.0075} \right) [-2.26 + 2 \times 3.08 \log_{10} R \right. \\ & \left. - 3 \times 2.8(\log_{10} R)^2] [1 - \gamma R] \right\} \times \Delta[\rho_w(443)]_N. \end{aligned} \quad (10.5)$$

To use Eq. (10.5) we need to invert Eq. (10.1), which is easy to accomplish (numerically). Figure 10.2 provides the factor in the curly brackets (“Factor”) in Eq. (10.5) as a function of C_P . Note that for a fixed $\Delta[\rho_w(443)]_N$ the magnitude of the error grows almost linearly with increasing C_P . Now, $[\rho_w(443)]_N$ becomes smaller than $[\rho_w(490)]_N$ for $C \approx 0.6 \text{ mg/m}^3$, so this algorithm would typically be used only up to this concentration. That means that the factor in the curly brackets is at most ~ -110 . So if we want the error in atmospheric correction to be roughly equivalent to the error in the bio-optical algorithm due to natural variability, i.e., $\sim 20\%$, then we require $|\Delta[\rho_w(443)]_N| \lesssim 0.002$. In that case, in a root-mean-square sense, the total error – natural plus atmospheric correction – would be at most $\sim 30\%$. Thus, we take the *goal* of the water-leaving radiance retrieval/atmospheric correction algorithm to be $|\Delta[\rho_w(443)]_N| \lesssim 0.002$. In clear ocean water (low C_P) this implies that the relative error in $[\rho_w(443)]_N$ in such cases should be no more than 5%.

10.3 A Single-Scattering Solution

We begin by approximating $\rho_{path}(\lambda_i)$ in the the limit that the optical thickness of the atmosphere is $\ll 1$, i.e., the single-scattering limit. Reflectances in this limit were developed in Chapter 9. The path reflectance is

$$\rho_{path}(\lambda_i) = \rho_r(\lambda_i) + \rho_{as}(\lambda_i), \quad (10.6)$$

where the aerosol contribution ρ_{as} is provided by⁵

$$\rho_{as}(\lambda) = \frac{\omega_a(\lambda)\tau_a(\lambda)p_a(\theta_v, \phi_v; \theta_0, \phi_0; \lambda)}{4 \cos \theta_v \cos \theta_0}, \quad (10.7)$$

⁵We have added the subscript “s” to the subscript “a” (for aerosol) to emphasize that this is *single* scattering for the aerosol. Note that in the SSA the Rayleigh scattering contribution ρ_r is given by the same formula, but with the subscripts “a” replaced by “r.”

$$p_a(\theta_v, \phi_v; \theta_0, \phi_0; \lambda) = P_a(\Theta_-, \lambda) + \left(r_f(\theta_v^{(s)}) + r_f(\theta_0) \right) P_a(\Theta_+, \lambda),$$

$$\cos \Theta_{\pm} = \mp \cos \theta_0 \cos \theta_v + \sin \theta_0 \sin \theta_v \cos(\phi_v - \phi_0),$$

where $P_a(\Theta, \lambda)$ is the aerosol scattering phase function for a scattering angle Θ , ω_a is the aerosol single scattering albedo, and $r_f(\alpha)$ is the Fresnel reflectance of the interface for an incident angle α , or in the case of $r_f(\theta_v^{(s)})$ for an incident angle of $\theta_v^{(s)} = \pi - \theta_v$. As usual, the angles represent the spherical coordinates of the propagation directions of the solar beam ($\hat{\xi}_0$) and of the light being viewed ($\hat{\xi}_v$), i.e., the remote radiometer is aimed in the direction $-\hat{\xi}_v$. Note that except for the angles and the Fresnel reflectances, all of the quantities depend on wavelength. We have provided examples of the accuracy of the single-scattering approximation in Chapter 9. In particular, the reader should examine the figures comparing the exact value of $\rho_{\text{path}} - \rho_r = (\rho_r + \rho_a + \rho_{ra}) - \rho_r$ with ρ_{as} (Figures 9.11 through 9.13). These show that, at least at the shorter wavelengths in the visible, e.g., 412 nm, accurate estimation of $\rho_{\text{path}}(\lambda_i)$ requires the complete solution to the radiative transfer equation, i.e., the inclusion of multiple scattering. Still, the SSA can serve as a roadmap to a more accurate approach.

Following the approach described above, we assume we are given the path reflectance at two bands in the NIR: λ_s and λ_l , where for SeaWiFS $\lambda_s = 765$ nm and $\lambda_l = 865$ nm, and for MODIS $\lambda_s = 748$ nm and $\lambda_l = 869$ nm. Given estimates of the surface atmospheric pressure and the wind speed, $\rho_r(\lambda)$ can be computed precisely and therefore $\rho_{as}(\lambda_s)$ and $\rho_{as}(\lambda_l)$ can be determined from the measurements of ρ_{path} at λ_s and λ_l .⁶ The parameter $\varepsilon(\lambda_s, \lambda_l)$, defined by

$$\varepsilon(\lambda_s, \lambda_l) \triangleq \frac{\rho_{as}(\lambda_s)}{\rho_{as}(\lambda_l)} = \frac{\omega_a(\lambda_s)\tau_a(\lambda_s)p_a(\theta_v, \phi_v; \theta_0, \phi_0; \lambda_s)}{\omega_a(\lambda_l)\tau_a(\lambda_l)p_a(\theta_v, \phi_v; \theta_0, \phi_0; \lambda_l)}, \quad (10.8)$$

can then be estimated. Assume for the moment that we can determine the value of $\varepsilon(\lambda_i, \lambda_l)$ for a spectral band at λ_i from $\varepsilon(\lambda_s, \lambda_l)$, e.g., through *extrapolation*, then $\rho_{as}(\lambda_i) = \varepsilon(\lambda_i, \lambda_l)\rho_{as}(\lambda_l)$, and

$$\begin{aligned} t_w(\lambda_i)\rho_w(\lambda_i) &= \rho_t(\lambda_i) - \rho_r(\lambda_i) - \varepsilon(\lambda_i, \lambda_l)\rho_{as}(\lambda_l) \\ &= \rho_t(\lambda_i) - \rho_r(\lambda_i) - \varepsilon(\lambda_i, \lambda_l)[\rho_t(\lambda_l) - \rho_r(\lambda_l)]. \end{aligned}$$

This shows that a key element to finding $\rho_w(\lambda_i)$ is the accurate extrapolation of $\varepsilon(\lambda_i, \lambda_l)$ from $\varepsilon(\lambda_s, \lambda_l)$.⁷

⁶Since we are ignoring Sun glitter, this assumes that $t_{wc}(\lambda_i)\rho_{wc}(\lambda_i)$ has also been provided.

⁷It is important to note that p_a in the definition of $\varepsilon(\lambda_s, \lambda_l)$ is *not* $P_a(\Theta_-)$ as has implicitly assumed by some authors, i.e., it involves both forward, $P_a(\Theta_+)$, and backward, $P_a(\Theta_-)$, scattering. Although $P_a(\Theta_-)$ is dominant, because $P_a(\Theta)$ is strongly peaked in near-forward directions (Chapter 4) the surface-reflected term $P_a(\Theta_+)$ can make a significant contribution to $\rho_{as}(\lambda)$, i.e., as much as 30% in some geometries.

10.3.1 Application to the CZCS

The CZCS was the first ocean color scanner in orbit. It had spectral bands at 443, 520, 550, and 670 nm. The earliest atmospheric correction algorithm for CZCS was based on single scattering. Because there were no NIR bands on CZCS,⁸ the algorithm could not be operated as described in the previous section. Table 10.1 shows that $\rho_w(670)$ can generally be taken to be zero if the pigment concentration (C_P) is sufficiently low, therefore the single scattering algorithm was typically operated with $\lambda_l = 670$ nm, and it was assumed $\rho_w(\lambda_l) = 0$. On CZCS, there was no shorter wavelength (λ_s) for which $\rho_w = 0$; however for $C_P \lesssim 0.25$ mg/m³, Figure 9.20 suggests that $[\rho_w(550)]_N$ is approximately a known constant, so $t^* \rho_w(550)$ can be estimated and subtracted from $\rho_t(550) - \rho_r(550)$ to yield $\rho_{as}(550)$.⁹ This can be used along with $\rho_{as}(670)$ to estimate $\varepsilon(550, 670)$ for “clear water” regions in a scene, i.e., regions for which $C_P \lesssim 0.25$ mg/m³, providing a basis for extrapolation to 520 and 443 nm by assuming the ε varies in a known manner with λ . Then it was further assumed that the resulting $\varepsilon(550, 670)$ was valid for the entire image, so retrieval of $[\rho_w(\lambda_i)]_N$ at 443, 520, and 550 could be effected for the entire image. Figure 10.3 provides an example of atmospheric correction of the CZCS blue and green bands in the Middle Atlantic Bight. Here, $\varepsilon(\lambda, \lambda_l)$ was taken to be proportional to $(\lambda_l/\lambda)^n$, where n was determined from $\varepsilon(550, 670)$ in the “clear water” region. Note that the intense haze layer seen crossing the diagonal in the images of L_t is absent from the images of L_w , revealing rich underlying horizontal structure in water-leaving radiance. Although, there are difficulties applying this procedure routinely, e.g., (1) the image of interest may contain no “clear water,” (2) the ε ’s may vary over the image because of variations in aerosol type, or (3) C_P may not be small enough to take $\rho_w = 0$ at 670 nm, Figure 10.3 does indicate that the single scattering algorithm shows considerable promise.

10.3.2 Application to MODIS/SeaWiFS

Since the key to application of the single scattering algorithm to the second-generation sensors, which *do* have spectral bands in the NIR, is the extrapolation from $\varepsilon(\lambda_s, \lambda_l)$ to $\varepsilon(\lambda_i, \lambda_l)$, it is important to try to understand the expected spectral behavior of $\varepsilon(\lambda_i, \lambda_l)$, i.e., the spectral behavior of ρ_{as} , to be able to make a valid extrapolation. We will try to do this in two ways: (1) by examining actual imagery for which $[\rho_w]_N$ is measured contemporaneously throughout the spectrum, so $\rho_a(\lambda)$, the multiple scattering analog to ρ_{as} , i.e., $\rho_t - \rho_r - t_w \rho_w$, can be determined, and (2) to examine the behavior of $\varepsilon(\lambda_i, \lambda_l)$ as

⁸The CZCS did have a spectral band centered at 750 nm having a spectral width of 100 nm. However, it was useless for atmospheric correction because of its low radiometric sensitivity compared to the visible bands.

⁹Since t_w is unknown at this point, we replace it with t^* (actually, t_r^*).

predicted from realistic aerosol models.

10.3.2.1 Spectral Behavior of ρ_a from SeaWiFS Data

We first look at the behavior of $\rho_a(\lambda)$ as revealed by data taken by the SeaWiFS sensor coincident with surface measurements of ρ_w . These are provided in Figure 10.4. They were obtained in the clear waters (low Chlorophyll) in the vicinity of the Hawaiian Islands. These data were taken on eight different days (eight separate satellite overpasses) with the ship-borne measurement coincident with the satellite measurement. In Figure 10.4 the measurements (ρ_t and ρ_w) are in the top panels with some derived quantities in the bottom panels. The measurement of ρ_w shows a spectrum that is typical of clear ocean water (Table 10.1), i.e., high values in the blue progressing rapidly to negligible values in the red and NIR. Clearly, the assumption that $\rho_w \approx 0$ in the red and NIR is valid for this region. The lower left panel is the residual reflectance after subtraction of the Rayleigh component of the total, i.e., $\rho_t - \rho_r$, with ρ_r computed using a full multiple scattering code, but using the standard atmospheric pressure rather than the actual. Subtracting the water-leaving reflectance (modified by the Rayleigh component of the diffuse transmittance t_r^*) then yields the aerosol component:

$$\rho_a = \rho_t - \rho_r - t_r^* \rho_w,$$

shown in the lower right panel. The diffuse transmittance t_r^* was used as an approximation, as both the aerosol properties and the angular distribution of ρ_w are unknown. We note that the aerosol component is weakly dependent on wavelength. As we shall see below, a weak dependence of the aerosol reflectance on wavelength, i.e., $\varepsilon(\lambda_i, \lambda_l) \approx 1$, is characteristic of a marine aerosol at high relative humidity. Note that, even in this area remote from large land masses and where the aerosol is mostly generated by breaking waves, the aerosol reflectance in the NIR, which is proportional to the aerosol concentration, still varies by a factor of four.

10.3.2.1 Spectral Behavior of ρ_a from Aerosol Models

We now try to understand what aerosol properties influence the behavior of $\varepsilon(\lambda_i, \lambda_l)$ by examining what is revealed about this quantity by several of the aerosol models described in detail in Chapter 4. We begin with the Shettle and Fenn [1979] models. Sample results for $\varepsilon(\lambda_i, \lambda_l)$, where λ_l is taken to be 865 nm (SeaWiFS), are presented in Figure 10.5 (left panel). Note that ε is a function of the Sun-viewing geometry, so these results only apply to the specific geometry indicated ($\theta_v^{(s)} \approx 1^\circ$ and $\theta_0 = 60^\circ$). These computations show a variation of ε with aerosol model. Most of the spectral variation is due to the spectral variation of the aerosol optical thickness, τ_a (Eq. (10.8)), with the rest due to the weak (but

significant) spectral variation of the phase function. Note that if $\varepsilon(\lambda_i, \lambda_l) \propto \exp[k(\lambda_l - \lambda_i)]$, where k is a constant, then $\varepsilon(\lambda_i, \lambda_l)$ would yield a straight line if plotted on Figure 10.5. This shows that over the spectral range considered (about a factor of 2 in wavelength) $\varepsilon(\lambda_i, \lambda_l)$ is nearly an exponential function of $\lambda_l - \lambda_i$, for the Shettle and Fenn [1979] models. Similar computations are presented for Haze C models (Chapter 4) in the right panel of Figure 10.5. Twelve separate Haze C models are shown: $\nu = 2, 3$, and 4, with the refractive index of the particles taken to be that of liquid water (labeled “H₂O”), close to that of the dust component of the Tropospheric model ($1.53 - 0.008i$, labeled “dust”), non-absorbing minerals ($1.50 - 0i$, labeled “1.50”), and absorbing minerals (labeled “min.”) as might be from windblown desert aerosols. Only for the absorbing minerals does the refractive index depend significantly on wavelength. We note that with the exception of the minerals, the exponential variation of ε with wavelength is clearly evident, and almost independent of the refractive index. An important observation from Figure 10.5 (right panel) is that, *in general, $\varepsilon(765, 865)$ depends mostly on the size distribution, and cannot be utilized to discriminate with certainty between weakly- and strongly-absorbing aerosols with similar size distributions.*

10.3.2.2 Test of SSA Algorithm with Synthetic Data

Let’s now examine the accuracy of the single-scattering algorithm for a sensor with two NIR bands, assuming an exponential spectral variation of $\varepsilon(\lambda_i, \lambda_l)$. We simulated atmospheres using an array of aerosol models: (1) the aerosol optical properties were taken from the Tropospheric, Coastal, and Maritime models at RH = 80%, denoted, respectively, as T80, C80, and M80; (2) the Shettle and Fenn [1979] Urban model at RH = 80% (U80) was used as an example of an aerosol with strong absorption and nearly wavelength-independent absorption index, e.g., a carbonaceous aerosol that might be characteristic of urban pollution (at 865 nm $\omega_a = 0.9934, 0.9884$, and 0.9528 , respectively, for the Maritime, Coastal, and Tropospheric models at 80% RH, but only 0.7481 for the Urban model with RH = 80%.); and (3) the twelve Haze C models described in the previous section.

Pseudo data values of $\rho_t(\lambda)$ were generated using a two-layer “exact” scalar radiative transfer code in which the aerosol was all in the lower layer and the molecular (Rayleigh) scattering all in the upper layer, a distribution of aerosols similar to that typically found over the open ocean, where most of the aerosol is confined to the marine boundary layer. The aerosol had the properties of each of the 16 individual models described above. The water surface was flat and all photons that penetrated the interface were assumed to be absorbed in the ocean, so $\rho_w(\lambda) = 0$ in the pseudo data.

There were seven Sun-viewing geometries: $\theta_0 = 20^\circ, 40^\circ$, and 60° , with $\theta_v^{(s)} \approx 1^\circ$ and

$\phi_v - \phi_0 = 90^\circ$, i.e., viewing near nadir, the sensor's scan center; and $\theta_0 = 0^\circ, 20^\circ, 40^\circ$, and 60° , with $\theta_v^{(s)} \approx 45^\circ$ and $\phi_v - \phi_0 = 90^\circ$. The set with $\theta_v^{(s)} \approx 45^\circ$ would represent viewing near the scan edge for most ocean color scanners. The simulations were performed for $\lambda = 443, 555, 765$, and 865 nm, which enables estimation of the accuracy of the algorithm in retrieving ρ_w in the blue and green, and the concomitant accuracy in the retrieved pigment concentration. Two aerosol optical thickness were used at 865 nm: $\tau_a(865) = 0.1$ and 0.2 , a typical value and high value for the open ocean, respectively. The $\tau_a(\lambda_i)$ at the other three wavelengths were determined from the spectral variation of the extinction coefficient for the individual model. Computations of these spectral variations are provided in Figure 10.6.

Because the actual $\rho_w(\lambda_i)$ was taken to be zero in the pseudo data the error in atmospheric correction, $\Delta(t\rho_w)$, is equal to the error in the predicted path radiance. This error is¹⁰

$$\begin{aligned}\Delta(t\rho_w(\lambda_i)) &= \rho_t(\lambda_i) - \rho_{path}(\lambda_i) = \rho_t(\lambda_i) - \rho_r(\lambda_i) - \varepsilon^{(e)}(\lambda_i, \lambda_l)\rho_{as}(\lambda_l) \\ &= \rho_t(\lambda_i) - \rho_r(\lambda_i) - \varepsilon^{(e)}(\lambda_i, \lambda_l)(\rho_t(\lambda_l) - \rho_r(\lambda_l)),\end{aligned}\tag{10.9}$$

where $\varepsilon^{(e)}(\lambda_i, \lambda_l)$ is the estimated value of $\varepsilon(\lambda_i, \lambda_l)$. If we assume an exponential variation of $\varepsilon(\lambda_i, \lambda_l)$ with λ_i , this estimate is

$$\varepsilon^{(e)}(\lambda_i, \lambda_l) \triangleq \exp[k(\lambda_l - \lambda_i)] = \exp\left[\left(\frac{\lambda_l - \lambda_i}{\lambda_l - \lambda_s}\right) \log_e \left(\frac{\rho_t(\lambda_s) - \rho_r(\lambda_s)}{\rho_t(\lambda_l) - \rho_r(\lambda_l)}\right)\right].$$

The Rayleigh reflectance $\rho_r(\lambda_i)$ was computed using the same radiative transfer code and geometry, but without aerosol. Figure 10.7 provides the error in the retrieved normalized water-leaving reflectance at 443 nm for each geometry and value of $\tau_a(865)$. Since $t\rho_w = tt_E[\rho_w]_N$, the error in the normalized water-leaving reflectance was derived from $\Delta(t\rho_w(443))$ assuming that $t(443) = t_r^*(443)$ and $t_E(443) = t_r^*(443)$ (see Footnote 20).

For non-absorbing aerosols (Figure 10.7, open symbols), the performance of this simple algorithm at 443 nm is excellent, with only a few cases falling outside the desired ± 0.002 error. In some cases, e.g., $\nu = 4$ for which Figure 10.6 shows that $\tau_a(443) \approx 0.35$ and 0.70 for $\tau_a(865) = 0.1$ and 0.2 , respectively, even with high values of $\tau_a(443)$ the algorithm works well in some geometries in the absence absorption. The large negative errors for $\nu = 4$ occur at the scan edge with $\theta_0 = 60^\circ$, i.e., the geometry with the most multiple scattering. When the aerosols are absorbing the errors are mostly negative and are significantly larger for $\tau_a(865) = 0.2$ than 0.1 . There are two sources of error for absorbing aerosols: (1)

¹⁰Henceforth, we will indicate the diffuse transmittance t_w by t (without the subscript), unless it would cause confusion.

the exponential extrapolation of $\varepsilon(765, 865)$ to $\varepsilon(443, 865)$ would lead to an overestimation of $\varepsilon(443, 865)$ (Figure 10.5), which overestimates $\rho_{as}(443)$, yielding a negative retrieval error; (2) even weak absorption as in the T80 model for which the extrapolation *does* work well (Figure 10.5, left panel), the absorption itself leads to significant error in the single-scattering approximation (i.e., the error is due to the omission of higher-order scattering).

The error in $[\rho_w(550)]_N$ in relation to the error in $[\rho_w(443)]_N$ is provided in Figure 10.7 (lower panels). The observed improvement in atmospheric correction at 550 compared to 443 nm results from (1) the ε determination requires a smaller extrapolation at 550 nm, and (2) there is less multiple scattering at 550 nm as both τ_a (Figure 10.6) and τ_r are smaller. Clearly, the error at 550 nm is typically much less than that at 443 nm. The tendency is for $\Delta[\rho_w(550)]_N \sim (1/3)\Delta[\rho_w(443)]_N$ for case with non-absorbing aerosols and $\Delta[\rho_w(550)]_N \sim (1/4)\Delta[\rho_w(443)]_N$ when the aerosol is strongly absorbing. Occasionally $|\Delta[\rho_w(550)]_N| \gtrsim |\Delta[\rho_w(443)]_N|$, but this is rare.

At this point it is useful to recall the results of AERONET measurements over the oceans that were discussed in Chapter 4. Quoting from that chapter: “... recent data from AERONET island stations provide mean values for the aerosol optical thickness $\tau_a(500)$ and the Angstrom power p for several of the world’s oceans. The global average is $\tau_a(500) = 0.108$ and $p = 0.573$. The Southern Ocean has the clearest atmosphere ($\tau_a(500) = 0.060$ and $p = 0.380$) and lowest variability, while the Atlantic Ocean has the most turbid atmosphere ($\tau_a(500) = 0.190$ and $p = 0.604$) and the highest variability. An important conclusion to be drawn from these studies is that for most of the World oceans, $\tau_a(500) \leq 0.2$.” Note that for a given value of p , if the Haze C distribution were appropriate, $\nu = p + 2$. So, for aerosols such as described in this quote, $\varepsilon(765, 865) < 1.1$, and Figure 10.7 (top left panel, open symbols) shows that the single scattering CZCS-type algorithm should be capable of retrieving $[\rho_w(443)]_N$ with the desired accuracy ($\Delta[\rho_w(443)]_N < \pm 0.002$).

Although the single scattering approach is seen to work well for sufficiently small optical depth (Figure 10.7) and for non-absorbing aerosols — typically the case over the open ocean — we desire an algorithm that has broader applicability, e.g., near the coast, where the aerosol optical depth is higher and the weakly-absorbing T80 or the extreme U80 aerosols would more likely be present. Hence, we are led to consider the possibility of a full multiple scattering approach.

10.4 Multiple Scattering

What are the effects of multiple scattering that are relevant to our problem? Noting that the single-scattering algorithm provides an excellent correction in many cases, we need to

ask, how does multiple scattering affect the algorithm? Recall that in multiple scattering (neglecting whitecaps and Sun glitter)

$$\rho_t - \rho_r - t\rho_w = \rho_a + \rho_{ra},$$

while in single scattering

$$\rho_t - \rho_r - t\rho_w = \rho_{as}.$$

Clearly the effect of multiple scattering on the algorithm is revealed through the relationship between $\rho_a + \rho_{ra}$ and ρ_{as} . Figure 10.8 shows this relationship for two cases described in Chapter 9: the Tropospheric model with RH = 50% (T50) and the Maritime model with RH = 99% (M99). These models were chosen because of the important difference in their size distributions leading to significant differences in their scattering phase functions (Figure 10.9). Of the aerosol models considered in this work, T50 has the least forward scattering, while M99 is one of the most forward scattering.¹¹ Interestingly, the M99 model shows a large deviation from single scattering with the value of $\rho_a + \rho_{ra}$ being about 40% greater than ρ_{as} ; however, the deviation is nearly independent of wavelength. This increase is due to the strong forward scattering. A photon that interacts once has a certain probability of backscattering out of the atmosphere and this leads to ρ_{as} , but photons that forward scatter through small angles have a similar probability to be backscattered. Thus, for M99, allowing more than one scattering per photon significantly increases the probability of backscattering. In contrast, the T50 model shows a smaller, but wavelength dependent deviation from single scattering. In this case the forward scattering is smaller while the side scatter is larger, so there are fewer forward scattered photons to be backscattered at the second scattering. In addition, in contrast to M99, the T50 model has some absorption (albeit weak). Figure 10.10 shows computations similar to those in Figure 10.8 obtained using the T50 model phase function (for 865 nm, Figure 10.9), but with ω_a values of 0.6, 0.8, and 1.0, rather than the values given by the model. They show that decreasing ω_a decreases $\rho_a + \rho_{ra}$ for a given ρ_{as} , and indicate that strong aerosol absorption can have a significant impact on the relationship between $\rho_a + \rho_{ra}$ and ρ_{as} .¹²

Unlike the single-scattering algorithm, for which aerosol models are not utilized at all, the dependence of multiple scattering on the aerosol properties, suggest that models will be required in order to include their effects in the ρ_w retrieval algorithm.

¹¹Chapter 4 provides examples of how the phase functions for models such as these vary with wavelength.

¹²Note that $\rho_{as} \propto \omega_a \tau_a$, which is the *scattering* optical depth. Thus, a given ρ_{as} implies a given value of the scattering optical thickness of the layer.

10.4.1 The Multiple-Scattering Retrieval Algorithm

Clearly, a way must be found to deal with multiple scattering. However, the success of the single-scattering algorithm at low values of τ_a suggests that we should mimic it as closely as possible. That is the path we will follow. In this development, we use the shorthand notation $\rho_A \triangleq \rho_a + \rho_{ra}$, i.e., ρ_A is the multiple scattering counterpart to ρ_{as} . The only link between the aerosol and its physical properties that can be determined from $\rho_t(\lambda)$ are the values of $\rho_A(\lambda_s)$ and $\rho_A(\lambda_l)$, both of which contain the effects of multiple scattering. It is clear that, for a given wavelength, Sun-viewing geometry and model (say the j^{th}), we can write ρ_A as a function of ρ_{as} . An example is

$$\rho_A^{(j)} = a^{(j)}[\rho_{as}^{(j)}]^1 + b^{(j)}[\rho_{as}^{(j)}]^2 + c^{(j)}[\rho_{as}^{(j)}]^3 + \dots, \quad (10.10)$$

where $a^{(j)}$, $b^{(j)}$ and $c^{(j)}$ are constants that depend on the model and the Sun-viewing geometry. In the single-scattering approximation, $a^{(j)} = 1$ and $b^{(j)} = c^{(j)} \dots = 0$. Let us use the following notation for such a relationship:

$$\rho_A^{(j)}(\lambda) = \rho_A^{(j)} \{ \rho_{as}^{(j)}(\lambda) \}, \quad (10.11)$$

where the big curly brackets mean $\rho_A(\lambda)$ for the j^{th} model given as a function of $\rho_{as}^{(j)}$ for the j^{th} model. This notation specifies ρ_A as a function of ρ_{as} for the j^{th} model. It can be inverted to yield ρ_{as} as a function of ρ_A for the j^{th} model, i.e.,

$$\rho_{as}^{(j)}(\lambda) = \rho_{as}^{(j)} \{ \rho_A^{(j)}(\lambda) \}. \quad (10.12)$$

Figure 10.8 is an example of these relationships for the T50 and M99 models in the given Sun-viewing geometry. Now, for a given pixel we have the measured values of ρ_A at λ_s and λ_l which we denote as $\rho_A^M(\lambda_s)$ and $\rho_A^M(\lambda_l)$. Then for the j^{th} model we can find

$$\rho_{as}^{(j)}(\lambda_s) = \rho_{as}^{(j)} \{ \rho_A^M(\lambda_s) \} \quad \text{and} \quad \rho_{as}^{(j)}(\lambda_l) = \rho_{as}^{(j)} \{ \rho_A^M(\lambda_l) \},$$

i.e., the values of ρ_{as} at λ_s and at λ_l that would be valid if the actual aerosol were identical to that for the j^{th} model. For the j^{th} model then, we can find

$$\varepsilon^{(j)}(\lambda_s, \lambda_l) = \frac{\rho_{as}^{(j)} \{ \rho_A^M(\lambda_s) \}}{\rho_{as}^{(j)} \{ \rho_A^M(\lambda_l) \}},$$

the value of $\varepsilon(\lambda_s, \lambda_l)$ if the j^{th} aerosol model were correct. Now Figure 10.8 shows that that there is almost a linear relationship between ρ_A and ρ_{as} , i.e., the first term in Eq. (10.10) is dominant, and that the multiple scattering effects are nearly the same for both the NIR bands. This suggests that the retrieved value of $\varepsilon^{(j)}(\lambda_s, \lambda_l)$ is going to be close to the true value, independent of the aerosol model. That is, if we assume several (N)

different models (“candidates”) for the aerosol, but one model say the i^{th} is correct, then each model, e.g., the j^{th} , will yield an $\varepsilon^{(j)}(\lambda_s, \lambda_l)$ that will be close in value to the true $\varepsilon^{(i)}(\lambda_s, \lambda_l)$. Since the ε ’s resulting from each of the trial aerosol models are close to one another, rather than pick a single model to estimate this quantity, as it is likely none of the aerosol models are correct, it seems more reasonable to estimate $\varepsilon(\lambda_s, \lambda_l)$ though

$$\varepsilon(\lambda_s, \lambda_l) = \frac{1}{N} \sum_{j=1}^N \varepsilon^{(j)}(\lambda_s, \lambda_l),$$

where $\varepsilon^{(j)}(\lambda_s, \lambda_l)$ is the value of $\varepsilon(\lambda_s, \lambda_l)$ derived by assuming that the j^{th} aerosol model is correct. Such an estimate will be most closely bracketed by two of the models, i.e., one with a slightly smaller value and one with a slightly larger value. Call these $\varepsilon^{(-)}(\lambda_s, \lambda_l)$ and $\varepsilon^{(+)}(\lambda_s, \lambda_l)$, respectively. Then

$$\rho_{as}^{(-)}(\lambda_i) = \varepsilon^{(-)}(\lambda_i, \lambda_l) \rho_{as}^{(-)}(\lambda_l) \quad \text{and} \quad \rho_{as}^{(+)}(\lambda_i) = \varepsilon^{(+)}(\lambda_i, \lambda_l) \rho_{as}^{(+)}(\lambda_l),$$

so

$$\rho_A^{(-)}(\lambda_i) = \rho_A^{(-)} \{ \rho_{as}^{(-)}(\lambda_i) \} \quad \text{and} \quad \rho_A^{(+)}(\lambda_i) = \rho_A^{(+)} \{ \rho_{as}^{(+)}(\lambda_i) \}.$$

Thus, we have two values for $\rho_A(\lambda_i)$. The procedure usually adopted going forward from here is to assume that $\rho_A(\lambda_i)$ falls between $\rho_A^{(-)}(\lambda_i)$ and $\rho_A^{(+)}(\lambda_i)$ in the *same proportion* as $\varepsilon(\lambda_s, \lambda_l)$ falls between $\varepsilon^{(-)}(\lambda_s, \lambda_l)$ and $\varepsilon^{(+)}(\lambda_s, \lambda_l)$. Some kind of assumption such as this is required, but as we shall see, this one is not always true. The N aerosol models are called “candidate” models, and to the extent that the candidate models are similar to the actual aerosol, this particular assumption works quite well.

10.4.2 Simulated Test of the Multiple-Scattering Algorithm

For our tests of the algorithm with simulated data (and for the initial application to SeaWiFS and MODIS) twelve candidate aerosol models were used: the Maritime, Coastal, and Tropospheric models with RH = 50, 70, 90, and 99%, labeled T50 to M99. Functions similar to Eq. (10.10) and the inverse functions were constructed by solving the radiative transfer equation for each model for $\theta_0 = 0$ to 80° in increments of 2.5° at 33 values of $\theta_v^{(s)}$ for eight values of $\tau_a(\lambda_i)$ from 0.05 to 0.8. Fourier analysis was used for the azimuthal dependence as described in Chapter 2. The structure of the atmosphere was the same as the simulated test data described earlier: a two-layer atmosphere with aerosols in the lower layer and molecular scattering in the upper lower.

For the purpose of the test, pseudo data were created as in Section 10.3.2.3 using the [Shettle and Fenn \[1979\]](#) Tropospheric, Coastal, Maritime, and Urban models at RH = 80%,

denoted by T80, C80, M80, and U80, respectively, and the same two-layer atmosphere model. It is important to note that, although the size and refractive index distributions of T80, C80, and M80 are similar to members of the candidate aerosol set, they are *not identical* to any members of the candidate set. In contrast, the U80 model is completely different from any of the candidates, as it exhibits strong absorption.

Figure 10.11 provides the comparison between the single-scattering and multiple-scattering algorithms for the T80, C80, M80, and U80 pseudo data at the seven Sun-viewing geometries described earlier with $\tau_a(865) = 0.2$. The multiple-scattering algorithm improves the retrieval of $[\rho_w(443)]_N$ for the C80, and M80 cases, and significantly for the T80 cases; however, the U80 retrievals, although somewhat improved over single scattering, are still very poor. Even though the size distribution of the U80 model is similar to the candidate's, the fact that its absorption properties are significantly different from the candidate's, causes almost as large an error in the retrieval of $[\rho_w(443)]_N$ as neglecting multiple scattering. As in Figure 10.7 (bottom panels), Figure 10.11 (bottom panels) shows that the relationship between $\Delta[\rho_w(550)]_N$ and $\Delta[\rho_w(443)]_N$ is similar to that in the SSA.

The error in the pigment concentration induced by $\Delta[\rho_w(550)]_N$ and $\Delta[\rho_w(443)]_N$ in the multiple-scattering algorithm is provided in Table 10.2. For Table 10.2 the errors $\Delta[\rho_w(550)]_N$ and $\Delta[\rho_w(443)]_N$ for the T80, C80, and M80 cases were added to values of $[\rho_w(550)]_N$ and $[\rho_w(443)]_N$ that are characteristic of three pigment concentrations (0.10, 0.47, and 0.91 mg/m³) yielding retrieved reflectances that included the atmospheric correction error. These were used in Eq. (10.1) to derive C_P for each of the seven Sun-viewing geometries. For each pigment concentration, the retrieved values of C_P were averaged over all seven geometries and the three aerosol models, and the standard deviation was computed. This was carried out for $\tau_a(865) = 0.1, 0.2$, and 0.3 . As one would expect, the quality of the retrievals is better for the lower values of $\tau_a(865)$. The magnitude of the errors show that the algorithm performs very well indeed,¹³ and suggests that the multiple-scattering algorithm can yield excellent retrievals of C_P ; however, the candidate aerosol models must be similar in size and composition to the aerosol actually present.

10.4.3 Estimation of Aerosol Optical Depth τ_a

There is considerable interest now in studying the global distribution of aerosols because of their role in climate forcing and biogeochemical cycling. Earth-orbiting satellites are ideal for such studies. The aerosol concentration is $\propto \tau_a$. In this section we show that τ_a can be

¹³For $\tau_a(865) = 0.3$ and a true value of C_P of 0.91 mg/m³, one retrieved value of C_P was ≈ 9 mg/m³ ($\theta_0 = 60^\circ$, $\theta_v \approx 45^\circ$, T80, for which $\tau_a(443) \approx 0.75$ and $\tau_a(550) \approx 0.6$). This value was not included in the average or the standard deviation computation.

Table 10.2: Mean value of C_P obtained for seven viewing geometries and three aerosol models (M80, C80, and T80). The number in parenthesis is the standard deviation divided by the mean (in %). Note that these errors are over and above any intrinsic error in Eq. (10.1). (From Gordon [1997].)

$\tau_a(865)$	$C_{\text{True}} = 0.10$ (mg/m ³)	$C_{\text{True}} = 0.47$ (mg/m ³)	$C_{\text{True}} = 0.91$ (mg/m ³)
0.1	0.101 (1.6)	0.466 (3.4)	0.912 (9.1)
0.2	0.100 (3.1)	0.470 (4.7)	0.940 (12.8)
0.3	0.098 (5.5)	0.493 (15.3)	0.936 (25.3)

retrieved with a simple extension of the atmospheric correction algorithm.

Even in the single scattering approximation, one notes from Eq. (12.29) that it is not possible to estimate τ_a without assuming a model for the aerosol to provide ω_a and P_a . The assumption of an incorrect model can produce significant errors (up to a factor of 2–3) in the recovered τ_a . As in atmospheric correction, we will try to avoid using an incorrect model in the retrieval of τ_a by utilizing the only other aerosol information available on a pixel-by-pixel basis — the spectral variation of ρ_{as} in the NIR. The τ_a -retrieval algorithm is a straightforward extension of the atmospheric correction algorithm. The correction algorithm provides two candidate models based on $\varepsilon(765, 865)$ and these specify two sets of P_a and ω_a values for two estimates of τ_a . The estimated value of τ_a is then determined from the weighted average of the two estimates as in the correction algorithm. This scheme was tested using the T80, C80, and M80 simulations from the previous section. Tables 10.3 and 10.4 provide the % error in the retrieved $\tau_a(865)$ for three aerosol models at the center and the edge of a typical scan as a function of θ_0 . The true value of $\tau_a(865)$ was 0.2 or 0.4. All the calculations were carried out for $\phi_v - \phi_0 = 90^\circ$. From the tables, we can see that the error in the retrieved aerosol optical thickness is typically within $\pm 10\%$ (and usually considerably less) for most of the cases examined. As with the atmospheric correction, the candidate aerosol models have to be realistic. However, in the case of τ_a , the model must produce a realistic phase function as well as single-scattering albedo. For example, if the size distribution and refractive indices for the model were precise, but the particles were not spherical as assumed, then the Mie-computed phase function could be significantly different from the actual one (Figure 4.30), yielding a correspondingly significant error in τ_a . Operation of the algorithm for the purpose of atmospheric correction does *not* require such accuracy in the phase function.

Table 10.3: Error in retrieved $\tau_a(865)$ for viewing at the center and edge of the scan. The true value of $\tau_a(865)$ is 0.20.

Position	θ_0	Error (%) in $\tau_a(865)$		
		M80	C80	T80
Center	20°	+17.4	+0.09	+0.63
	40°	−1.53	−2.88	−0.41
	60°	+2.96	−10.5	−2.41
Edge	0°	+0.55	−3.64	−0.88
	20°	+1.31	−4.74	−1.28
	40°	+2.41	−9.27	−2.54
	60°	+3.71	−14.0	−4.18

Table 10.4: Error in retrieved $\tau_a(865)$ for viewing at the center and edge of the scan. The true value of $\tau_a(865)$ is 0.40.

Position	θ_0	Error (%) in $\tau_a(865)$		
		M80	C80	T80
Center	20°	+16.9	+0.32	+0.19
	40°	−1.03	−4.57	+0.72
	60°	+3.78	−8.18	+2.05
Edge	0°	+1.12	−4.13	+1.04
	20°	+1.87	−4.94	+1.18
	40°	+3.41	−7.58	+1.69
	60°	+6.49	−7.80	+2.77

10.4.4 The Diffuse Transmittance

There remains to derive the diffuse transmittance including the multiple scattering. As developed in Chapter 9, the diffuse transmittance is the transmittance relating water-leaving radiance from the surface (in a given direction) to the water-leaving radiance at sensor (in the *same* direction). It can be computed simply by introducing radiance leaving the sea surface (with a given angular distribution) as a boundary condition at the bottom of the atmosphere, and solving the radiative transfer equation for the radiance exiting the top of the atmosphere. However, there is another way to carry out the computation that involves using the reciprocity principle.

Recall from Chapter 2, where we considered a volume V bounded by a surface S , and two separate radiative transfer problems involving the given medium within V . In problem 1 the radiance incident on S from outside V was given by $L_1(\vec{\rho}, \hat{\xi}, \lambda)$ and the internal sources (sources within V) were specified by $Q_1(\vec{r}, \hat{\xi}, \lambda)$. Likewise, in problem 2 the radiance

incident on S from outside V was $L_2(\vec{\rho}, \hat{\xi}, \lambda)$ and the internal sources were specified by $Q_2(\vec{r}, \hat{\xi}, \lambda)$. Then, under the assumption that the volume scattering function is time reversal invariant, the *reciprocity principle* relates the solutions $L_1(\vec{r}, \hat{\xi}, \lambda)$ and $L_2(\vec{r}, \hat{\xi}, \lambda)$ of these two problems to one another:

$$\begin{aligned} & \int_S dS \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| \left[\frac{L_1(\vec{\rho}, \hat{\xi}, \lambda) L_2(\vec{\rho}, -\hat{\xi}, \lambda)}{m^2(\vec{\rho})} - \frac{L_1(\vec{\rho}, -\hat{\xi}, \lambda) L_2(\vec{\rho}, \hat{\xi}, \lambda)}{m^2(\vec{\rho})} \right] d\Omega(\hat{\xi}) \\ &= \int_{4\pi} d\Omega(\hat{\xi}) \int_V \left[\frac{Q_2(\vec{r}, -\hat{\xi}, \lambda) L_1(\vec{r}, \hat{\xi}, \lambda)}{m^2(\vec{r})} - \frac{Q_1(\vec{r}, \hat{\xi}, \lambda) L_2(\vec{r}, -\hat{\xi}, \lambda)}{m^2(\vec{r})} \right] dV, \end{aligned} \quad (10.13)$$

where S is the bounding surface of V and has a normal (outward) \hat{n} , and m is the refractive index. In this equation, $L_1(\vec{\rho}, -\hat{\xi}, \lambda)$ and $L_2(\vec{\rho}, -\hat{\xi}, \lambda)$ are the radiances *exiting* the volume at the boundaries, and the $\hat{\xi} \bullet \hat{n} < 0$ on the solid angle integral on the left-hand-side indicates that the integration is restricted to directions *into* the medium. Points within V are located by \vec{r} , while points on the boundary of V are located by $\vec{\rho}$.¹⁴ So, if there are no sources within V , this reduces to

$$\begin{aligned} & \int_S dS \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| \left[\frac{L_1(\vec{\rho}, \hat{\xi}, \lambda) L_2(\vec{\rho}, -\hat{\xi}, \lambda)}{m^2(\vec{\rho})} \right] d\Omega(\hat{\xi}) \\ &= \int_S dS \int_{\hat{\xi} \bullet \hat{n} < 0} |\hat{\xi} \bullet \hat{n}| \left[\frac{L_1(\vec{\rho}, -\hat{\xi}, \lambda) L_2(\vec{\rho}, \hat{\xi}, \lambda)}{m^2(\vec{\rho})} \right] d\Omega(\hat{\xi}). \end{aligned} \quad (10.14)$$

It is straightforward to use this to develop a rigorous relationship for the diffuse transmittance of the atmosphere, valid for all orders of multiple scattering. Let the volume V comprise the whole atmosphere with upper surface at the top of the atmosphere, and lower surface *just beneath* the sea surface. Then, for problem 1 choose the incident radiance on the top of the atmosphere (TOA) to be that of the solar beam, i.e., $L_1(\vec{\rho}_T, \hat{\xi}) = F_0 \delta(\hat{\xi} - \hat{\xi}_0)$, where $\vec{\rho}_T$ is a TOA point and $\hat{\xi}_0$ is the direction of propagation of the solar beam. We assume that there is no upward radiance incident at $\vec{\rho}_B$, a point just beneath the sea surface, i.e., $L_1(\vec{\rho}_B, \hat{\xi}) = 0$ for $\hat{\xi} \bullet \hat{n} < 0$. For problem 2 we let $L_2(\vec{\rho}_T, \hat{\xi}) = 0$ for $\hat{\xi} \bullet \hat{n} < 0$ (no incident radiance on the TOA), and $L_2(\vec{\rho}_B, \hat{\xi})$ be specified for $\hat{\xi} \bullet \hat{n} < 0$, i.e., a specified upward radiance distribution incident on the bottom surface. Then applying Eq. (10.14), we have

$$L_2(\vec{\rho}_T, -\hat{\xi}_0) = \frac{1}{F_0 |\hat{\xi}_0 \bullet \hat{n}|} \int_{\Omega_d} |\hat{\xi} \bullet \hat{n}| \frac{L_1(\vec{\rho}_B, \hat{\xi}) L_2(\vec{\rho}_B, -\hat{\xi})}{m_w^2} d\Omega(\hat{\xi}),$$

where Ω_d is the full solid angle in the downward direction, and m_w is the refractive index of water. Letting $\hat{\xi}'_0$ be the direction of the refracted solar beam in the water for a flat

¹⁴Note that S cannot fall on a surface where the refractive index changes discontinuously, e.g., the air-water interface, because m is undefined there.

surface,

$$\frac{L_2(\vec{\rho}_T, -\hat{\xi}_0)}{L_2(\vec{\rho}_B, -\hat{\xi}'_0)} = \frac{1}{F_0|\hat{\xi}_0 \bullet \hat{n}|m_w^2} \int_{\Omega_d} |\hat{\xi} \bullet \hat{n}| L_1(\vec{\rho}_B, \hat{\xi}) \frac{L_2(\vec{\rho}_B, -\hat{\xi})}{L_2(\vec{\rho}_B, -\hat{\xi}'_0)} d\Omega(\hat{\xi}). \quad (10.15)$$

If we replace $L_2(\vec{\rho}_B, -\hat{\xi}'_0)$ in Eq. (10.15) by $L_w(\vec{\rho}_B, -\hat{\xi}_0)$ (Chapter 2, Eq. (2.14)) the water-leaving radiance just *above* the sea surface, i.e.,

$$L_2(\vec{\rho}_B, -\hat{\xi}'_0) = \frac{m_w^2}{t_f(\hat{\xi}_0)} L_w(\vec{\rho}_B, -\hat{\xi}_0), \quad (10.16)$$

then, Eq. (10.15) becomes

$$\frac{L_2(\vec{\rho}_T, -\hat{\xi}_0)}{L_w(\vec{\rho}_B, -\hat{\xi}_0)} = \frac{1}{F_0|\hat{\xi}_0 \bullet \hat{n}|t_f(\hat{\xi}_0)} \int_{\Omega_d} |\hat{\xi} \bullet \hat{n}| L_1(\vec{\rho}_B, \hat{\xi}) \frac{L_2(\vec{\rho}_B, -\hat{\xi})}{L_2(\vec{\rho}_B, -\hat{\xi}'_0)} d\Omega(\hat{\xi}) \triangleq t(-\hat{\xi}_0). \quad (10.17)$$

The quantity $t(-\hat{\xi}_0)$ is what we defined as the diffuse transmittance in Chapter 9, i.e., Eq. (9.11). Equation (10.17) provides an alternate method for computing the diffuse transmittance. In this method we illuminate the top of the atmosphere by the solar beam in the direction $\hat{\xi}_0$ and solve the radiative transfer equation for the downward radiance *just beneath* the water surface, when the water is assumed to be totally absorbing (so $L_1(\vec{\rho}_B, \hat{\xi}) = 0$ for $\hat{\xi} \bullet \hat{n} < 0$) to derive $L_1(\vec{\rho}_B, \hat{\xi})$ for $\hat{\xi} \bullet \hat{n} > 0$. Then, given the actual upward radiance distribution $L_2(\vec{\rho}_B, -\hat{\xi})$ in the water (for the same atmosphere and illumination) use Eq. (10.17) to compute $t(-\hat{\xi}_0)$.¹⁵

We have seen in Chapter 6 the upwelling radiance just beneath the water surface is expected to have only a weak variation with direction, so it is useful to consider the case where $L_2(\vec{\rho}_B, -\hat{\xi})$ is independent of $\hat{\xi}$, i.e., the upwelling radiance is uniform just *beneath* the water surface. Then

$$t(-\hat{\xi}_0) = \frac{E_{d1}(\vec{\rho}_B)}{F_0|\hat{\xi}_0 \bullet \hat{n}|t_f(\hat{\xi}_0)} \triangleq t^*(-\hat{\xi}_0), \quad (10.18)$$

where

$$E_{d1}(\vec{\rho}_B) \triangleq \int_{\Omega_d} |\hat{\xi} \bullet \hat{n}| L_1(\vec{\rho}_B, \hat{\xi}) d\Omega(\hat{\xi})$$

¹⁵We have derived Eq. (10.14) for the case of a flat interface. For a wind-ruffled interface the equation (10.15) is still valid and relates the TOA radiance to the upwelling *subsurface* radiance, but Eq. (10.16) can no longer be rigorously applied. However, we have seen in Chapter 7 (Figures 7.14 and 7.15) that Eq. (10.16) is still a very good approximation as long as the angle $\theta_0 \lesssim 60^\circ$. The surface roughness will of course affect $L_1(\vec{\rho}_B, \hat{\xi})$, so it must be recomputed for each surface.

This is the version of the diffuse transmittance that is used in processing SeaWiFs and MODIS data.¹⁶ Computation of t^* requires an aerosol model and τ_a . Fortunately, the

¹⁶In Chapter 9 we developed a single scattering version of the diffuse transmittance and applied it to a situation in which the water-leaving radiance itself was uniform, i.e., the upward radiance was uniform *after* transmission through the water surface from below (we also called this t^*). This situation can also be treated with the reciprocity principle. Let the volume V again be the whole atmosphere with one surface (B) just beneath the water surface and the other surface (T) at the TOA. Let \hat{n}_B and \hat{n}_T be the associated *outward* normals on the surfaces at $\vec{r} = \vec{\rho}_B$ and $\vec{r} = \vec{\rho}_T$, respectively. The water below the surface is assumed to be totally absorbing. We consider two problems. Problem 1: The top boundary is illuminated by the solar beam of irradiance F_0 propagating in the direction $\hat{\xi}_0$, with no radiance incident on the bottom boundary and no sources within the medium. Problem 2: There is a plane source with a particular angular distribution within the medium, but no radiance incident from the outside on the top and bottom boundaries. These give the following terms in Eq. (10.13):

$$\begin{aligned}
 L_1(\vec{\rho}_T) &= F_0 \delta(\hat{\xi} - \hat{\xi}_0) \\
 L_1(\vec{\rho}_B) &= 0 & (\hat{\xi} \cdot \hat{n}_B < 0) \\
 Q_1(\vec{r}, \hat{\xi}) &= 0 \\
 L_2(\vec{\rho}_T) &= 0 & (\hat{\xi} \cdot \hat{n}_T < 0) \\
 L_2(\vec{\rho}_B) &= 0 & (\hat{\xi} \cdot \hat{n}_B < 0) \\
 Q_2(\vec{r}, \hat{\xi}) &= 0 & (\hat{\xi} \cdot \hat{n}_T < 0) \\
 Q_2(\vec{r}, \hat{\xi}) &= (\hat{\xi} \cdot \hat{n}_T) \delta(\vec{r} - \vec{\rho}_2) L_2 & (\hat{\xi} \cdot \hat{n}_T > 0),
 \end{aligned}$$

where the source is located on the plane $\vec{r} = \vec{\rho}_2$, L_2 is a constant, and the reason for the particular angular distribution of Q_2 will be apparent later. Inserting these into Eq. 3 yields,

$$\begin{aligned}
 L_2(\vec{\rho}_T, -\hat{\xi}_0) &= \frac{L_2}{F_0 |\hat{\xi}_0 \cdot \hat{n}_T|} \int_{-\hat{\xi} \cdot \hat{n}_T > 0} -(\hat{\xi} \cdot \hat{n}_T) L_1(\vec{\rho}_2, \hat{\xi}) d\Omega(\hat{\xi}) \\
 &= \frac{L_2}{F_0 |\hat{\xi}_0 \cdot \hat{n}_T|} \int_{\hat{\xi} \cdot \hat{n}_B > 0} (\hat{\xi} \cdot \hat{n}_B) L_1(\vec{\rho}_2, \hat{\xi}) d\Omega(\hat{\xi}) \\
 &= \frac{L_2}{F_0 |\hat{\xi}_0 \cdot \hat{n}_T|} E_{d1}(\vec{\rho}_2) \implies \frac{L_2(\vec{\rho}_T, -\hat{\xi}_0)}{L_2} = \frac{E_{d1}(\vec{\rho}_2)}{F_0 |\hat{\xi}_0 \cdot \hat{n}_T|}.
 \end{aligned}$$

where $E_{d1}(\vec{\rho}_2)$ is the downwelling irradiance on the plane in Problem 1. Thus,

$$t^*(-\hat{\xi}_0) = \frac{L_2(\vec{\rho}_T, -\hat{\xi}_0)}{L_2} = \frac{E_{d1}(\vec{\rho}_2)}{F_0 |\hat{\xi}_0 \cdot \hat{n}_T|}. \quad (10.19)$$

It remains to interpret the constant L_2 . Recall the definition of the intensity density:

$$Q = \frac{dJ}{dV} = \frac{d^4 P}{dV d\Omega} = \frac{d^4 P}{dz dA d\Omega} \implies \frac{d^2 P}{dA} = \int Q dz d\Omega,$$

so integrating over the upper hemisphere provides the upwelling irradiance and integrating over the lower hemisphere, the downwelling irradiance. Inserting the values of Q_2 yields $E_u(\vec{\rho}_2) = \pi L_2$, which is the upwelling irradiance associated with a *uniform* radiance L_2 in the upward direction. The Q_2 is therefore the intensity density that provides a uniform upwelling radiance on the plane at $\vec{r} = \vec{\rho}_2$, which we place just *above* the water surface, and the left-hand-side of Eq. (10.19) is the diffuse transmittance when the water-leaving radiance is uniform just *above* the interface. This is the approach used to compute the

parts of the atmospheric correction algorithm described earlier can furnish both. Equation (10.18) provides a simple method for deriving t^* in a Monte Carlo solution to the radiative transfer equation: simply inject N photons in the direction $\hat{\xi}_0$ into the top of the atmosphere over a *totally absorbing* ocean and detect the number n that penetrate the surface, then $t^*(-\hat{\xi}_0) = n/(Nt_f(\theta_0))$.

When we consider absorbing aerosols (Section 10.5) and their influence on atmospheric correction (it's very large), we will find that the aerosol vertical structure plays a major role in atmospheric correction. Fortunately, t^* is almost independent of the aerosol vertical structure, even in the case of strongly absorbing aerosols. Figure 10.12 compares the computed t^* 's for a two layer atmosphere with U50 aerosols isolated in the lower layer, with an atmosphere having the aerosols uniformly mixed with air. It shows that to a high degree, the vertical structure has little influence on the value of t^* .

Finally, we note that one should use t rather than t^* for the diffuse transmittance, and if we use t^* (as is common), we need to understand the magnitude of the induced error. To estimate t we need $L_2(\vec{\rho}_B, -\hat{\xi})$ in Eq. (10.15). This is the upwelling radiance just beneath the water surface which we designated as $L_u(0, \mu, \phi; u_0, \phi_0)$ in Chapter 6. In this case, we only need the shape of L_u as it is normalized by its value at a particular angle in Eq. (10.15). We can use the quasi-single scattering model (Chapter 6) as a first approximation of the shape of $L_u(0, \mu, \phi; u_0, \phi_0)$. In this model the shape is completely determined by the scattering phase function at scattering angles greater than about 48° . Here we will use two scattering phase functions: the Rayleigh phase function to represent very clear ocean water and the Petzold (Phase-T) phase function to represent waters with a high particle concentration. Figure 6.38 in Chapter 6 provides an example of an angular distribution computed using the quasi-single scattering model, and also the deviations from the model incurred as multiple scattering becomes more dominate. Figures 10.13-10.16 provide the error incurred using t^* in place of t as a function of viewing angle in recovering ρ_w for a variety of aerosol optical thicknesses and solar zenith angles. In these figures the scan plane is perpendicular to the solar beam ($\phi_v - \phi_0 = 90^\circ$). For $\tau_a \approx 0.1$ (typical of the open-ocean), we see that the error in these extreme cases rarely exceeds 3%. We use the word “extreme” here because multiple scattering within the water will smooth out the variations in the QSSA-values of $L_u(0, \mu, \phi; u_0, \phi_0)$, so the errors in the figures can be considered upper limits for the given phase function. Also, for the Petzold phase function, which is more representative of ocean particles, the variation of $L_u(\hat{\xi})$ with $\hat{\xi}$ is smaller which places it closer to a uniform radiance distribution moving t closer to t^* .

“exact” results for t^* in Chapter 9, Figures 9.16 – 9.18. The reader should be able to verify that when the right-hand-side is computed in the single scattering approximation, the result is identical to Eq. (9.18) in Chapter 9 when L_w is uniform.

10.4.5 Ancillary Data Required to Operate the Algorithm

Several sets of ancillary data are required to operate the $[\rho_w]_N$ retrieval algorithm in the most effective manner. We briefly discuss each ancillary data set below. Some are now available as measured global data sets from other space-borne sensors, others as the output of numerical weather models.

10.4.5.1 Ozone Optical Thickness

In the radiative transfer model a three-layer atmosphere was assumed. On top was the Ozone layer, which was non-scattering, in the middle was a molecular-scattering (Rayleigh) layer and on the bottom was the aerosol. The Ozone optical thickness $\tau_{Oz}(\lambda)$ is required to compute the two-way transmittance of the various reflectances ρ_r , ρ_w , etc., through the Ozone layer. The Ozone absorption is small, but the concentration is highly variable over the planet (Figure 10.17). From Chapter 4, Figure 4.28, we see that $\tau_{Oz} \approx 0.044$ at the absorption maximum (near 600 nm) for an Ozone concentration (C_{Oz}) of 343 DU (Dobson Units), or $k_{Oz} = 1.28 \times 10^{-4} \text{ DU}^{-1}$. Since the two-way Ozone transmittance along a path specified is $T_{Oz} = \exp(-k_{Oz}C_{Oz}M)$, where M is the air mass ($u^{-1} + |u_0|^{-1}$), we find the relative error in T_{Oz} is

$$\frac{\Delta T_{Oz}}{T_{Oz}} = -k_{Oz}\Delta C_{Oz}M.$$

For not too large viewing or Sun angles $T_{Oz} \approx 1$, and $\Delta T_{Oz} \approx -2k_{Oz}\Delta C_{Oz}$. Thus, for $\Delta T_{Oz} \lesssim 1\%$, $\Delta C_{Oz} \lesssim 39 \text{ DU}$.¹⁷ The accuracy of a TOMS-like satellite-borne Ozone sensor exceeds this, so it, or a similar instrument, could be used to assess the effect of Ozone.

10.4.5.2 Surface Atmospheric Pressure P_0

The atmospheric pressure is needed to compute the Rayleigh optical thickness (τ_r) necessary for the computation of ρ_r , the diffuse transmittance t , and the direct transmittance T . It is readily available from global weather models. Given the value of τ_{r0} , the Rayleigh optical thickness at the standard atmospheric pressure P_0 of 1013.25 mb (Chapter 4, Eq. (4.8)) at any surface pressure P , the Rayleigh optical depth is

$$\tau_r = \frac{P}{P_0}\tau_{r0}.$$

¹⁷Not that this Ozone concentration error yields a transmission error of 1% only in the green-yellow part of the spectrum ($\sim 600 \text{ nm}$). The error in the blue and NIR would be significantly less (Figure 10.17).

Note that an error in P translates directly to an error in τ_r , i.e., $\Delta\tau_r/\tau_r = \Delta P/P$, and since most of ρ_r is the result of single scattering (Chapter 9, Figures 9.7 and 9.8), the error in τ_r translates directly into error in ρ_r : $\Delta\rho_r/\rho_r = \Delta P/P$. Because ρ_r makes up the bulk of ρ_t in the blue, it is necessary to calculate it with an error of no more than about 0.5% or less. This requires $\Delta P \lesssim 5$ mb.

10.4.5.3 Wind Speed W and Wind Vector \vec{W}

The wind speed can be used in the computation of ρ_r , i.e., if one desires to consider surface roughness in its computation. It is necessary for an estimation of $[\rho_{wc}]_N$ as described in Chapter 7.¹⁸ The wind vector \vec{W} is required for any assessment of Sun glint, e.g., construction of a glint mask, to remove areas contaminated by Sun glint from the imagery before processing.

10.4.5.4 Sea Surface Temperature and Atmospheric Stability

These quantities allow the computation of the Richardson number (Ri) used in newer estimates of the surface slope variance as part of an assessment of Sun glitter (see Chapter 7).

10.4.5.5 Relative Humidity RH

The surface relative humidity (RH) is not actually needed to operate the algorithm; however, it could be useful for constraining the candidate aerosol models (Section 10.4.1).

10.4.6 Annotated Flow Diagram of the Complete Algorithm

Many of the quantities required for the correction algorithm are stored in lookup tables (LUTs). These include the relationship between $\rho_a + \rho_{ra}$ and ρ_{as} for various optical thicknesses, sun-viewing geometry and models as well as those relating t^* to the same variables. Figure 10.18 provides a simplified flow diagram describing how the original SeaWiFS/MODIS algorithm was operated, along with the ancillary inputs and LUTs.

¹⁸The value of $[\rho_{wc}]_N$ is the remote sensing augmented reflectance $RSAR$ from Chapter 7. It is taken to be $3 \times 10^{-6} W^{2.55}$ throughout the visible for $W \geq 6$ m/s and zero for $W < 6$ m/s.

10.4.7 Non-zero $[\rho_w]_N$ in the NIR

A fundamental assumption in the atmospheric correction algorithm is that $[\rho_w]_N = 0$ in the NIR (748 and 869 nm for MODIS, 765 and 865 nm for SeaWiFS). However, it is well known that at high Chlorophyll *a* concentrations, even in Case 1 waters, there is a small, but non-zero, water-leaving reflectance in the NIR. Therefore, it is necessary to estimate $[\rho_w]_N$ in the NIR.

For Case 1 waters, we have an AOP model that does a good job relating L_u/E_d to the concentration of Chlorophyll *a* (Chapter 6) or to the pigment concentration. The relationship between L_u/E_d and $[\rho_w]_N$ is provided in the Appendix. The estimate of $[\rho_w]_N$ in the NIR can be incorporated in an atmospheric correction algorithm in the following manner. First, an atmospheric correction is carried out assuming that $[\rho_w]_N = 0$ in the NIR. The retrieved reflectances are then used in a ratio algorithm, e.g., similar to Figure 10.1, to estimate the concentration of Chlorophyll *a*. The estimate of C is then used to estimate $[\rho_w]_N$ in the NIR using the above relationships. This $[\rho_w]_N$ is then used to estimate $t\rho_w$ in the NIR, which is subtracted from ρ_t , and the atmospheric correction algorithm operated again using the revised ρ_t , etc.

For Case 2 waters, in which high concentrations of scattering material (both biogenic and non-biogenic) can lead to $[\rho_w]_N \neq 0$ in the NIR, a similar procedure can be employed. However, in this case specific relationships must be developed between the principal scattering constituents and the IOPs for each application area, as these relationships are typically unique to the individual study area. Alternatively, if the sensor has additional spectral bands in the short-wave-infrared (SWIR) such bands can be used in the algorithm described above (MODIS has such bands at 1.2, 1.6 and 2.1 μm). However, one must bear in mind that there is a pitfall with the SWIR approach: the aerosol optical properties in the SWIR are dominated by large particles and extrapolation of $\rho_a + \rho_{ra}$ from the SWIR to the visible may not be representative of its actual values in the visible, as the extrapolation can be a factor of 3 to 4 in wavelength compared to a factor of 2 with typical operation of the algorithm. Nevertheless, there has been considerable success using the SWIR for atmospheric correction of MODIS in turbid coastal waters, where $[\rho_w]_N$ can be a significant contributor to ρ_t in the NIR.

10.5 Absorbing Aerosols

We have seen in earlier examples that the multiple-scattering algorithm can fail when the aerosol is strongly absorbing, even though the particle size distribution may be similar to

that of the candidate aerosol models. Figure 10.11 shows that even when multiple scattering is included in the algorithm, the correction for atmospheres containing strongly-absorbing aerosols (e.g., U80) is not really improved over the single-scattering algorithm; however, there is considerable improvement for weakly-absorbing aerosols (e.g., T80). Because of the manner in which the twelve candidate models in Section 10.4.2 were developed (see Chapter 4) they force a definite relationship between ω_a and $\varepsilon(\lambda_i, \lambda_l)$. Physically this is due to a relation between the absorption properties (ω_a) and the size distribution (ε) that exists in these models. For the geometry shown in Figure 10.5, Figure 10.19 provides this relationship for the 12 candidates (M, C and T) along with the similar relationship for the urban models (U). The T80 model falls within the range of the candidates and a good correction is realized; however, the urban model U80, as well as all the other Urban models, are far outside the range of the candidates. No wonder the retrieval of ρ_w is so poor for U80. Practically, some kind of $\varepsilon - \omega_a$ relationship is necessary for atmospheric correction because we only have one piece of information available on which to base the choice of aerosol models: $\varepsilon(\lambda_i, \lambda_l)$. Unfortunately, as we noted earlier, $\varepsilon(\lambda_i, \lambda_l)$ depends mostly on the size distribution and much less on the refractive index (Figure 10.5 and 10.19) and, as such, it is a poor indicator of the presence or absence of aerosol absorption. Beyond this one piece of information we must rely on other knowledge, such a climatology, to make a judicious choice of a model set to use for correction. The MCT set seems to be adequate for the open ocean most of the time. However, if it is known for example that during a given time period (season) there is a high probability of urban-type aerosols resulting from pollution being present, e.g., at times off the U.S. East Coast in summer, then the MCT candidates would be a poor choice. Another observation from Figure 10.19: simply adding some strongly-absorbing aerosols into the candidate set will not work, as there is no way given $\varepsilon(\lambda_s, \lambda_l)$ to predict ω_a . It is clear that we almost need to know *a priori* whether the aerosol is strongly or weakly absorbing, or more precisely, its $\varepsilon - \omega_a$ relationship. In addition, there is a further complication: the effect of strongly-absorbing aerosols is dependent on the aerosol's vertical concentration profile, and strongly-absorbing aerosols such as wind-blown dust and carbonaceous urban aerosols are rarely contained wholly within the marine boundary layer, as is assumed in the multiple scattering algorithm.

In this section, first we provide an example showing that when candidate aerosol models are *restricted* to those with the appropriate absorption properties, the correction algorithm yields suitable results. Next we examine in some detail the influence of the vertical distribution of the aerosol. Finally, we discuss strategies for finding appropriate models and incorporating them into a correction algorithm.

10.5.1 Atmospheric Correction with a Restricted Set of Candidate Models

Before continuing, it is important to ask: if candidate aerosol models containing the appropriate absorption properties (along with the *correct* vertical distribution) are used in the correction procedure as developed thus far, will the uncertainty of the resulting $[\rho_w]_N$ retrieval be within acceptable limits? To examine this question, we constrained the algorithm to utilize only four candidate models, the Urban models U50, U70, U90, and U99 shown in Figure 10.19. Then the algorithm was tested using pseudo data simulated with the U80 model, which falls approximately where the “U” symbol appears on Figure 10.19. The results showing the error in $[\rho_w(443)]_N$ as a function of the aerosol optical thickness of U80 at 865 nm are provided in Figure 10.20. Note that even with $\tau_a(865)$ as high as 0.4, there are only two cases that fall outside the allowable range of error. These occur at the scan edge with $\theta_0 = 60^\circ$ — the most difficult geometry to correct. This shows that when models with the proper absorption properties and the correct vertical distribution are used as candidates, even in the case of strong aerosol absorption, acceptable retrievals are obtained. As mentioned in the previous section, one cannot just add appropriate candidates to the model mix, i.e., add some absorbing models to the non- or weakly-absorbing models, because the value of $\varepsilon(\lambda_s, \lambda_l)$ is only weakly dependent on aerosol absorption (Figure 10.5, right panel and Figure 10.19), so there would be no way to sort out the strongly from the weakly absorbing models using ρ_t in the NIR. There must be some model *restriction* when strongly-absorbing aerosols are suspected.¹⁹ But this is not the only problem.

10.5.2 Aerosol Vertical Structure

The reflectance of the atmosphere in the single-scattering approximation is independent of the vertical distribution of the aerosol. However, this is not the case when multiple scattering becomes important. To demonstrate the importance of the vertical distribution we provide a simple example. Recall that the algorithm, as described thus far, places all of the aerosol at the bottom of a two-layer radiative transfer model, with all of the Rayleigh scattering in the upper layer. We compare the correction for a case where the pseudo data were created for an atmosphere that actually *has* the assumed structure, with that for which the pseudo data were created for an atmosphere in which the aerosols are uniformly mixed with air throughout the atmosphere, e.g., a one-layer atmosphere. Figure 10.21 (left panel) provides such a comparison for the M80 and T80 aerosol models with $\tau_a(865) = 0.2$. In this case the correction algorithm was operated with the original twelve

¹⁹The situation is not totally hopeless. One could, for example, use a standard set of models to effect a correction, and if unrealistic results are obtained, e.g., some reflectance are *negative*, one could try a limited set of more absorptive candidates.

candidate aerosol models. Note that the an incorrect assumption regarding the vertical structure, even in the extreme of uniform mixing, does not lead to serious error — only the T80 model at the scan edge for $\theta_0 = 60^\circ$ is slightly outside the acceptable error range. We note in passing that the T80 aerosol does have some (weak) absorption ($\omega_a(443) \approx 0.975$). Figure 10.21 (right panel) provides a similar comparison when the U70 and U80 models are used to create the pseudo data set. For this the candidates were restricted to the four used for Figure 10.20: U50, U70, U90, and U99. For U80 the error for the uniformly mixed case becomes excessive: over an order of magnitude larger than the two-layer case. But wait, look at U70, which is actually *one of the candidate aerosol models*! When the vertical structure is correct for U70, of course the error is negligible, but when the U70 aerosol is uniformly mixed, i.e., when the incorrect vertical structure is assumed, the error for U70 becomes even larger than that for U80.

Let's look at a more realistic vertical distribution for aerosols. Consider an aerosol layer that thickens from being confined just near the surface to being mixed higher and higher in the atmosphere, keeping the total aerosol concentration (τ_a) constant. The simulations were performed again for the U80 model using a two-layer radiative transfer model to create pseudo data with aerosol *plus* Rayleigh scattering in the lower layer and *only* Rayleigh scattering in the upper layer. The fraction of the Rayleigh scattering optical thickness assigned to the lower layer was that for aerosol-layer thicknesses of 0, 1 km, 2 km, 4 km, 6 km, and ∞ . The aerosol concentration was kept constant with $\tau_a(865) = 0.2$. The multiple-scattering algorithm was then operated with this pseudo data using the same four candidate models as in Figure 10.21 (right panel). Figure 10.22 provides the results in the usual manner. Obviously progressively thickening of the aerosol layer leads to a progression of increasingly larger error in $[\rho_w(443)]_N$.

This influence of vertical structure on the algorithm when the aerosol is strongly absorbing can be understood qualitatively in a simple manner. The algorithm is based on the assumption that all the aerosol is in a thin layer near the surface. As this layer thickens (mixing molecular scattering with aerosol scattering in the layer) the Rayleigh scattering within the aerosol layer will cause an increase in the average path length of photons through the layer, compared to that for just aerosol alone. This causes an increase in absorption and a decrease in ρ_t , which when processed by the algorithm results in value of $[\rho_w(443)]_N$ that is too small, and may even be negative.²⁰

Unfortunately, there is still an additional complication. Figure 10.23 shows the spectral behavior of $\rho_a + \rho_{ra}$ as a strongly absorbing aerosol is mixed higher into the atmosphere for a *fixed* value of $\tau_a(865)$ of 0.2. Clearly, for a given τ_a , ρ_t will decrease as the thickness of the aerosol layer increases. This decrease is relatively more in the visible than in the

²⁰The appearance of negative $[\rho_w(\lambda)]_N$'s in the blue is indeed often one of the signatures of a strongly absorbing aerosol.

NIR, so as the layer thickens, the algorithm will predict values of $\rho_a + \rho_{ra}$ in the visible that are too large, yielding an over correction, $\Delta[\rho_w(443)]_N < 0$. It is critical to note again that for strongly-absorbing aerosols the behavior of $\rho_a + \rho_{ra}$ in the NIR provides little or no information regarding the vertical distribution of the aerosol or its spectral variation, and as such, the behavior of $\rho_a + \rho_{ra}$ in the NIR provides *no* indication of its behavior in the visible. The predictability of $\rho_a + \rho_{ra}$ in the visible from the NIR is the foundation of the atmospheric correction algorithm for non- or weakly-absorbing aerosols, suggesting a new approach is needed to deal with strongly-absorbing aerosols.

10.5.3 Strategy for Dealing With Strongly-Absorbing Aerosols

How then will we ever be able to perform atmospheric correction in the presence of strongly-absorbing aerosols? Actually, Figure 10.23 provides a valuable clue. It shows that in the case of urban aerosols, and whatever the vertical distribution, the absorption is manifest in a progressively decreasing $\rho_a + \rho_{ra}$ with decreasing wavelength (i.e., red to blue) compared to what would be expected for non- or weakly-absorbing aerosols. Thus, to detect aerosol absorption we need to look in the blue part of the spectrum. However, the blue region of the spectrum is the one mostly influenced by the constituent concentration in the water, e.g., the pigment concentration, etc. So, how do we separate the effects of water color variations from aerosol color variations in the total reflectance measured at the sensor? That is, in

$$\rho_t(\lambda) - \rho_r(\lambda) = \rho_a(\lambda) + \rho_{ra}(\lambda) + t(\lambda)\rho_w(\lambda),$$

how do we apportion spectral variations seen in $\rho_t(\lambda) - \rho_r(\lambda)$ between $\rho_a(\lambda) + \rho_{ra}(\lambda)$ and $t(\lambda)\rho_w(\lambda)$? Just as we have models describing the spectral variation of $\rho_a(\lambda) + \rho_{ra}(\lambda)$, we are going to need to have a model describing the spectral variation of $\rho_w(\lambda)$ in terms of the water's constituent concentrations. Such models are available and described in Chapter 6. The strategy employed for dealing with strongly-absorbing aerosols is to find a set of water constituent concentrations and an aerosol model and concentration (optical thickness) as well as vertical distribution that explains the given $\rho_t(\lambda) - \rho_r(\lambda)$. That is the subject of this section.

We will begin describing a straightforward method for effecting this called the *spectral matching algorithm* (SMA). The method is tested through simulations in which the urban aerosol is dominant as in Section 10.5.2 and 10.5.3. Then we will apply the algorithm to retrieve ρ_w in the presence of Saharan dust over the Eastern North Atlantic. To effect the dust application, we will need to develop appropriate dust models which involves finding models that reproduce the radiative effects of the dust. This is described in detail. The resulting dust models are then used to perform actual water-constituent retrievals.

10.5.3.1 The Spectral Matching Algorithm (SMA)

In the spectral matching algorithm (SMA) the properties of the ocean and the atmosphere are retrieved simultaneously. As in each of the algorithms discussed so far, it is assumed that the whitecap and Sun glint contributions to $\rho_t(\lambda)$ have been removed and that $\rho_w(\lambda_l) = 0$. No assumption is made regarding $\rho_w(\lambda_s)$. Then from the measured $\rho_t(\lambda)$ and the computed $\rho_r(\lambda)$ we can form $\rho_t(\lambda) - \rho_r(\lambda)$, which we take as our experimental input data, and

$$\rho_t(\lambda) - \rho_r(\lambda) = \rho_a(\lambda) + \rho_{ra}(\lambda) + t(\lambda)\rho_w(\lambda).$$

Since $\rho_w(\lambda_l) = 0$, the difference $\rho_t(\lambda_l) - \rho_r(\lambda_l)$ provides $\rho_a(\lambda_l) + \rho_{ra}(\lambda_l)$, from which, for a given aerosol model (e.g., the i^{th}), one can find the value of the aerosol optical depth, $\tau_a^{(i)}(\lambda_l)$, that reproduces $\rho_a(\lambda_l) + \rho_{ra}(\lambda_l)$. Then from $\tau_a^{(i)}(\lambda_l)$ and the aerosol model, $\rho_a^{(i)}(\lambda_j) + \rho_{ra}^{(i)}(\lambda_j)$ can be determined for all spectral bands j .²¹ This provides the quantity

$$t(\theta_v, \lambda_j)\rho_w^{(i)}(\lambda_j) = \rho_t(\lambda_j) - \rho_r(\lambda_j) - \rho_a^{(i)}(\lambda_j) - \rho_{ra}^{(i)}(\lambda_j),$$

i.e., the retrieved values of $t(\theta_v, \lambda_j)\rho_w^{(i)}(\lambda_j)$ *assuming that the i^{th} aerosol model is correct*. The idea is now to compare the value of $t(\theta_v, \lambda_j)\rho_w^{(i)}$ with that predicted by a model of $t(\theta_v, \lambda)\rho_w(\lambda)$. This model, which is briefly described in the Appendix, provides $[\rho_w(\lambda)]_N$ as a function of the pigment concentration C_P and a particle scattering parameter b^0 . It is used to compute $[\rho_w(\lambda, C_P, b^0)]_N$ for a discrete set of values of C_P and b^0 falling within the typical range of variation. These provide trial values of

$$t(\theta_v, \lambda)\rho_w(\theta_v, \lambda) = t(\theta_v, \lambda)t_E(\theta_0, \lambda)[\rho_w(\lambda, C_P, b^0)]_N.$$

But here we assume that $t(\theta_v, \lambda)$ can be replaced by $t^{*(i)}(\theta_v, \lambda)$ and noting that $t_E(\theta_0, \lambda) \approx t^{*(i)}(\theta_0, \lambda)$, we have the modeled²²

²¹Note, $t^{*(i)}(\theta, \lambda_j)$ can be determined as well.

²²At this point it is useful to examine the relationship between t_E and t^* . These are given by $t_E = E_d^+/F_0 \cos \theta_0$ and $t^* = E_d^-/t_f F_0 \cos \theta_0$, where E_d^+ is the irradiance incident on the water *just above* (+) the interface, E_d^- is the irradiance incident on the water *just below* (-) the interface and t_f is the Fresnel transmittance of the solar beam. As in Chapter 6, Section 6.4.4.3, if we let $g = E_d^{S+}/E_d^+$, where E_d^{S+} is the irradiance incident on the water surface from the sky, then it follows that

$$t^* = \left[\frac{gT^S + (1-g)t_f}{t_f} \right] t_E,$$

where T^S is the irradiance transmittance of sky light through the interface. Thus, the relationship between t_E and t^* is governed by g and T^S : g depends on the properties of the atmosphere and T^S on the angular distribution of the sky radiance and the water surface roughness. As an example of the difference between t_E and t^* we examine the Rayleigh component of each at 412 nm. (Recall that Rayleigh scattering is much more important than aerosols in both quantities and is largest at the shortest wavelengths of ocean color sensors. The table below provides the ratio t^*/t_E , at 412 nm, as a function of θ_0 derived from exact

$$t(\theta_v, \lambda) \rho_w(\theta_v, \lambda) = t^{*(i)}(\theta_v, \lambda) t^{*(i)}(\theta_0, \lambda) [\rho_w(\lambda, C_P, b^0)]_N.$$

Next, we examine the residual

$$\delta(i, C_P, b^0) \triangleq 100\% \times \sqrt{\frac{1}{n-1} \sum_{j=1}^n \left[\frac{t^{*(i)}(\theta_v, \lambda_j) t^{*(i)}(\theta_0, \lambda_j) [\rho_w(\lambda_j, C_P, b^0)]_N - t(\theta_v, \lambda_j) \rho_w^{(i)}(\lambda_j)}{t^{*(i)}(\theta_v, \lambda_j) t^{*(i)}(\theta_0, \lambda_j) [\rho_w(\lambda_j, C_P, b^0)]_N} \right]^2},$$

where n is the number of visible wavelengths. The value of δ is computed for each aerosol model and set of ocean parameters. One might suggest that the set of parameters i , C_P , and b^0 , that yield the smallest $\delta(i, C_P, b^0)$ should be chosen as the best, i.e., the solution to the problem; however, as it is unlikely that the “correct” model is one of the set of candidates, averaging over some number of the best retrievals (i.e., retrievals with the lowest values of $\delta(i, C_P, b^0)$) to obtain the retrieved ocean and aerosol parameters seems more reasonable. Here that number is ten.

Figure 10.24 shows the result of tests of this algorithm using atmospheric pseudo data generated with the M80, C80, T80, and U80 aerosol models. The aerosol optical thickness at 865 nm is 0.2 for all cases. The pseudo data for $\rho_w(\lambda)$ were generated using the model for the water-leaving reflectance given in the Appendix. The same water model is used to generate the pseudo data and to operate the SMA; so this may represent an optimum situation. Sixteen aerosol models are used as candidates: M, C, T and U models with RH = 50, 70, 90, and 99%. Note the test models are not part of the candidate set. For this test, the vertical structure of the aerosol in the generated pseudo data is the same as that in the candidate models — again, the optimal situation. The figure clearly demonstrates that with the above limitations — correct water model and correct aerosol vertical distribution — the SMA performs very well. Of particular importance is that the *algorithm has no difficulty determining the presence of strongly absorbing aerosols*. In a real-world operation, the algorithm could incorporate vertical structure by having candidate models with a range of prescribed vertical structure.

computations for a flat interface.

θ_0	t^*/t_E
0°	0.988
20°	0.987
40°	0.984
60°	0.986
70°	1.018

Clearly, the error in replacing t_E by t^* is $\lesssim 1.5\%$.

10.5.3.2 Application of the SMA — A Case Study

Large dust plumes often flow off the African Continent over the Northeast Atlantic. Figure 10.25 shows a dramatic example of such a plume captured in a SeaWiFS image. The islands in the image provide evidence that this plume is not confined to the surface, but extends to at least 2 km. There is no way one could hope to effect retrieval of ρ_w in the dense parts of a plume such as this, as the water surface is totally obscured; however, even outside the densest parts, where the water surface is clearly visible, i.e., the blue portions of the image, the dust can confound the retrieval algorithm discussed in Section 10.4 because it strongly absorbs in the blue. Dust-contaminated imagery such as this can provide a challenging, real-world, test of the SMA. However, to effect such a test, we need a realistic model of the spectral scattering and absorption properties of the dust. In addition, as Figure 10.25 shows, we cannot assume the aerosol is confined to a thin layer near the surface as we did in the tests in Section 10.5.4.1.

In what follows, we first develop models of the dust that are based on SeaWiFS observations of a dust event (not the one in Figure 10.25), and then use these models in the SMA to perform atmospheric correction of imagery in the presence of wind-blown dust.

10.5.3.2.1 Development of Aerosol Models for Windblown Dust

To develop and tune models that are appropriate to the Saharan dust zone, we need to examine a dust event in which the dust concentration is strong enough that the contribution to $\rho_a + \rho_{ra} + t\rho_w$ from $t\rho_w$ is insignificant compared to $\rho_a + \rho_{ra}$. Such an event is available and is shown in Figure 10.26, which provides $\rho_A(865) \triangleq \rho_a(865) + \rho_{ra}(865)$ from SeaWiFS imagery off the coast of Africa. Figure 10.27 provides values of $\rho_A(\lambda)$ extracted from the image along the dashed track line in Figure 10.26. Interestingly, there is a steady *decrease* in $\rho_A(\lambda)$ progressing from the red to the blue. Note that the areas in the image with the highest dust concentration have $\rho_A(865) \approx 0.25$. This is extraordinarily high considering that a 100% diffusely reflecting surface (e.g., a piece of white paper) would have a ρ of unity. It is also an order of magnitude or more higher than $\rho_A(865)$ typically found over the open ocean (Figure 10.4). Also note the track from 23°N to 26°N, where $\rho_A(865)$ increases by over a factor of two, but $\rho_A(412)$ hardly increases at all, and $\rho_A(443)$ increases only about 25%. This is in the region where the absorption is so high that ρ_A increases only slightly with a large increase in τ_a , i.e., ρ_A is approaching an asymptotic value as $\tau_a \rightarrow \infty$ (saturation). In contrast, between 28°N and 33°N ρ_A at 412 and 443 seems to follow the variations in $\rho_A(865)$. This suggests that there are aerosol differences between these two latitude regions on the track, either in aerosol's physical properties (particle size or refractive index) or in the aerosol's vertical distribution. But before further analysis of this data, we need to show that the water's contribution to $\rho_A(\lambda)$ can be safely neglected.

To estimate the contribution of $t\rho_w$ at the other bands (it's taken to be zero at 865 nm), we need an estimate of ρ_w and an estimate of t . For this region it is reasonable to assume that $C_P \approx 0.5 \text{ mg/m}^3$. The model in the Appendix provides $[\rho_w(\lambda)]_N$ for this C_P . Then $t\rho_w = t(\theta_v)t_E(\theta_0)[\rho_w(\lambda)]_N$, but since the angular distribution of ρ_w is unknown, we take it to be uniform, so $t(\theta_v) = t^*(\theta_v)$. As before, we also take $t_E(\theta_0) = t^*(\theta_0)$, yielding $t\rho_w = t^*(\theta_v)t^*(\theta_0)[\rho_w(\lambda)]_N$. So we need to be able to estimate $t^*(\theta)$, and this requires the aerosol optical depth. From Figure 10.26 we can see that the track along which the data were extracted is located near the edge of the scan with the sun near the zenith (as indicated by the sun glitter mask). This suggests that $\theta_0 \approx 0$, $\theta_v^{(s)} \approx 45^\circ$, and $\phi_v - \phi_0 \approx 90^\circ$. To get a rough estimate of $\tau_a(865)$ consider Figure 10.8 for the M99 model. Figure 10.8 is for $\theta_0 = 60^\circ$ and $\theta_v^{(s)} = 0^\circ$; however, all of the top-of-atmosphere ρ 's satisfy the reciprocity principle in a particularly simple form: $\rho(\hat{\xi}_v, \hat{\xi}_0) = \rho(\hat{\xi}_0, \hat{\xi}_v)$, so Figure 10.8 also applies to $\theta_0 = 0^\circ$ and $\theta_v^{(s)} = 60^\circ$. Ignoring the difference for the moment between $\theta_v^{(s)} = 45^\circ$ and $\theta_v^{(s)} = 60^\circ$, Figure 10.8 suggests that $\rho_a + \rho_{ra} \approx 1.5 \rho_{as}$. We can calculate ρ_{as} directly from Eq. (12.29) using the phase function for M99 in Figure 10.9, yielding $\rho_{as}(865) \approx 0.053 \tau_a(865)$. Then, for the region where $\rho_a + \rho_{ra} \approx 0.1$ along the track, i.e., between 26 and 32°N , we find that the required value of $\tau_a(865)$ would be approximately 1.25. Thus, it is reasonable to expect for these regions, and for much of the data extracted along the track line in Figure 10.26, that $\tau_a(865) > 1$. We will, *conservatively*, assume $\tau_a(865) = 1$ in computing the diffuse transmittances, fully realizing that it could be, and likely is larger, making t^* and $t^*\rho_w$ even smaller.

Using $\tau_a(865) = 1$ and the formulas for t^* in Chapter 9, the resulting values of $t\rho_w$ are provided in Table 10.5 for two aerosol models: M99 and BDS1. Recall (Chapter 4) that BDS1 is an aerosol model that has been used extensively to estimate the optical properties of Saharan dust and as such has significant absorption, especially in the blue. In the case of BDS1, the t^* 's were computed "exactly," while those for M99 were approximated using the formulas in Chapter 9. Comparing the values of $\rho_A(\lambda)$ in Figure 10.27 with $t\rho_w(\lambda)$ in Table 10.5 suggests that $t\rho_w(\lambda)$ contributes at most a few percent to $\rho_A(\lambda)$, especially considering that $\tau_a(865) = 1$ is a *conservative* estimate. Thus, we will assume henceforth in this section that $\rho_A(\lambda) = \rho_a(\lambda) + \rho_{ra}(\lambda)$ along the track line shown in Figures 10.26 and 10.27, i.e., we ignore $t\rho_w$.

Let's now see how maritime and dust models compare when applied to the data in Figure 10.27. If we are given $\rho_A(865)$, then for a given aerosol model we can compute $\rho_a(\lambda) + \rho_{ra}(\lambda)$ as follows:

$$\rho_A(865) \xrightarrow{\text{Model}} \rho_{as}(865) \xrightarrow{\text{Model}} \tau_a(865) \xrightarrow{\text{Model}} \tau_a(\lambda) \xrightarrow{\text{Model}} \rho_a(\lambda) + \rho_{ra}(\lambda).$$

Figure 10.28 shows the results of this procedure using the M90 and the BDS1 models. For the computations in Figure 10.28 the vertical structure of the aerosol was the same in

Table 10.5: Estimated $t\rho_w$ for data along the track line in Figure 10.26. Two aerosol models are used in the estimate: M99 and BDS1. Also given are the estimated values of $[\rho_w]_N$ for value of $C_P = 0.5 \text{ mg/m}^3$.

λ (nm)	$[\rho_w]_N$	$t\rho_w$ M99	$t\rho_w$ BDS1
412	0.0081	0.0043	0.0025
443	0.0084	0.0047	0.0031
490	0.0103	0.0058	0.0044
510	0.0090	0.0052	0.0040
555	0.0057	0.0036	0.0028

the two cases — all confined to the lower layer of a two-layer atmosphere. The difference between the results for the two models is striking. In contrast to the experimental data, as expected the M90 model with very little absorption shows practically no spectral variation, while the BDS1 model captures the main features of Figure 10.27. There is no doubt that the dust is absorbing, and that it becomes more absorbing as one progresses from the red to the blue.

Although the BDS1 model with the dust confined to the surface captures the qualitative features of the data along the track, we need to see if it is in quantitative agreement. Also, we need to know quantitatively the influence of vertical structure. For this purpose, we considered the BDS models (Chapter 4) with the following two-layer vertical structures: all of the aerosol confined to a thin layer at the surface with all of the molecular scattering above it (V00); the aerosol mixed uniformly with air to an altitude of 2, 4, 6, or 8 km with only the air above it (respectively, V02, V04, V06, or V08); and the aerosol uniformly mixed with air throughout the entire atmosphere (VUU). Exact computations of $\rho_a + \rho_{ra}$ were then carried out for each wavelength and compared with the data extracted along the track line shown in Figure 10.26. Figure 10.29 provides an example of this extraction at 443 nm. The extracted data clearly show the two latitude regimes discussed earlier: one where $\rho_A(443)$ varies more or less linearly with $\rho_A(865)$ (Region 1) and one where it varies also linearly, but very slowly with $\rho_A(865)$ (Region 2). The comparison between exact computations, for two aerosol models, and the extracted data for 443 nm is provided in Figure 10.30. Plotted along with the data on the figure (left panel) are the model results for BDS1 and BDS3 having the vertical distributions V00, V04, and VUU. The results for BDS1 and V00 clearly suggest that this model is too absorbing over the Region 1 part of the track. The other vertical distributions shown are even more absorbing in Region 1. The BDS1 model can work in Region 2, but that requires it be mixed to heights above 4-6 km. However, the BDS3 model, which differs from BDS1 only by having twenty times more particles in the largest size mode, can explain the Region 2 data only if the aerosol is not mixed as high a 4 km. To try to achieve a better fit in Region 1, a new model BDB

was developed with the same size distribution as BDS but with a smaller absorption index (imaginary part of the refractive index). (The physical properties of the BDB model is described in Chapter 4.) Figure 10.30 (right panel) shows the results for this model. It fits Region 1 (BDB1) well and can also fit Region 2 if the aerosols are mixed high enough in the atmosphere with more large particles (BDB3). The absorption index as a function of wavelength for the BDB model was derived by carrying out the exercise described here for all of the other SeaWiFS bands. Because the data in Region 2 are closer to the most intense part of the plume, and hence closer to the source, it is reasonable to expect the largest-sized particles to still be in suspension. Farther away from the plume one might well expect some of the largest particles have settled out. So a variation of the particle size distribution over the area is not unlikely. However, as these dust particles are likely all from one source one would expect that the composition (refractive index) are all the similar, unless it depends significantly on particle size, i.e., a different absorption index in each size class. Thus, one would expect either the BDS or BDB models might be appropriate to this dust event, but not both, as they only differ in absorption index.

10.5.3.2.2 Application of SMA to Imagery Containing Windblown Dust

Now that we have seemingly appropriate models to use then windblown dust is present off the West African Coast, we can use these as candidates in the SMA in this region to effect retrieval of C_P . To carry this out, 18 candidate models have been used: BDS1, BDS2, BDS3, BDB1, BDB2, and BDB3, each with three vertical distributions — V02, V04, and V06. The results are then compared with the standard SeaWiFS algorithm (STD) described in Section 10.4. The STD estimates the concentration of Chlorophyll a , while the SMA estimates the pigment concentration C_P . Typically when measured at sea using fluorometry C_P is about 35% greater than C . However, as the $[\rho_w]_N$ model is not perfect, it can yield C_P 's that are actually lower than C . We will not try to reconcile these differences here, but will use the exercise only to assess the ability of the SMA to effect a reasonable atmospheric correction. Figure 10.31 compares the STD and the SMA for two days: October 3, 1997 (dusty) and September 29, 1997 (dust free). The area between 13°N and 22°N in the STD image for October 3 contains the dust and no retrievals are obtained there because of the SeaWiFS dust avoidance mask (i.e., these regions are not processed); however the SMA does retrieve reasonable values of C_P there, as evidenced by the similarity with the retrievals from the clear day (September 29). The speckled white areas in the imagery are clouds and are masked out (not processed) by both algorithms. The SMA and the STD use different cloud masking algorithms (SMA is more conservative), so in some areas there appear to be more clouds in the SMA than in the STD. The consistency between the SMA retrievals for the two days suggests success in atmospheric correction. Figure 10.32 provides the average aerosol optical depth retrieved over the time period September 30 to October 7 and shows that the mean $\tau_a(865)$ is close to unity over much of this plume, while outside it is ~ 0.15 . However, outside the plume, the dust models are

probably not realistic (and non-absorbing aerosol are *not* included as candidates), so the τ_a retrieval is likely too large outside the plume, e.g., in Figure 10.10 compare the $\omega_0 = 1$ with the $\omega_0 = 0.8$ curves.

This example of the development of models for, and application of, the SMA to atmospheric correction in a region of strongly-absorbing aerosols has been presented as a case study. The goal was to define a procedure when such aerosols are encountered. To effect atmospheric correction one must first develop site-specific aerosol models, as the nature of the aerosol from one region to another is variable, i.e., Saharan dust does not have the same chemical composition or optical properties as dust from the Gobi Desert. Next, one must have some understanding regarding the vertical distribution of the aerosol. Finally, one needs a reliable bio-optical model to be able to predict $[\rho_w(\lambda)]_N$ as a function of the constituent concentrations. Only by possessing knowledge of all three of these components can one hope for adequate atmospheric correction.

10.5.3.3 Further Comments Regarding Absorbing Aerosols

In the example above, one of the retrieved quantities was C_P , but it was compared with C . This can be remedied by using a $[\rho_w]_N$ model based on C as developed in Chapter 6. Another approach is to use the SMA only to provide the best aerosol model at a given pixel, and then use that particular aerosol model in the STD algorithm for estimating $[\rho_w(\lambda)]_N$. The estimated $[\rho_w(\lambda)]_N$ can then be used in the standard bio-optical algorithms to retrieve C . This procedure was used to study the distribution of C in the Arabian Sea during the Southwest Monsoon time period when much of the Northern Arabian Sea is obscured by dust. Figure 10.33 compares the results of the SMA and STD algorithms for this region using the dust models developed in the previous section. Note the considerable increase in coverage for the summer months by the SMA and the excellent agreement between the two in regions outside the dust.

The vertical distribution of the aerosol requires further comment. In the SMA we have used vertical distributions for which the dust is extant from the surface to a specific altitude, e.g., 2, 4, 6 km, etc. But what if the dust exists in a *layer* above the surface? This can lead to further complications. Consider Figure 10.34, which provides the $\rho_A(412) - \rho_A(865)$ relationship for two vertical distributions: V04 and V26. For V04 the aerosol is uniformly mixed with air from the surface to 4 km, while for V26 the same optical depth of aerosol is mixed from 2 km to 6 km with no aerosol from the surface to 2 km. Thus, the layers have the same physical (and optical) thickness, but differ only in their position vertically in the atmosphere. The difference in the computed $\rho_A(412)$ is striking. The V26 result is actually very close to the result for V08. Clearly, in general we not only need to know the thickness of the layer, *we need to know where it is in the atmosphere*.

The strong influence of the vertical distribution of particle absorption on the reflectance of the ocean-atmospheric system can be understood with a simple example. Consider an aerosol with $\omega_a = 0$, i.e., totally absorbing. If it is in a thin layer at the surface, only that part of ρ_r that interacts with the surface (the Fresnel reflectance terms in Eq. (12.29)) is influenced by the aerosol. They account for only about 5% of ρ_r . If the aerosol is in a layer at the top of the atmosphere, the solar beam is attenuated before creating the Rayleigh reflectance, and then the created ρ_r is attenuated again propagating to the TOA, i.e., $\rho_r \rightarrow \rho_r \exp[-M \tau_{\text{Abs}}]$, where M is the airmass $[(\cos \theta_0)^{-1} + (\cos \theta_v^{(s)})^{-1}]$ and τ_{Abs} is the absorption optical depth of the aerosol. The latter ρ_r is much smaller than the former: the TOA-reflectance becomes smaller as the absorbing layer moves higher into the atmosphere, as in Figure 10.34.

Increasingly, LIDAR systems are being placed on Earth observing systems for the purpose of measuring the cloud and aerosol backscattering. These, like LITE (described in Chapter 4), can provide the aerosol concentration as a function of altitude. Newer systems, that measure the polarization of the backscattered light as well as the magnitude, can give an indication of the aerosol type, e.g., mineral (dust) as opposed to maritime (water). Thus, vertical distributions of aerosol will be available for used by future ocean color scanners. Incorporation of such information into atmospheric correction schemes in an efficient manner is left to future investigators — perhaps the reader.

10.6 Summary and Discussion

In this chapter we have derived an atmospheric correction procedure using the single-scattering analysis and its algorithm as a guide. This algorithm was particularly appealing because it only required a knowledge of the spectral variation of ρ_{as} , which could be estimated from that in the NIR. However, because of the severe accuracy requirements on the retrieved $[\rho_w]_N$ in the visible, we showed that the single-scattering algorithm often was not up to the task, and that attention to multiple scattering was required. As multiple scattering effects depend significantly on the aerosol phase function and single scattering albedo, more complete information is required to address them. This forced the use of detailed models of the aerosol — considerably more information than was required in the single scattering case. Since inclusion of multiple scattering required solving the radiative transfer equation for the particular geometry and aerosol concentration, this required off-line solution of the equation and the storage of, and the use of, the spectral variation of ρ_A in the form of lookup tables that are accessed at each pixel to be corrected. Fortunately, for non- or weakly-absorbing aerosols — the usual situation — the spectral variation of ρ_A is nearly independent of the vertical distribution of the aerosol, greatly reducing the complexity of the lookup tables. Simulations indicate that the final multiple-scattering

algorithm can retrieve $[\rho_w]_N$ with the needed accuracy, but only in situations in which the aerosol is non-absorbing or weakly absorbing. This is borne out by several years of successful processing of SeaWiFS imagery. It is important to note that the aerosol models that are used in the algorithm need not be precise. For atmospheric correction they need only provide the correct spectral variation of ρ_A , i.e., $\varepsilon(\lambda_i, \lambda_l)$. In contrast, to retrieve the aerosol optical depth, it is necessary that the models provide a good representation of the actual aerosol phase function and aerosol absorption properties, as ρ_A is approximately $\propto \omega_a \tau_a [P(\Theta_-) + r(\theta_0) + r(\theta_v)P(\Theta_+)]$. Error in ω_a or $P(\Theta)$, especially in the backward directions (Θ_-), translates directly to error in τ_a . The reader will note that our development has been based on scalar radiative transfer. Extension to vector radiative transfer (i.e., including polarization) only requires recalculation of the lookup tables using a vector radiative transfer code, and this has been done.

We then examined the problems associated with atmospheric correction in the presence of strongly-absorbing aerosols, using urban aerosols as an example. An algorithm was proposed — the spectral matching algorithm — in which atmospheric correction and the retrieval of constituent concentrations are effected in a single process. A case study was provided in which the strongly-absorbing aerosol was windblown Saharan dust. In this study, satellite imagery from the specific region in question was used to develop appropriate aerosol models to use in the correction process. With these models the algorithm successfully retrieved the pigment concentration in dusty areas off the West Coast of Africa during a dust outbreak. In addition, combining the dust algorithm with the standard algorithm allowed estimation of the concentration of Chlorophyll *a* in the Arabian Sea during the Southwestern Monsoon.

What we have so far then is an algorithm that will successfully retrieve the water-leaving reflectance in most cases. However, there are additional complications that must be faced in special circumstances. These include, but are of course not limited to, (1) the effect of the curvature of the earth when imagery are acquired at large θ_0 , e.g., high-latitude imagery, (2) the effect of thin cirrus clouds as well as stratospheric aerosols resulting from volcanic sources, and (3) the failure to address the problem of replacing t by t^* . In addition, there are many issues that must be faced concerning unintended instrumental effects such as the sensitivity of the instrument to the polarization state of the radiance being measured, and the out-of-band response of the instrument. Dealing with these and other complications would take us too far afield from the main subject of this monograph, but they are dealt with in references provided in the Bibliographic Notes.

10.7 Appendix: The SMA Water-Leaving Reflectance Model

In Chapter 7, Section 4 the normalized-water-leaving radiance $[L_w]_N$ was related to the subsurface upwelling radiance through

$$[L_w(\hat{\xi}_v, \hat{\xi}_s)]_N = \Re(\hat{\xi}_v, \hat{\xi}_s) \bar{F}_0 \left(\frac{L_u(\hat{\xi}'_v, \hat{\xi}'_s)}{E_d(\hat{\xi}'_s)} \right),$$

where

$$\Re(\hat{\xi}_v, \hat{\xi}_s) = \frac{t_f(\hat{\xi}_v, \hat{\xi}'_v) t_f(\hat{\xi}_s, \hat{\xi}'_s)}{m_w^2 (1 - rR(\hat{\xi}'_s))}.$$

Recall that the water-leaving reflectance $[\rho_w]_N$ is related to the water-leaving radiance through $[\rho_w]_N = \pi[L_w]_N/\bar{F}_0$, so modeling $[\rho_w]_N$ requires modeling L_u/E_d .²³

The model used in the SMA was developed when the modeling of AOPs in terms of IOPs was in its infancy, and therefore is far less sophisticated than those in Chapter 6. The radiative transfer part was developed from relationships found in Section 6.4.9 of Chapter 6, in which early studies suggested that

$$\frac{L_u}{E_d} = \sum_{N=1}^2 g_n \frac{b_b}{a + b_b} \quad \text{and} \quad \frac{K_d(0)}{D_0} \approx a + b_b,$$

so, taking only the first term in the sum,

$$\frac{L_u}{E_d} \approx g_1 D_0 \frac{b_b}{K_d(0)}.$$

Here, $g_1 \approx 0.1$ and typically $1 < D_0 < 1.2$ and is nominally taken to be 1.1. The downwelling irradiance attenuation coefficient was modeled by [Baker and Smith \[1982\]](#) for Case 1 waters as

$$K = K_w + K_c, \tag{10.20}$$

where K_w and K_c are the partial contributions to K due to water and phytoplankton, with

$$K_c = k_c C_P \exp \left[- [k'_c \log_{10}(C_P/C_{P0})]^2 \right] + 0.001 C_P^2,$$

where k_c , k'_c , and C_{P0} are dependent on wavelength and have been tabulated for $C_P < 10$ mg/m³. The SMA water-leaving reflectance model assumes $K_d(0) = K$, as modeled in Eq.

²³It also requires modeling R in \Re . This is easy since $R = QL_u/E_d$, and because of the smallness of R (typically < 0.1) Q can be taken anywhere between 3 and 6 with little affect on the result. So modeling L_u/E_d also provides the needed model for R .

(10.20). The model for b_b is a cruder version of those discussed in Chapter 5. We note that $b_b = b_{bw} + b_{bp}$, and write $b_{bp} = B_p b_p$. Then, b_p is taken to be $b_p = b^0 C_P^{0.62}$, where $0.12 < b^0 < 0.45 \text{ m}^{-1}$, with a mean of 0.3 m^{-1} and C_P is in mg/m^3 . The backscattering probability varies from 2% at $C_P = 0.1 \text{ mg}/\text{m}^3$ to 0.5% at $C_P = 20 \text{ mg}/\text{m}^3$ (within absorption bands, e.g., 443 nm, $B_p = 0.3\%$ at $C_P = 20 \text{ mg}/\text{m}^3$). Between these C_P limits, B_p follows a power-law variation with C_P . The particle scattering coefficient is taken to vary with wavelength as λ^{-1} . Thus, the water-leaving reflectance becomes a function of three parameters: b_p^0 , C_P and λ .

Examples of the quality of the model in representing $[\rho_w(\lambda)]_N$ as a function of C_P are provided in Figure 10.35, for the mean and the extreme values of b^0 . Figure 10.36 shows the ratio $[\rho_w(443)]_N/[\rho_w(550)]_N$ as a function of C_P for $b^0 = 0.2 \text{ m}^{-1}$. Clearly the model captures much of the variation of $[\rho_w(\lambda)]_N$ with C_P .

10.8 Bibliographic Notes

Most of the work described in the chapter was carried out by the author and co-workers. A significant portion of the material (some verbatim) is drawn directly from Gordon [1997], albeit with some enhancements.

10.1 Introduction

The data shown in Figure 10.1, converted to $[\rho_w]_N$ from $[L_w]_N$, was taken from Gordon et al. [1988]. The data were also presented in Gordon [1997], although most the $[L_w]_N$ data had already been presented in Gordon and Clark [1981], where the concept of normalized water-leaving radiance was first introduced. The global aerosol network, AERONET, is describe in Holben et al. [1998].

10.3 A Single Scattering Solution

The single scattering solution was first provided in Gordon [1978b] and Gordon et al. [1983], and in the form presented in this section by Wang and Gordon [1994]. Figure 10.3 is from Gordon et al. [1983]. The data in Figure 10.4 were taken close to the Hawaiian Islands on the SeaWiFS initial calibration-validation cruise organized and directed by Dennis Clark.

10.4.1 The Multiple Scattering Retrieval Algorithm

The multiple scattering algorithm first appeared in Gordon and Wang [1994] in basically

the form presented here.

10.4.2 Simulated Test of the Multiple-Scattering Algorithm

The tests here are from [Gordon \[1997\]](#).

10.4.4 The Diffuse Transmittance

The development of t and t^* in this section follows that in [Yang and Gordon \[1997\]](#). A demonstration of the magnitude of the error $(t - t^*)/t$ for actual SeaWiFS data is presented in [Gordon and Franz \[2008\]](#), and suggests that the actual error in replacing t by t^* is $\sim 1\%$ as long as the viewing angle is less than about 60° .

10.4.7 Non-zero $[\rho_w]_N$ in the NIR

The “black pixel” approximation in the NIR is examined by [Siegel et al. \[2000\]](#) and [Bailey et al. \[2010\]](#) along with an iterative method to overcome its failure. [Wang and Shi \[2007\]](#) and [Wang et al. \[2009\]](#) discuss using the SWIR in addition to the NIR for atmospheric correction in turbid coastal environments.

10.5 Absorbing Aerosols

[Gordon et al. \[1997b\]](#) first introduced the spectral matching algorithm for absorbing aerosol. Much of the material in this section is from [Gordon \[1997\]](#).

10.5.3.2.1 Development of Appropriate Aerosol Models for Windblown Dust

The material here is mostly from [Moulin et al. \[2001\]](#).

10.5.3.2.2 Application of SMA to Imagery Containing Windblown Dust

This is from [Moulin et al. \[2001\]](#).

10.5.3.2.3 Further Comments Regarding Absorbing Aerosols

The summer month panels of Figure [10.33](#) appeared in [Banzon et al. \[2004\]](#).

10.6 Summary and Discussion

The extension of the standard atmospheric correction to large values of the solar zenith angle for which the curvature of the Earth is important is provided in [Ding and Gordon \[1994\]](#). They showed that the [Gordon and Wang \[1994\]](#) algorithm was still valid in such

situations as long as the Rayleigh contribution was computed taking into account the curvature of the Earth. Correction of the interference of the O₂ “A” absorption band on the 765 nm band of SeaWiFS is described in [Ding and Gordon \[1995\]](#) and a general method for dealing with out-of-band spectral sensitivity is provided in [Gordon \[1995\]](#). The problem of correcting sensor measurements for polarization sensitivity is developed in [Gordon et al. \[1997a\]](#). The general subject of the diffuse transmittance is given in [Yang and Gordon \[1997\]](#) and in the context of replacing t by t^* in [Gordon and Franz \[2008\]](#). Some effects of stratospheric aerosols and thin cirrus clouds are studied in [Gordon et al. \[1996\]](#).

Finally, an alternative to the SMA called the spectral optimization algorithm (SOA) has been developed for dealing with absorbing aerosols of the urban variety [[Chomko and Gordon, 1998, 2001](#); [Chomko et al., 2003](#)]. Here power-law size distributions are used, and a matrix of refractive indices, allowing essentially a continuum of models through interpolation. Classical optimization techniques are then used to effect a solution. This method was applied to Case 2 waters by [Kuchinke et al. \[2009\]](#) and [Kuchinke et al. \[2009\]](#).

10.7 Appendix: The SMA Water-Leaving Reflectance Model

This model was originally developed in [Gordon et al. \[1988\]](#). Although more sophisticated models now exist (e.g., see Chapter 6), this is the one that was used with the original SMA.

10.9 Figures

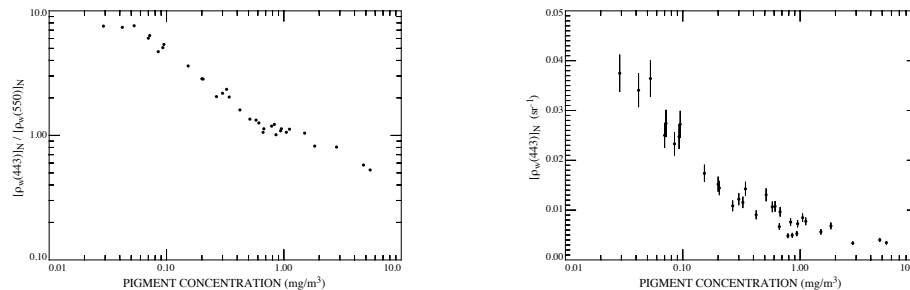


Figure 10.1: Normalized water-leaving reflectance ratio (443:550) as a function of pigment concentration (left panel). Normalized water-leaving reflectance (443) as a function of pigment concentration (right panel). From [Gordon \[1997\]](#), redrawn from [Gordon et al. \[1988\]](#).

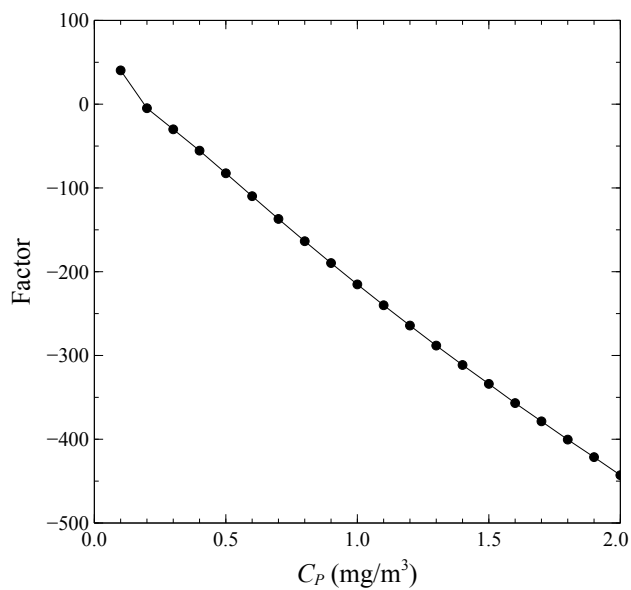


Figure 10.2: “Factor,” the term in the curly brackets in Eq. (10.5) is plotted here as a function of C_P .

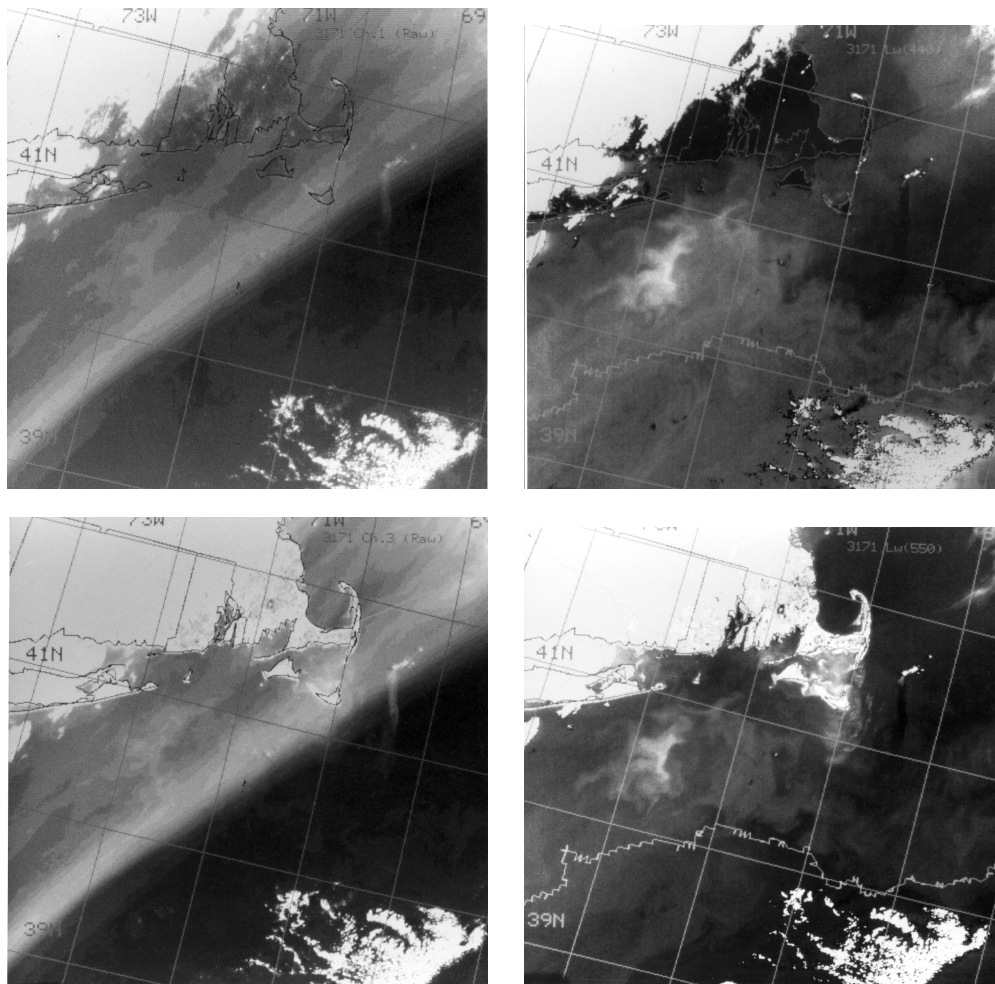


Figure 10.3: $L_t(443)$ (top left panel), $L_w(443)$ (top right panel), $L_t(550)$ (bottom left panel) and $L_w(550)$ (bottom right panel) for a CZCS image from Orbit 3171 over the Middle Atlantic Bight. The L_w retrieval was based on the method described in the text. A warm core ring located between the cloud and the lower right corner of the image was used as the “clear water” area for determining $\varepsilon(550, 670)$. From [Gordon et al. \[1983\]](#).

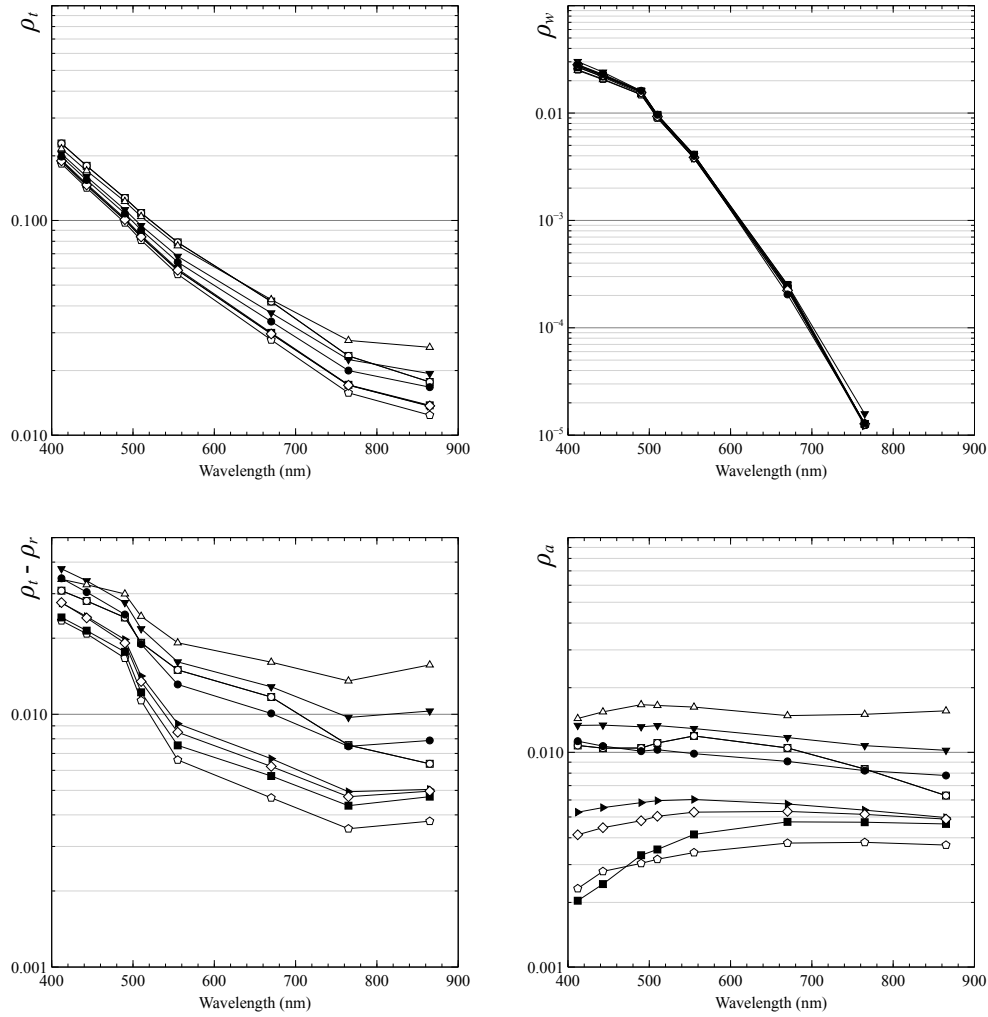


Figure 10.4: Top of atmosphere reflectance data ρ_t from eight SeaWiFS images (top left) taken in oligotrophic waters near Hawaii coincident with surface measurements of ρ_w (top right) obtained from a ship at the same location. The lower figures are the derived values of $\rho_t - \rho_r$ (bottom left) and $\rho_t - \rho_r - t\rho_w = \rho_a$ (bottom right).

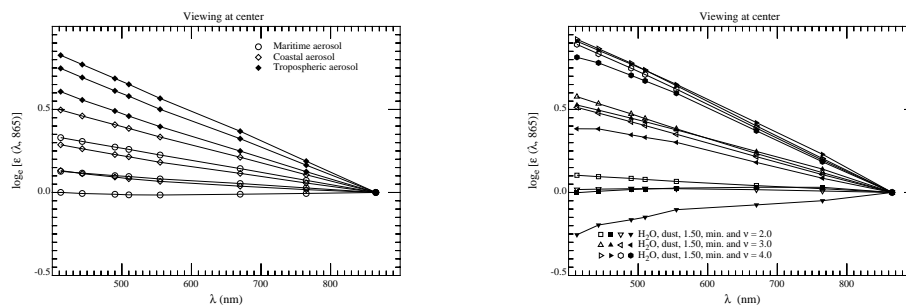


Figure 10.5: $\varepsilon(\lambda, 865)$ for near-nadir viewing with $\theta_0 = 60^\circ$. Left panel: Maritime (M), Coastal (C), and Tropospheric (T) aerosol models (the relative humidity, RH, values are 50, 80, and 98% from the upper to the lower curves, so the model for the top curve is referred to in the text as “T50” and the bottom as “M98”). Right panel: Haze C models. For both panels the open symbols are for models with practically no absorption. The filled symbols are for models with weakly- or strongly-absorbing aerosols. From [Gordon \[1997\]](#).

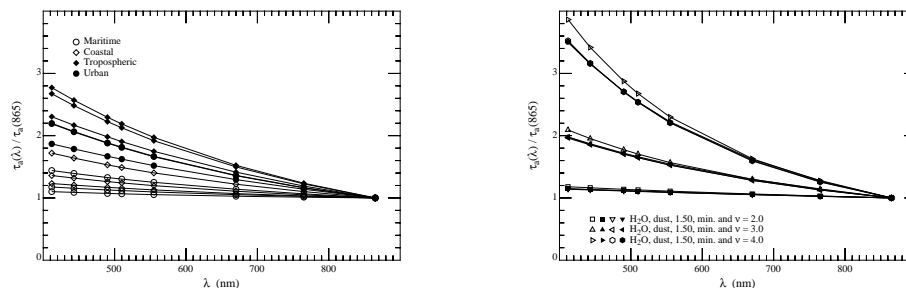


Figure 10.6: Spectral variation of the total (absorption plus scattering) aerosol optical depth τ_a . Left panel: Maritime, Coastal, Tropospheric and Urban aerosol models (the RH values are 50, 80, and 98% from the upper to the lower curves). Right panel: Haze C models. Absorbing aerosol models are indicated as in [Figure 10.5](#). From [Gordon \[1997\]](#).

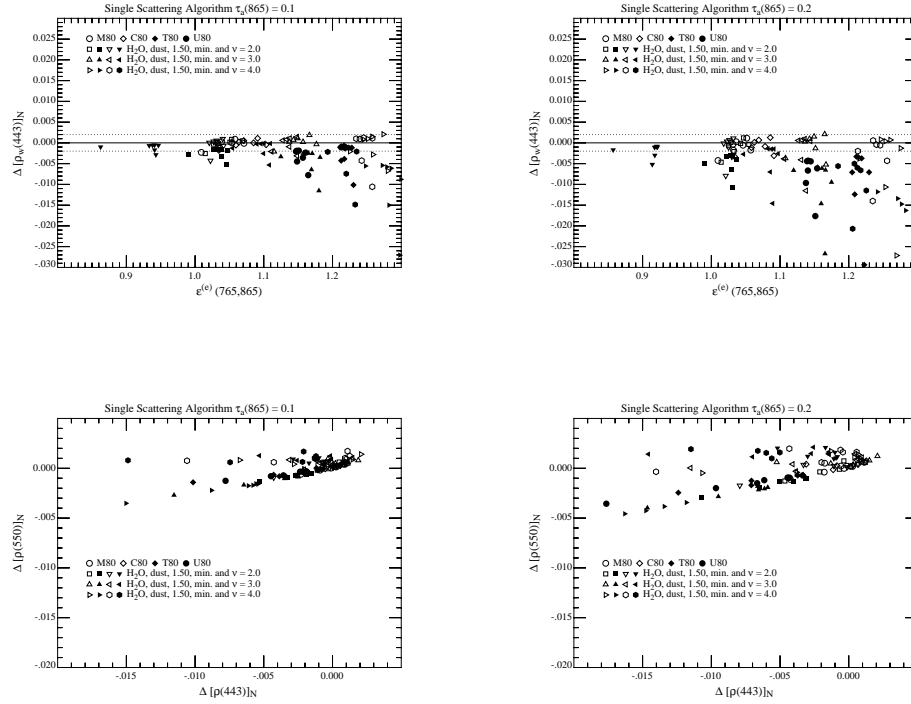


Figure 10.7: $\Delta[\rho_w(443)]_N$ as a function of $\varepsilon^{(e)}(765,865)$ and $\tau_a(865)$ (top) and $\Delta[\rho_w(550)]_N$ as a function of $\Delta[\rho_w(443)]_N$ (bottom) for all of the aerosol models and the seven Sun-viewing geometries: $\theta_v^{(s)} \approx 1^\circ$ and 45° with $\theta_0 = 0, 20^\circ, 40^\circ$, and 60° and $\phi_v - \phi_0 = 90^\circ$, but excluding $\theta_v^{(s)} \approx 1^\circ$ with $\theta_0 = 0$ because of Sun glitter. The dashed lines in the upper panels indicate the goal of ± 0.002 for the error in $[\rho_w(443)]_N$. Left panels: $\tau_a(865) = 0.1$. Right panels: $\tau_a(865) = 0.2$. Open symbols are for aerosol with very little (or no) absorption. Filled symbols are for aerosols with moderate to strong absorption. From Gordon [1997].

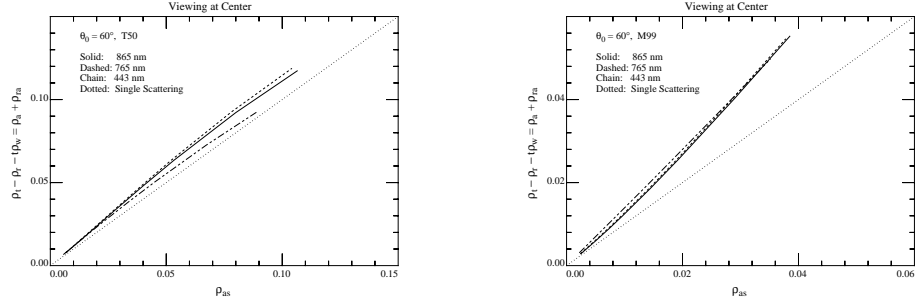


Figure 10.8: The $\rho_a(\lambda) + \rho_{ra}(\lambda)$ versus $\rho_{as}(\lambda)$ relationship at $\theta_0 = 60^\circ$ and near-nadir viewing. Left panel: T50. Right panel: M99. From [Gordon \[1997\]](#).

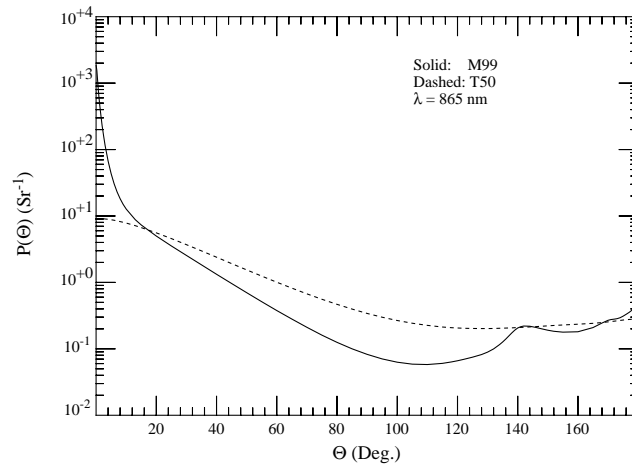


Figure 10.9: Phase functions for the T50 and M99 models at 865 nm. From [Gordon \[1997\]](#).

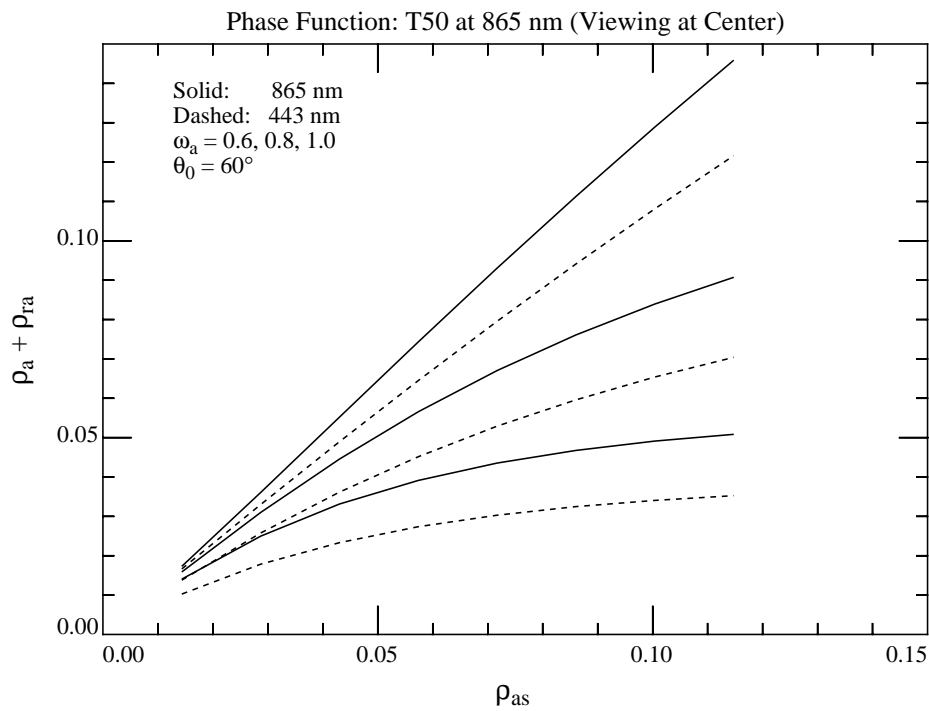


Figure 10.10: Computations of the $\rho_a + \rho_{ra}$, as a function of ρ_{as} , for examining the effects of aerosol absorption. This is similar to Figure 10.8, but obtained using the T50 model phase function (for 865 nm, Figure 10.9) with ω_a values of 0.6, 0.8, and 1.0 (increasing from bottom to top), rather than the values given by the model. From Gordon [1997].

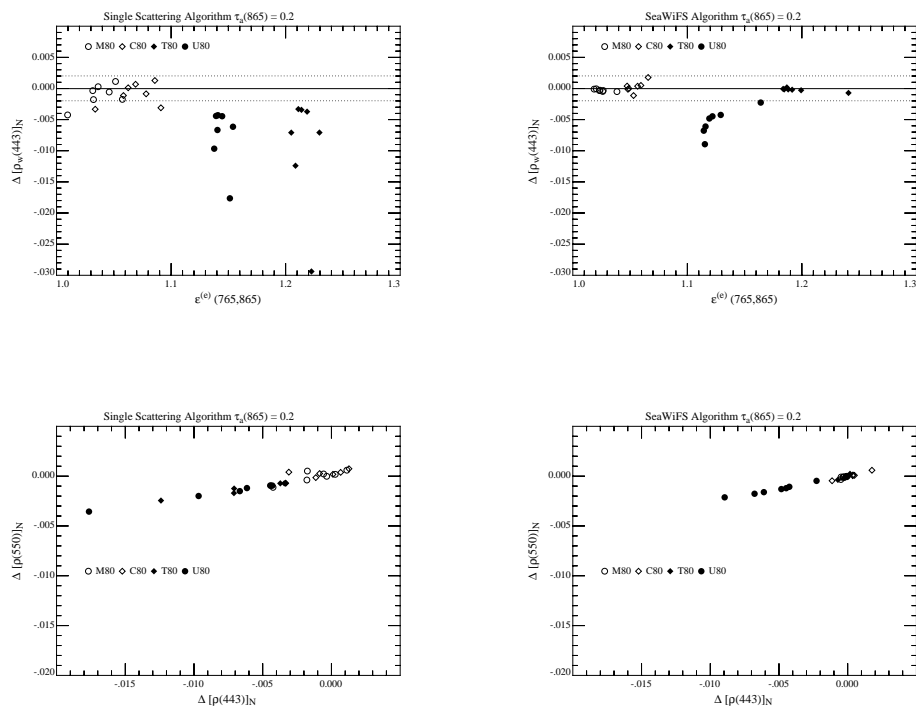


Figure 10.11: Similar to Figure 10.7, but compares single (CZCS) and multiple (SeaWiFS) scattering algorithms. Left panels: Single scattering algorithm. Right panels: Multiple scattering algorithm. Open symbols are for aerosol with very little (or no) absorption. Filled symbols are for aerosols with moderate to strong absorption. From Gordon [1997].

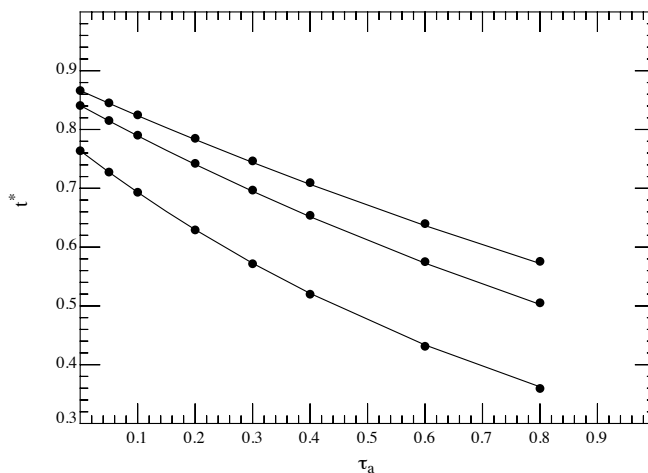


Figure 10.12: Effect of aerosol vertical structure on t^* . Computations of $t^*(\theta_0)$ for various τ_a and the U50 aerosol model. Lines correspond to placing the aerosol at the bottom of a two-layer atmosphere. Dots correspond to the aerosol being uniformly mixed with air molecules throughout the atmosphere. Curves from top to bottom correspond to $\theta_0 = 20^\circ$, 40° and 60° , respectively.

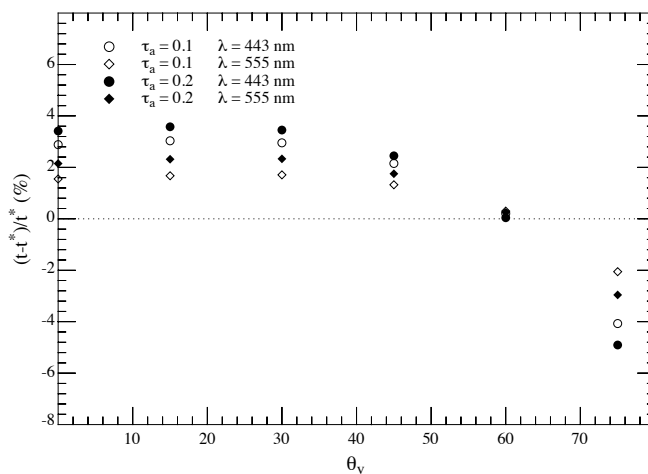


Figure 10.13: Percent error in the diffuse transmittance when t^* is used in place of t and the angular distribution of L_u is determined by the Petzold phase function (Phase-T). The viewing angle is θ_v (angle between zenith and the propagation vector) and the viewing azimuth angle is 90° relative to the sun. The aerosol is represented by the M90 model and $\theta_0 = 40^\circ$. From [Yang and Gordon \[1997\]](#).

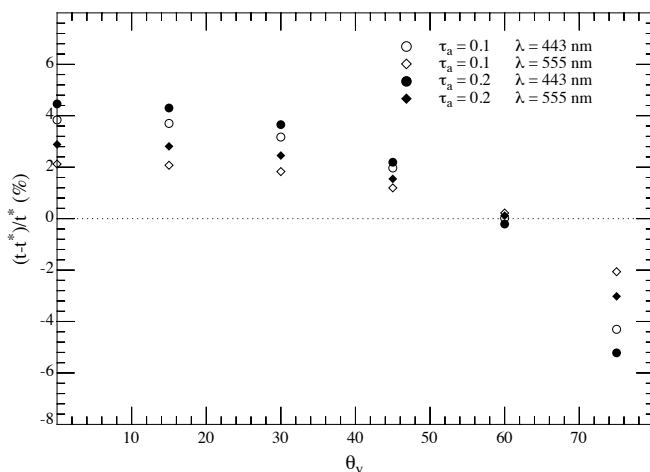


Figure 10.14: Percent error in the diffuse transmittance when t^* is used in place of t and the angular distribution of L_u is determined by the Petzold phase function (Phase-T). The viewing angle is θ_v (angle between zenith and the propagation vector) and the viewing azimuth angle is 90° relative to the Sun. The aerosol is represented by the M90 model and $\theta_0 = 60^\circ$. From [Yang and Gordon \[1997\]](#).

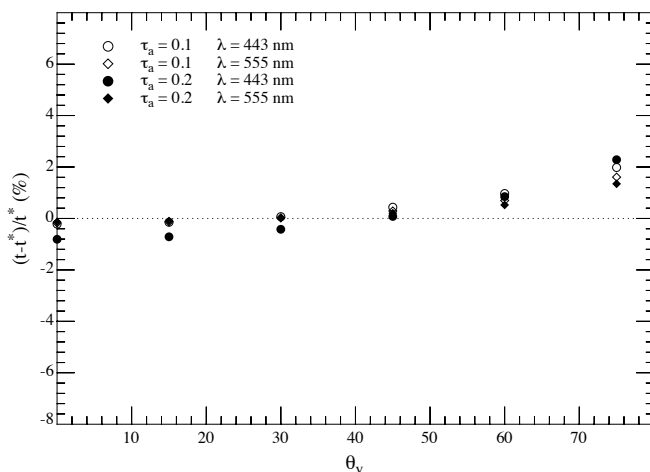


Figure 10.15: Percent error in the diffuse transmittance when t^* is used in place of t and the angular distribution of L_u is determined by the Rayleigh phase function. The viewing angle is θ_v (angle between zenith and the propagation vector), the viewing azimuth angle is 90° relative to the Sun, and $\theta_0 = 40^\circ$. From [Yang and Gordon \[1997\]](#).

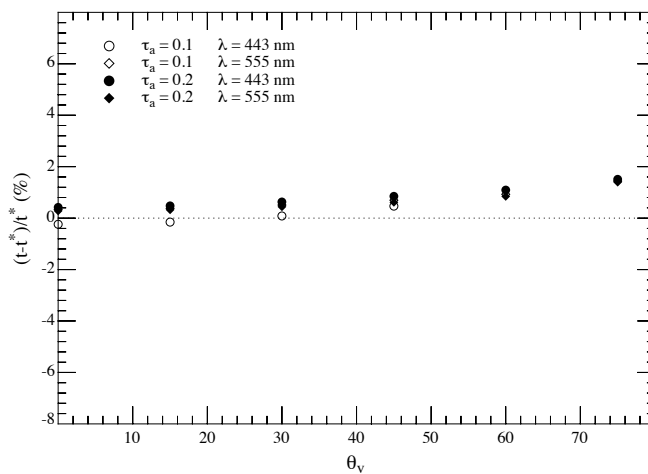


Figure 10.16: Percent error in the diffuse transmittance when t^* is used in place of t and the angular distribution of L_u is determined by the Rayleigh phase function. The viewing angle is θ_v (angle between zenith and the propagation vector), the viewing azimuth angle is 90° relative to the Sun, and $\theta_0 = 60^\circ$. From [Yang and Gordon \[1997\]](#).

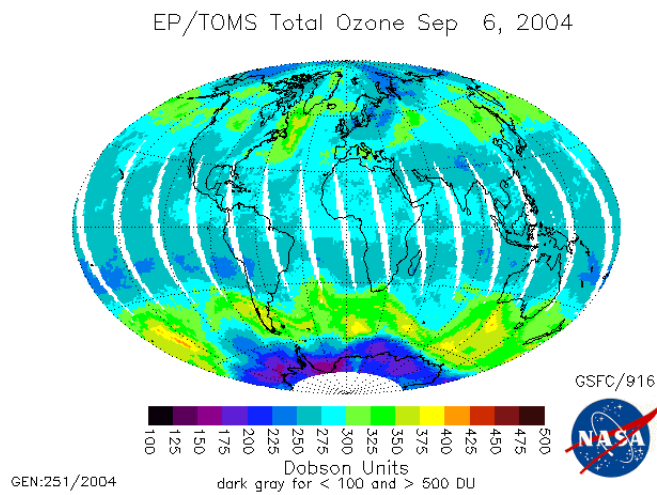


Figure 10.17: The concentration of Ozone for a single day as revealed by TOMS. Note the range of variability and the weather-like patterns in the concentration.

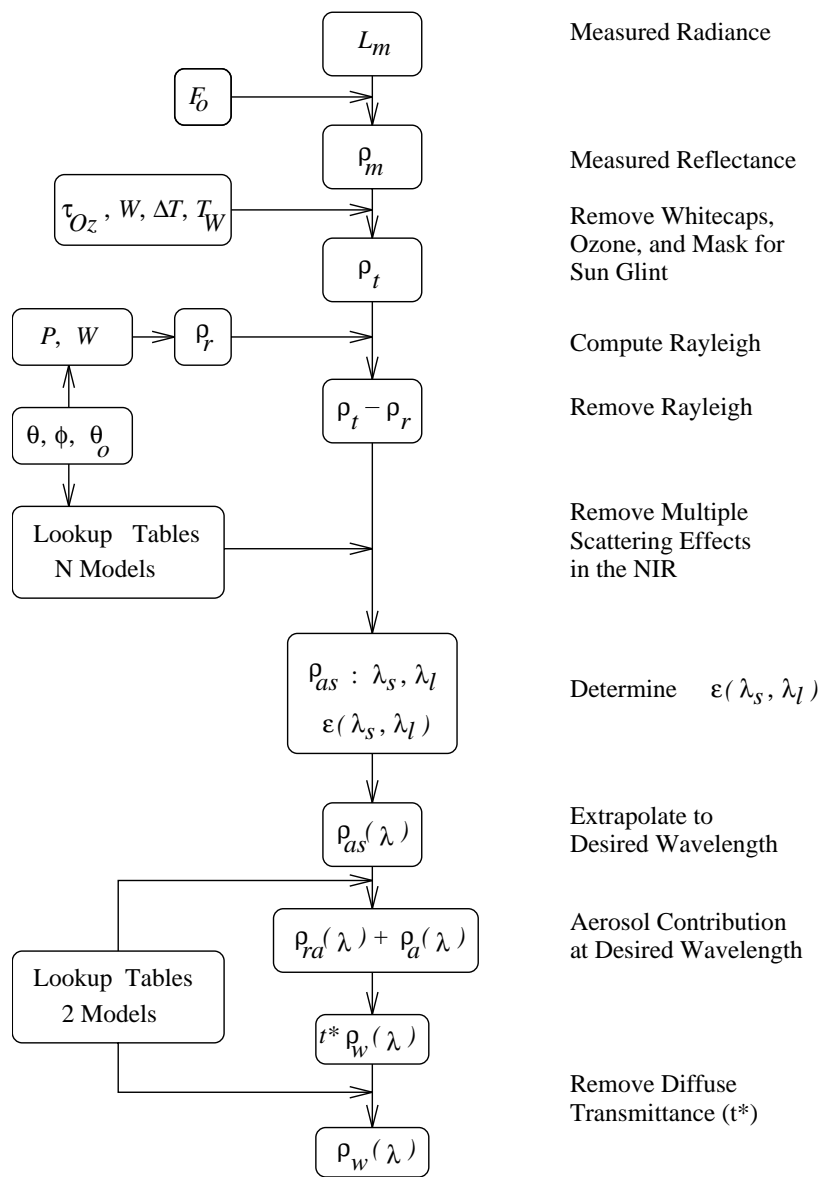


Figure 10.18: Annotated flow diagram of the algorithm.

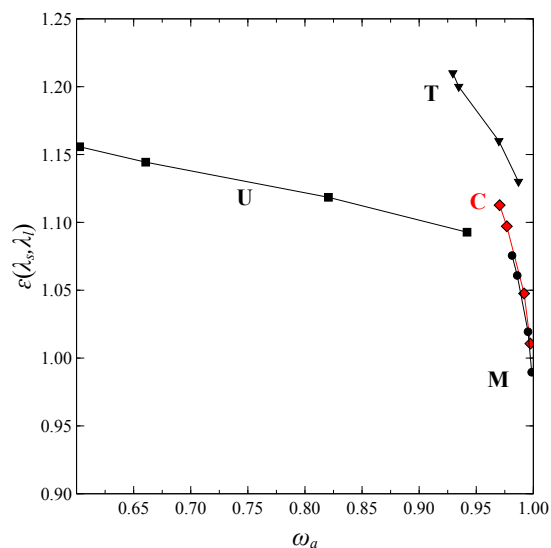


Figure 10.19: The relationship between $\varepsilon(\lambda_s, \lambda_l)$ and $\omega_a(\lambda_l)$ for the twelve original candidate aerosol models: M (circles), C (diamonds, red), and T (triangles). For each model the relative humidities are 99% for the lowest ε to 50% for the highest, with 90% and 70% in between. The model marked U (squares) is the urban model with U50 at the far left and U99 at the far right. The position of the “U” symbol on the plot gives the approximate value of ω_a for U80.

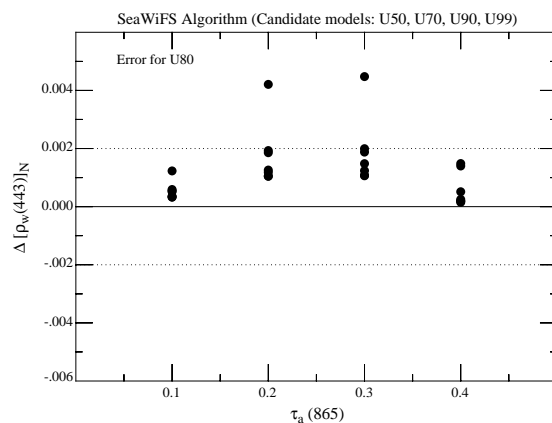


Figure 10.20: The error in $\Delta[\rho_w(443)]_N$ as a function of $\tau_a(865)$ for the U80 model for the seven sun-viewing directions, when the candidate aerosol models in the multiple-scattering algorithm are restricted to U50, U70, U90, and U99. From [Gordon \[1997\]](#).

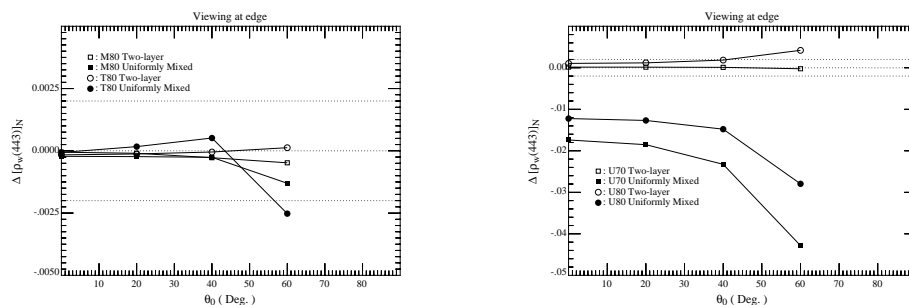


Figure 10.21: The effect of the vertical distribution of aerosol on the error $\Delta[\rho_w(443)]_N$ as a function of θ_0 at the edge of the scan $\tau_a(865) = 0.2$ using the multiple scattering algorithm that assumes that the “Two-layer” stratification is correct. Left panel: T80 and M80. Right panel: U70 and U80. From [Gordon \[1997\]](#).

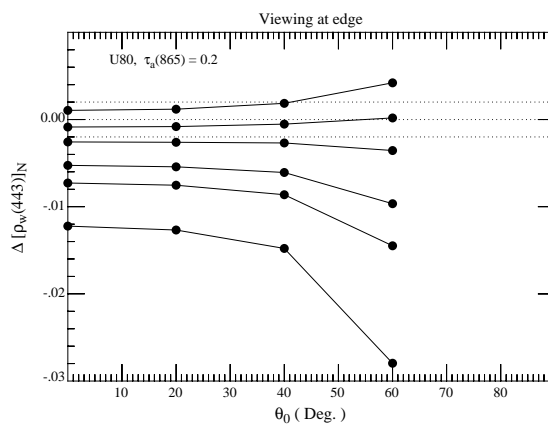


Figure 10.22: Effect of the vertical distribution of aerosol on the error $\Delta[\rho_w(443)]_N$ as a function of θ_0 at the edge of the scan for the U80 aerosol models with $\tau_a(865) = 0.2$. Curves from top to bottom refer to situations in which the aerosol is confined to a layer just above the surface, between the surface and 1, 2, 4, and 6 km, and uniformly mixed throughout the atmosphere. From [Gordon \[1997\]](#).

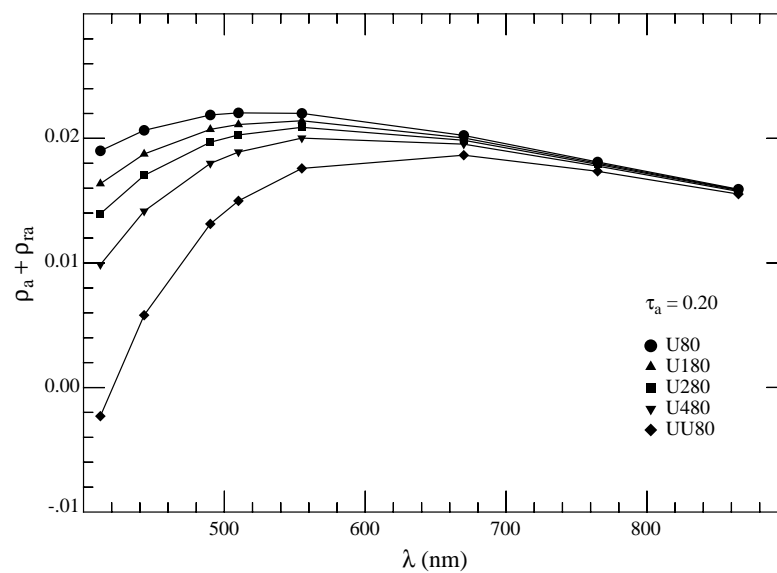


Figure 10.23: Influence of the physical thickness of the aerosol layer on the spectrum of $\rho_a + \rho_{ra}$. For U80 the aerosol is confined to a thin layer near the surface, while for U180, U280, U480, and UU80, the aerosol is uniformly mixed with air to a height of 1 km, 2 km, 4 km, and the whole atmosphere, respectively. Viewing is near nadir and $\theta_0 = 60^\circ$. ([Gordon et al. \[1997b\]](#).)

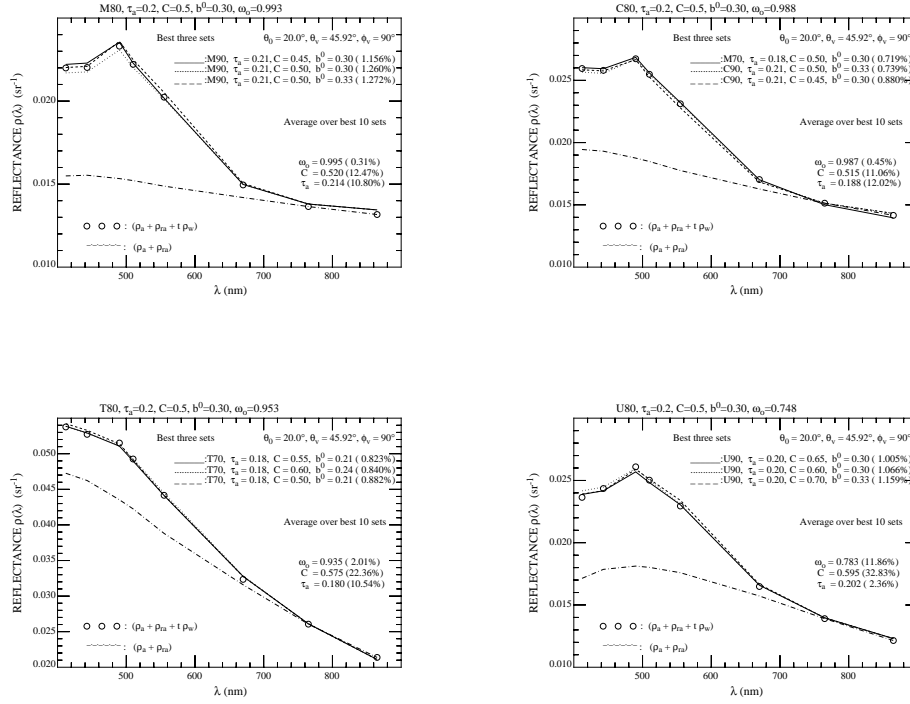


Figure 10.24: Results of tests of the SMA using M80, C80, T80 and U80 to create the pseudo data for ρ_t . Given along the top of each panel are the values of τ_a at 865 nm, C_P (called C on the figure) in mg/m 3 , b^0 in m $^{-1}$ and ω_a (called ω_0 on the figure) at 865 nm. The data points are $\rho_t - \rho_r$ and the lines close to the data points are the results of the three best models. The labels for these lines provide the retrieved model, τ_a , C_P , b^0 and, in parenthesis, δ in %. Note that the SMA has no problem distinguishing between very-weakly-absorbing aerosols (M80 and C80), weakly-absorbing aerosols (T80) and strongly-absorbing aerosols (U80). The averages for ω_a , C_P and τ_a over the ten best models and their relative standard deviation in %, are also provided. The lower curve (chain dashed) gives $\rho_a + \rho_{ra}$ for the best model. The parameter listed as θ_v is actually $\theta_v^{(s)}$. From [Gordon et al. \[1997b\]](#).

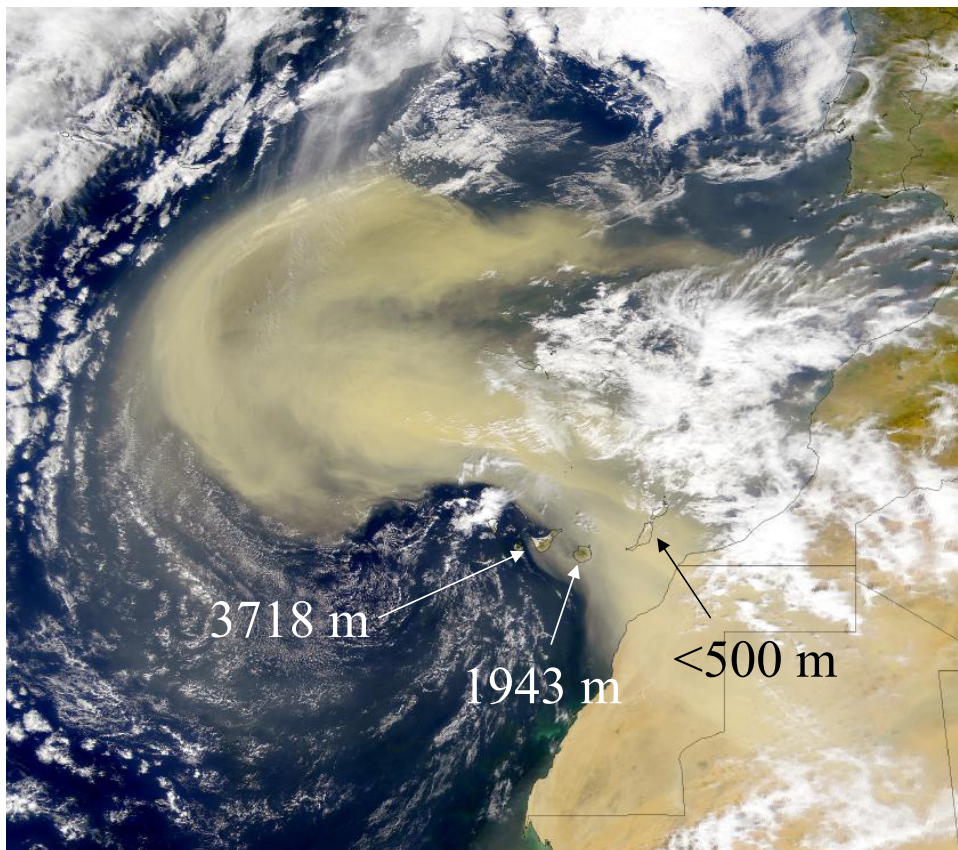


Figure 10.25: SeaWiFS image of a Saharan dust plume flowing from Africa into the Northeast Atlantic. The annotation provides the heights of various islands in the path of the plume suggesting that it extends to an altitude of at least 2 km, but probably not to 4 km. (SeaWiFS Project NASA/GSFC and ORBIMAGE)

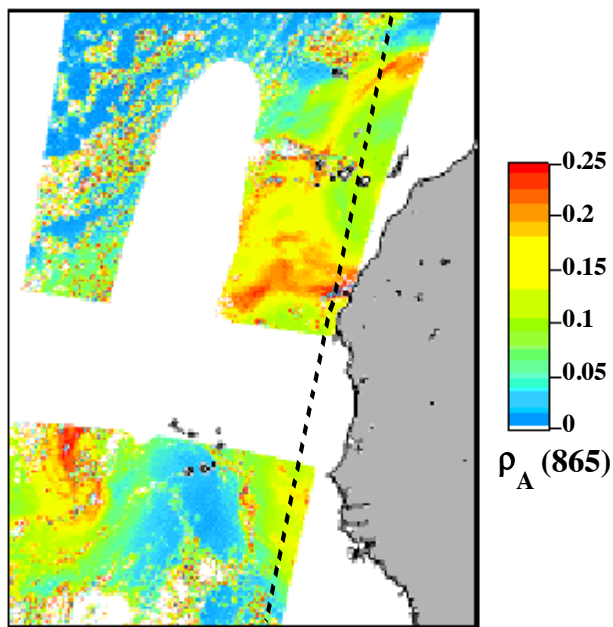


Figure 10.26: SeaWiFS image of $\rho_A(865) = \rho_a(865) + \rho_{ra}(865)$ showing a dust plume off the West Coast Africa on June 23, 1998. The vertical white area is the portion masked because of Sun glint, and the horizontal strip is the data that is missing because of the change in the sensor's tilt from north to south. From [Moulin et al. \[2001\]](#).

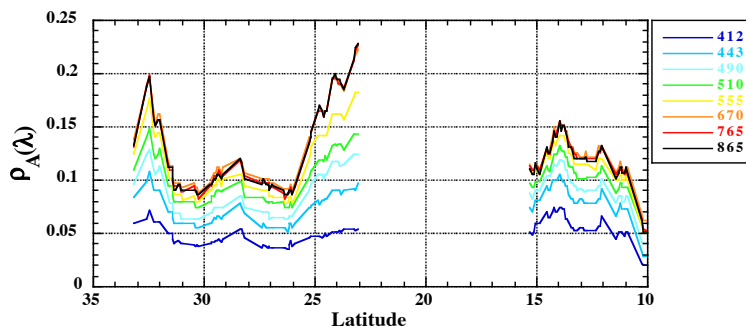


Figure 10.27: Extracted values of $\rho_A \equiv \rho_a + \rho_{ra}$ along the track line in Figure [10.26](#). Curves progress from bottom to top through 412 to 865 nm in numerical order. The quantity ρ_A is actually $\rho_A + t\rho_w$; however, here $t\rho_w \ll \rho_A$. From [Moulin et al. \[2001\]](#).

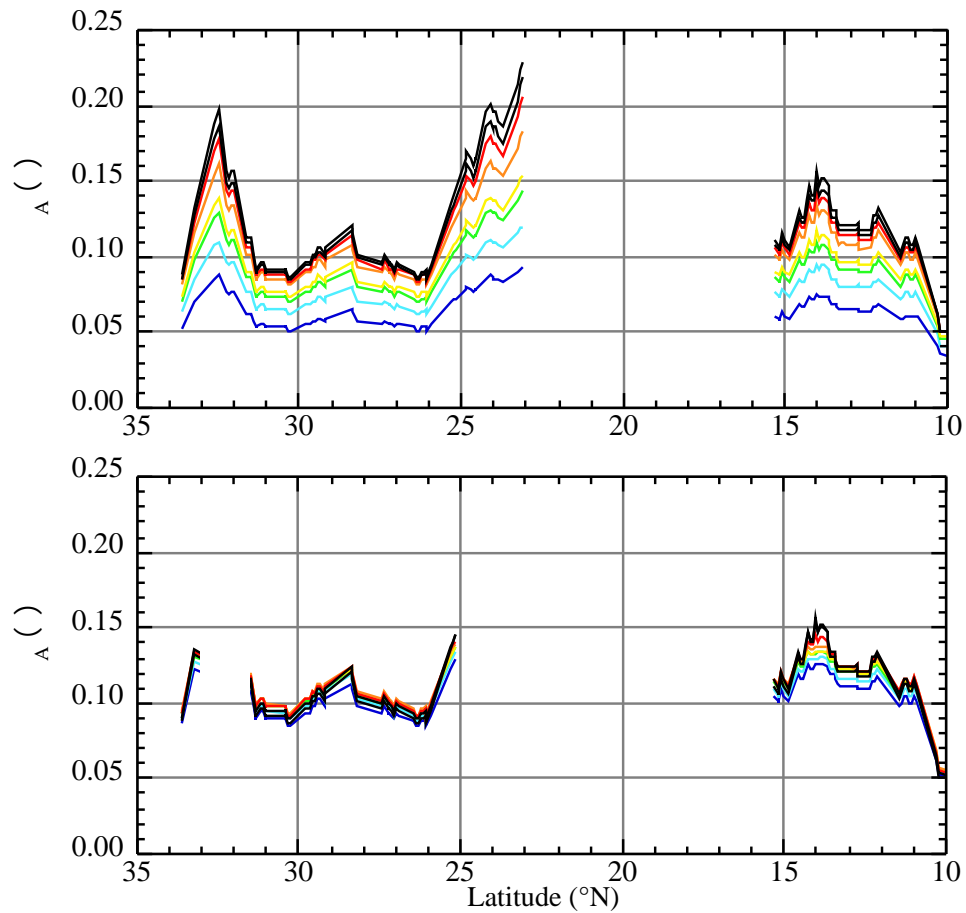


Figure 10.28: Computed ρ_A along the track line in Figure 10.26 for the BDS1 model (upper) compared to that of the M90 model (lower). As in Figure 10.27, curves progress from bottom to top through 412 to 865 nm in numerical order. From Moulin et al. [2001].

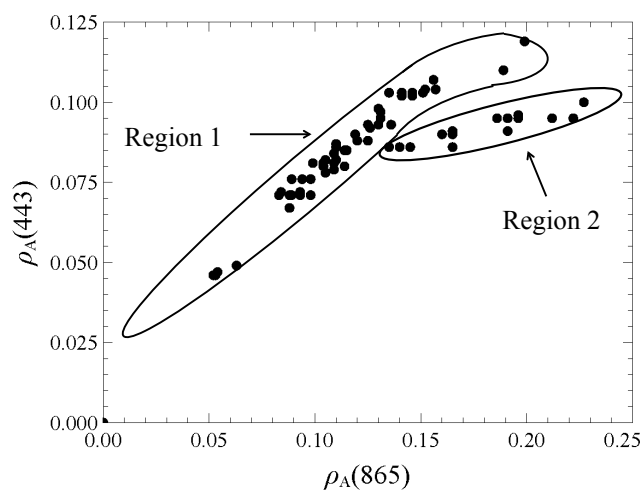


Figure 10.29: $\rho_A(443)$ as a function of $\rho_A(865)$ extracted from the track line in Figure 10.26. The enclosed points in Region 1, and Region 2 are described in the text.

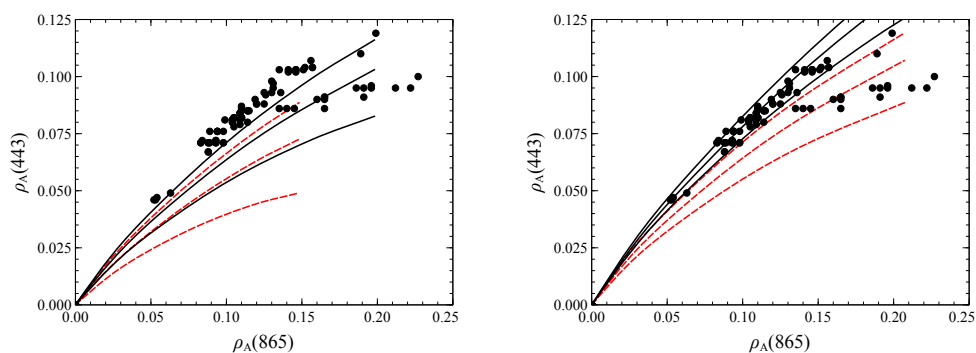


Figure 10.30: Data points: $\rho_A(443)$ as a function of $\rho_A(865)$ from Figure 10.29. Left panel: Relationships provided by the BDS1 (solid-black) and BDS3 (dashed-red) models superimposed. For each model, the upper curve is “V00,” the middle curve “V04” and the bottom curve “VUU.” Right panel: Same presentation as left panel but for BDB1 (solid-black) and BDB3 (dashed-red).

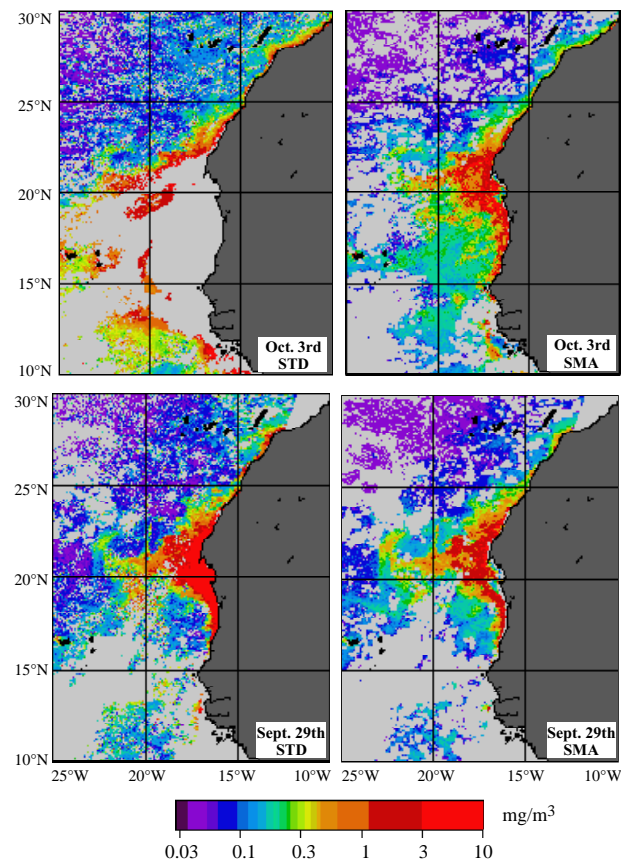


Figure 10.31: Comparison between the standard (STD, left panels) and the SMA (right panels) algorithms for processing SeaWiFS imagery. The retrieved quantity is C_P for the SMA and Chlorophyll a (C) for the STD. Typically, C_P is about 35% greater than C . Top panels are for a dusty day and bottom panels are for a clear day (very little dust). Reformatted from Moulin et al. [2001].

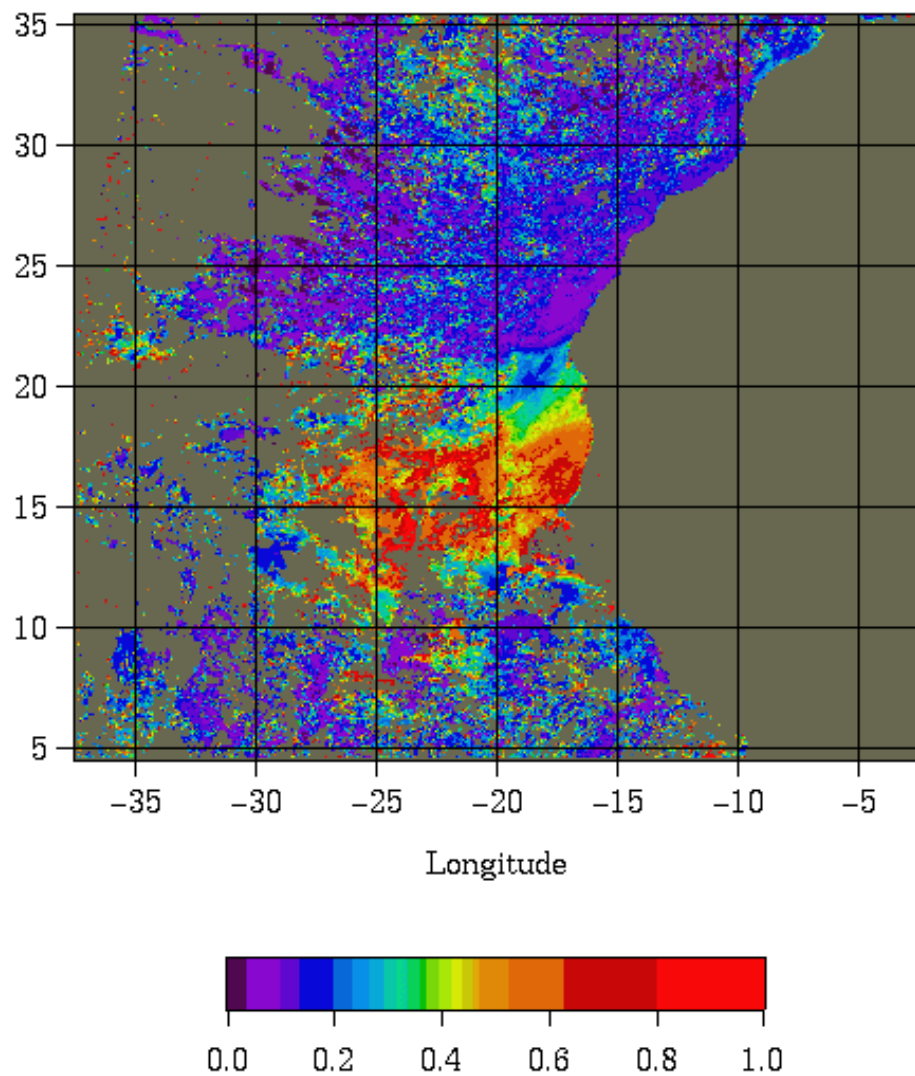
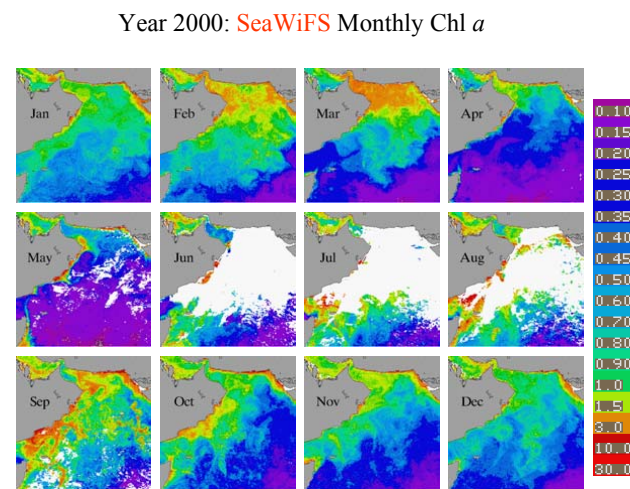
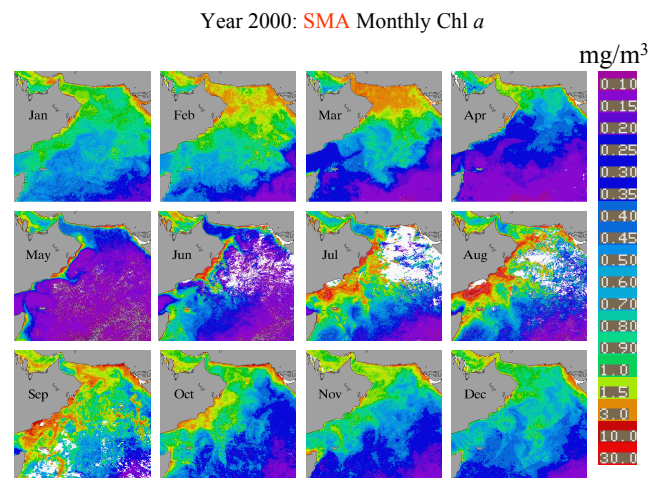


Figure 10.32: Average aerosol optical depth at 865 nm for the period Sept. 30 to Oct. 7 that includes the dust event in Figure 10.31 (top panels).



Standard processing leads to data gaps due to cloud/dust masking.

Figure 10.33: Monthly concentration of chlorophyll *a* derived from SeaWiFS imagery using the spectral matching algorithm (top) [Moulin et al., 2001] and the standard algorithm (bottom). The June, July and August panels appeared in Banzon et al. [2004].

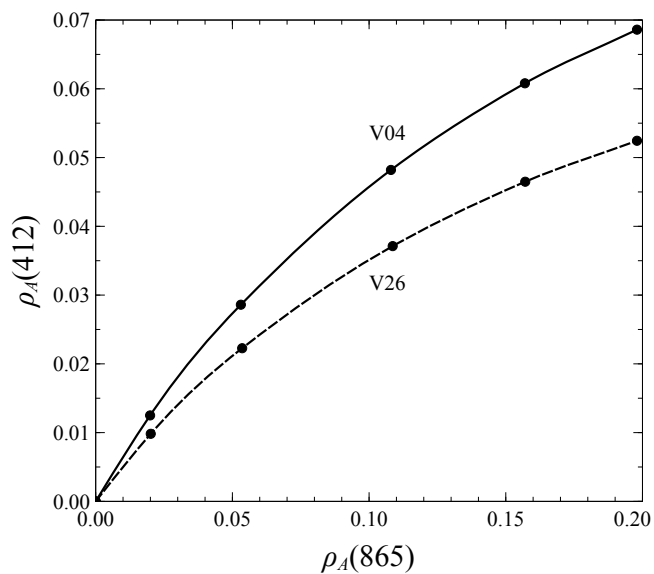


Figure 10.34: Computations of $\rho_A = \rho_a + \rho_{ra}$ at 412 nm as a function of $\rho_A(865)$ for two vertical distributions of the BDS1 model. Solid line is for the vertical distribution V04. Dashed line is for the vertical distribution V26.

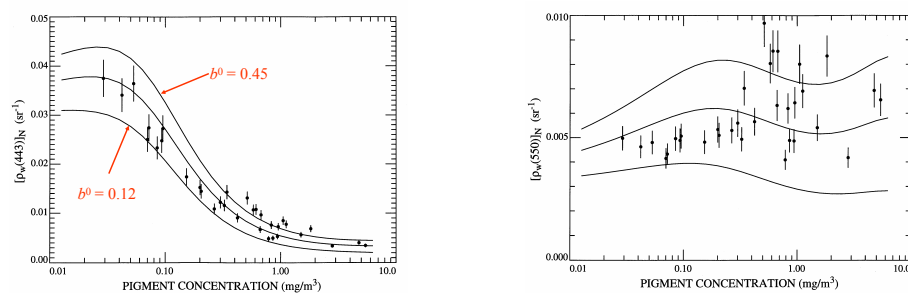


Figure 10.35: Modeled $[\rho_w(443)]_N$ (left panel) as a function of C_P compared to experimental data from Figure 10.1. The lines from bottom to top are for different values of b^0 : 0.12, 0.30, and 0.45 m^{-1} , respectively. Right panel is the same presentation but for $\lambda = 550$ nm. Redrawn from [Gordon et al. \[1988\]](#).

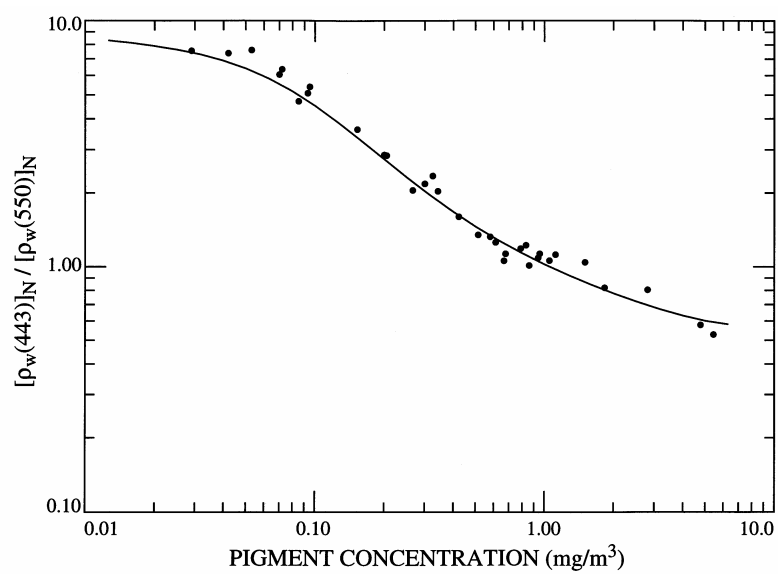


Figure 10.36: Modeled ratio $[\rho_w(443)]_N / [\rho_w(550)]_N$ as a function of C_P compared to the experimental data from Figure 10.1. The value of b^0 used to generate the line was 0.20 m^{-1} . Redrawn from Gordon et al. [1988].

Chapter 11

Basics of Ocean Color Sensors

11.1 Introduction

In this chapter we will present some of the basic issues involved in developing the instrumentation used to provide imagery of $\rho_w(\lambda)$. We will not get bogged down in the specific detail of any one instrument; however, along the way we will discuss and compare some aspects of the CZCS, SeaWiFS, MODIS, and VIIRS. In particular we will show examples of how some of the basic instrumentation problems have been solved. Because the author has extensive experience with U.S. sensors and their design, these will be focus of the discussion. We begin with consideration of some of the radiometric design questions that the scientist must pay particular attention to, e.g., the position and number of the spectral bands, the allowable noise in the various bands, and the analog-to-digital quantization of the data from each spectral band. Then we look at the imaging aspects of remote sensing, i.e., how space borne radiometers use scanning and the motion of the spacecraft to obtain a spectral image of the water surface. Finally, we examine two examples of ocean color scanners: CZCS and MODIS. In an appendix we consider some relevant aspects of satellite orbits.

11.2 Radiometric Design Considerations

The design of ocean color imagers is a collaborative effort by scientists and engineers. There are several considerations that are the responsibility of the scientific community based on the goals of the project and the manner in which the goals are to be achieved,

the latter of which is based in part on the algorithms to be used in processing data from the sensor. These considerations reflect the kind and character of the data to be provided by the sensor. Among them are the required number of spectral bands and their position in the spectrum, the acceptable level of noise in the sensor, and the quantization interval and saturation values of the measured radiances (reflectances). The responsibility of the engineer is to develop a system that meets these requirements and to establish any tradeoffs to be faced in the light of physical, time, and cost restraints. In this section we will discuss the scientific design considerations.

11.2.1 Spectral Bands

The positioning of the spectral bands is determined by the nature of the spectrum of the quantity one wishes to measure. In the case of most ocean color imagers the principal quantity one desires is the concentration of Chlorophyll *a* (or the pigment concentration C_P). Figure 11.1 shows the simulations of spectrum of water-leaving radiance (lower curves) and top-of-atmosphere radiance (upper curves) for both low (solid) and high (dotted) pigment concentrations in Case 1 waters.¹ As we have seen earlier (Chapter 6), the curves clearly show that the signature of high pigment concentration compared to low concentration is a depression of the spectrum in the blue accompanied by an enhancement in the red and very little change in the green. Superimposed over the spectrum are the spectral bands of CZCS (left panel) and SeaWiFS (right panel). We see that the spectral bands of CZCS are optimally positioned to record the most important features of the spectrum. The 750 nm band on CZCS, intended for use over land, is omitted because it was too insensitive radiometrically to be of value over the water. The fewer number of bands on CZCS compared to SeaWiFS reflects the fact that it was a proof-of-concept instrument. In the case of SeaWiFS we see a band added at 412 nm to help separate dissolved organic material from pigments, a band near 490 nm which is more sensitive to some accessory pigments (e.g., carotenoids, Chapter 5) than Chlorophyll *a*, and spectral bands at 765 and 865 nm that were selected specifically for the purpose of atmospheric correction.² The visible bands on both sensors are 20 nm wide. Although one might like spectral bands as narrow as possible, the radiant power in a spectral band is proportional to the spectral width, so the chosen width is a compromise to provide sufficient radiant power for the detectors. The

¹In the simulations, smoothed values of $F_0(\lambda)$ have been used, so the Fraunhofer lines do not appear; however, the absorption bands of H₂O and O₂ are not smoothed and are clearly visible in the figure.

²One notes that the spectral band at 765 nm overlaps the Oxygen “A” absorption band (band head at 759 nm). This adds additional variation to the signal in this band that was removable through modeling. Originally it was planned to divide this band into two bands (and sum their detected powers) to avoid the absorption, but that would have caused an unacceptable decrease in the reflected signal (radiant power) in the band combination, and the plan was abandoned.

NIR bands are 40 nm wide.³ MODIS has spectral bands similar to SeaWiFS (except 510 nm is moved to 531 nm), but with approximately half of the spectral width for each band. In addition it has a very sensitive band at 678 nm to be able to detect the Chlorophyll *a* fluorescence at 683 nm (Chapter 6).

11.2.2 Acceptable Levels of Sensor Noise

All detectors of electromagnetic radiation (and their associated electronics) display *noise*. Noise is a random fluctuation in the output of the detector, e.g., the current flowing in the detection circuit (Chapter 12). There are several sources of noise. Many are associated with the discrete or particle nature of light and matter. Light consists of a stream of photons, but this stream is not uniform. The emission of a photon from a source is a random event, so a beam of photons consists of a sequence of random events. The number of photons received by a detector in a given time is therefore not constant, but varies in time, so the output of the detector will vary in time.⁴ This variation is called *photon noise*. Photon noise depends on the radiance itself and is Poisson distributed.⁵ If the detector on the average receives \bar{N} photons per unit time, then the fluctuation (the square root of the variance) in the number received is given by $\sqrt{\bar{N}}$, i.e., if a stream of photon causes a detector to receive 100 photons per unit time on average, then the fluctuation will be about 10 photons per unit time, i.e., the number received will vary roughly between 90 and 110 during each unit time. In this case, the *signal to noise ratio* (SNR), i.e., the average signal divided by its standard deviation will be 10, while if the beam is brighter, say containing 10,000 photons per second, then the SNR will be 100, etc. The SNR is a convenient way of describing the performance of a detection system. The above example also shows that for photon noise the SNR can be adjusted simply by changing the sampling time period (also called the *integration time*), i.e., by increasing the sampling time by a factor of 100, the SNR increases by a factor of 10. It also exposes a fundamental tradeoff for detection systems for measuring weak light sources — longer integration times and lower noise versus

³The detectors used for SeaWiFS were silicon photodiodes, which are more efficient in the NIR than the visible. However, Figure 11.1 shows that the top-of-atmosphere radiance at 765 and 865 nm is an order of magnitude or more *less* than in the blue. This lower level of radiance required the larger spectral band width in the NIR compared to the visible. (Recall that the spectral radiant *power* falling on a detector is proportional to the spectral width $\Delta\lambda$.)

⁴What we normally term as *radiance* is really the measured instantaneous radiance averaged over a period of time sufficiently long to render the detector fluctuations irrelevant.

⁵For a Poisson distribution of photons, if the mean number of photons arriving during a time interval Δt is \bar{N} , then the probability of N photons arriving in Δt is

$$p(N) = \frac{\bar{N}^N}{N!} \exp[-\bar{N}].$$

It is easy to show that the variance in N is \bar{N} .

shorter integration times and higher noise.⁶

Even though we are not going to study noise in detail, to aid in the understanding of noise, we briefly describe two additional sources among others. The first is called *shot noise*. Shot noise in any electrical system is the result of the discrete nature of electric charge. Just as with photon noise, if a current (charge per unit time) consists of an average \bar{N} electrons per unit time then the variance in the number is \bar{N} , so the standard deviation is $\sqrt{\bar{N}}$. So, again, the SNR is $\sqrt{\bar{N}}$, which can be increased by increasing the unit of measurement time which increases \bar{N} . A second noise source is related to the thermal motion of electrons in any circuit element. It is called *Johnson noise*. To understand Johnson noise, consider an electrical resistor of resistance R connected to a sensitive voltmeter. The electrons in the resistor are in constant random (thermal) motion. Therefore, a fluctuating voltage will be measured by the voltmeter even in the absence of any net current through the resistor. Because thermal energy is proportional to the absolute temperature T , we expect that this noise source becomes more important as the temperature increases. This is in fact the case: the fluctuating current has an RMS value of $i_{\text{RMS}} = \sqrt{2k_B T / (\Delta t R)}$, where k_B is the *Boltzmann constant* (1.38×10^{-23} Watt second/Kelvin), and Δt is the integration time. Note that in this case the noise does not depend on the magnitude of the signal, i.e., the current flowing in the resistor, but longer integration times still lead to lower noise.

There is noise in all ocean color radiometers, and it is clear that sensors with lower noise will perform better than sensors with higher noise. However, the reduction of noise is usually accompanied by an increase in cost, so the *maximum* acceptable noise has to be specified to be below a given level. It is up to the scientist to specify what that level of performance must be to accomplish the scientific goals. It is up to the design engineer to determine how to obtain the given level of performance.

How can we specify the performance of a system as complicated as an ocean color sensor? A realistic approach is to require that the sensor be *as least as good or better* than the algorithms used to provide the desired products. We have seen that for the low- C_P algorithm⁷ using the blue (443 nm) to green (550 nm) ratio, the error in the estimate of C_P , i.e., $\Delta C_P / C_P$, is of the order of 20% (Chapter 10). For convenience we provide this algorithm from Chapter 10:

$$\log_{10} C_P = -0.04 - 2.26 \log_{10} R + 3.03(\log_{10} R)^2 - 2.80(\log_{10} R)^3, \quad (11.1)$$

with $R = [\rho_w(443)]_N / [\rho_w(550)]_N$. In addition, we found that to a good approximation

$$[\rho_w(443)]_N \approx 0.0075 C_P^{-0.48}, \quad (11.2)$$

⁶We shall see later that there is also a relationship between the integration time and the spatial resolution at the water surface. This is another example of a tradeoff to be faced in ocean color sensor design.

⁷Recall that this algorithm would be used for $0 \leq C_P \leq 0.6$ mg/m³.

and by inverting the first equation and combining it with the second we have a statistical approximation to $[\rho_w(550)]_N$ as a function of C_P . For the sensor be *as least as good or better* than this algorithm, we require that the error introduced by the sensor noise be a fraction f of $\Delta C_P/C_P$. For example, if we required that

$$\left(\frac{\Delta C_P}{C_P}\right)_{\text{Sensor}} = f \left(\frac{\Delta C_P}{C_P}\right)_{\text{Algorithm}}$$

then in a statistical sense, the total (sensor *plus* algorithm) uncertainty in the derived C_P would be

$$\begin{aligned} \left(\frac{\Delta C_P}{C_P}\right)_{\text{Total}} &= \sqrt{\left(\frac{\Delta C_P}{C_P}\right)_{\text{Sensor}}^2 + \left(\frac{\Delta C_P}{C_P}\right)_{\text{Algorithm}}^2} \\ &= \sqrt{1 + f^2} \left(\frac{\Delta C_P}{C_P}\right)_{\text{Algorithm}}. \end{aligned}$$

Thus, if $f = 1/2$ the total uncertainty is only about 22% if

$$\left(\frac{\Delta C_P}{C_P}\right)_{\text{Algorithm}} = 0.2.$$

This approach allows us to determine some requirements concerning the acceptable sensor noise. Since the manifestation of the noise is a noisy radiance, we refer to the noise as the *noise-equivalent radiance* or NER , or if working in reflectance units, the *noise-equivalent reflectance* $NE\Delta\rho$.

We will begin by deriving some estimates of the $NE\Delta\rho$ in the visible bands through an analytical approach.⁸ From Eq. (11.1) we have

$$\left(\frac{\Delta C_P}{C_P}\right)_{\text{Algorithm}} = \left\{ -2.26 + 2 \times 3.08 \log_{10} R - 3 \times 2.8 (\log_{10} R)^2 \right\} \frac{\Delta R}{R}, \quad (11.3)$$

with

$$\frac{\Delta R}{R} = \frac{\Delta[\rho_w(443)]_N}{[\rho_w(443)]_N} - \frac{\Delta[\rho_w(550)]_N}{[\rho_w(550)]_N}. \quad (11.4)$$

As we are talking about random noise at both wavelengths, we replace this by⁹

$$\frac{\Delta R}{R} = \sqrt{\left(\frac{\Delta[\rho_w(443)]_N}{[\rho_w(443)]_N}\right)^2 + \left(\frac{\Delta[\rho_w(550)]_N}{[\rho_w(550)]_N}\right)^2}. \quad (11.5)$$

⁸Later we will redo the computation using an approach in which we actually simulate the noise in each spectral band and examine the consequences.

⁹The reader should contrast the following form for $\Delta R/R$ with the situation in Chapter 10, where the same equation, Eq. (11.4), was used to determine the accuracy requirement for atmospheric correction. In that case the errors in $[\rho_w]_N$ for the two bands were perfectly correlated because of the nature of the atmospheric correction algorithm. Here, because the errors are random there is no correlation.

For low C_P , $[\rho_w(443)]_N$ is substantially larger than $[\rho_w(550)]_N$, so if the $\Delta[\rho_w]_N$'s are roughly the same, we can neglect the term for 443 nm in Eq. (11.5) compared to the 550 nm term, and use this to determine an allowable $\Delta[\rho_w(550)]_N$. For $C_P = 0.075 \text{ mg/m}^3$, $R \approx 5.15$, $[\rho_w(550)]_N \approx 0.005$, and the term inside the curly brackets in Eq. (11.3) is -2.207 . If we take $f = 0.5$ and recall that $(\Delta C_P/C_P)_{\text{Algorithm}} \approx 0.2$, we have finally

$$\Delta[\rho_w(550)]_N = 2.26 \times 10^{-4},$$

which we will refer to as $NE\Delta[\rho_w(550)]_N$. Now we estimate $\Delta[\rho_w(443)]_N$. For this we use $C_P = 0.6 \text{ mg/m}^3$ and keep *both* terms in the square root in Eq. (11.5). For this value of C_P , $[\rho_w(550)]_N \approx 7.8 \times 10^{-3}$, $[\rho_w(443)]_N \approx 9.6 \times 10^{-3}$, $R = 1.23$ and the term in the curly brackets is -1.79 . With the above result for $\Delta[\rho_w(550)]_N$, we find

$$\Delta[\rho_w(443)]_N = 4.58 \times 10^{-4},$$

which we will refer to as $NE\Delta[\rho_w(443)]_N$. Therefore, for low C_P ,

$$\frac{NE\Delta[\rho_w(443)]_N}{[\rho_w(443)]_N} = 0.0176 \quad \text{and} \quad \frac{NE\Delta[\rho_w(550)]_N}{[\rho_w(550)]_N} = 0.0453,$$

and for the higher value of C_P ,

$$\frac{NE\Delta[\rho_w(443)]_N}{[\rho_w(443)]_N} = 0.0477 \quad \text{and} \quad \frac{NE\Delta[\rho_w(550)]_N}{[\rho_w(550)]_N} = 0.0290.$$

The algorithm in Eq. (11.1) is used for $0 \leq C_P \leq 0.6 \text{ mg/m}^3$. For $C_P > 0.6 \text{ mg/m}^3$ 443 nm is replaced by 490 nm. The algorithm using 490 and 550 nm can then be used to estimate $NE\Delta[\rho_w(490)]_N$, etc.¹⁰

Let us recall for a moment the basic components of the reflectance in any spectral band:

$$\rho_t = \rho_r + \rho_A + t_v t_s [\rho_w]_N,$$

where t_v is the diffuse transmittance for the viewing angle and t_s is the diffuse transmittance for the solar zenith angle. These are to a good approximation given by (Chapter 9)

$$t_v = \exp[-\tau_r/(2.2 \times |\cos \theta_v|)] \quad \text{and} \quad t_s = \exp[-\tau_r/(2.2 \times \cos \theta_s)],$$

where θ_v and θ_s are the viewing and Sun angles, respectively. What we computed above and called $NE\Delta[\rho_w]_N$ is the maximum acceptable noise in $[\rho_w]_N$; however, what we really

¹⁰At this point it is important to understand the notation being used here and henceforth. When we write $\Delta\rho$ we are referring to an individual sample of $\Delta\rho$ from a gaussian distribution with standard deviation $NE\Delta\rho$, e.g., $N(0, NE\Delta\rho)$. The $NE\Delta\rho$ is a number which we either have already determined, as for $NE\Delta[\rho_w(443)]_N$, etc., or which we wish to determine through further analysis.

want for instrument design is the maximum acceptable noise in ρ_t , i.e., $\Delta\rho_t$. There is no “noise” in ρ_r , as it is a calculated quantity. There is noise in ρ_A that results from noise in the NIR bands propagated by the atmospheric correction procedure into the visible bands. Thus, after atmospheric correction, there are two sources of noise in the resulting $[\rho_w(\lambda)]_N$: sensor noise at λ which we refer to as *intrinsic* noise; and noise at λ resulting from noise in the NIR propagated to the visible through atmospheric correction which we call *induced* noise. Including the second form of noise at λ and noting that the induced and intrinsic noise sources are uncorrelated, we have estimated the maximum acceptable values of $\Delta[\rho_w(\lambda)]_N$, i.e.,

$$\Delta[\rho_w(\lambda)]_N^2 = (t_v t_s)^{-2} \left(\Delta\rho_A(\lambda)^2 + \Delta\rho_t(\lambda)^2 \right).$$

The left-hand-side of this is determined, so it is really up to the design engineer to decide how to divide the noise between the visible and the NIR bands; however, let’s try to understand the influence of $\Delta\rho_A(\lambda)$ on ΔC_P .

The magnitude of the induced noise in the visible, from the NIR through atmospheric correction, is dependent on the $NE\Delta\rho$ ’s in the NIR bands. This magnitude can be determined by applying the atmospheric correction algorithm with noise added to the NIR bands in a stochastic manner. Carrying out such a calculation analytically is not feasible, but it is simple to effect numerically. That is what we do next.

As with our analysis of the accuracy required for atmospheric correction, we shall use the single-scattering version of the algorithm and assume that $\rho_{as}(\lambda)$ varies exponentially with λ . In the algorithm we replace

$$\varepsilon(\lambda_s, \lambda_l) = \frac{\rho_{as}(\lambda_s)}{\rho_{as}(\lambda_l)} \quad \text{by} \quad \varepsilon'(\lambda_s, \lambda_l) = \frac{\rho_{as}(\lambda_s) + \Delta\rho_{as}(\lambda_s)}{\rho_{as}(\lambda_l) + \Delta\rho_{as}(\lambda_l)},$$

where the $\Delta\rho_{as}$ ’s are the added uncorrelated noise in the two NIR bands, and use

$$\varepsilon'(\lambda, \lambda_l) = \exp[k(\lambda_l - \lambda)],$$

with k determined from

$$\varepsilon'(\lambda_s, \lambda_l) = \exp[k(\lambda_l - \lambda_s)].$$

Then, $\rho_{as}(\lambda) = \varepsilon'(\lambda, \lambda_l)\rho_{as}(\lambda_l)$. As $[\rho_w]_N = 0$ in the NIR and there is no noise in ρ_r , the $\Delta\rho_{as}$ ’s in the NIR are in fact equal to the $NE\Delta\rho_t$ ’s there. The noise in the NIR bands is assumed to have normal distribution with zero mean and a standard deviation of $NE\Delta\rho_t$, i.e., it is an $N(0, NE\Delta\rho_t)$ distributed random variable. For simplicity, we have assumed that the $NE\Delta\rho_t$ ’s in the two NIR bands are the same.

We could sample repeatedly from $N(0, NE\Delta\rho_t)$ and compute the distribution of values induced in $\rho_{as}(\lambda)$ (the “noise”), which we call $\Delta\rho_{as}(\lambda)$, as a result of noise in the NIR bands.

If this induced noise were significantly less than the $NE\Delta[\rho_w(\lambda)]_N$, it could be ignored, otherwise it must be included in order to find the acceptable $NE\Delta\rho_t$ in the visible. But for simplicity, we will consider both intrinsic and induced noise sources simultaneously. We will use our above estimates for acceptable values of $NE\Delta[\rho_w(443)]_N$ and $NE\Delta[\rho_w(550)]_N$ (although they may be too large). We simulate the noise in the visible by repeatedly choosing $\Delta[\rho_w(\lambda)]_N$ from a normal distribution $N(0, NE\Delta[\rho_w(\lambda)]_N)$ for both wavelengths. The total computational program is

1. compute $\Delta\rho_{as}(\lambda)$ from by sampling $N(0, NE\Delta\rho_t)$ individually at both NIR wavelengths and applying the atmospheric correction algorithm
2. sample $\Delta[\rho_w(\lambda)]_N$ from $N(0, NE\Delta[\rho_w(\lambda)]_N)$ individually at both 443 and 550 nm,
3. form

$$\left(\Delta[\rho_w(\lambda)]_N\right)_{\text{Total}} = \left(\frac{\Delta\rho_{as}(\lambda)}{t_v t_s} + \Delta[\rho_w(\lambda)]_N\right)$$

individually at both wavelengths,

4. add the resulting $\left(\Delta[\rho_w(\lambda)]_N\right)_{\text{Total}}$ to the nominal $[\rho_w(\lambda)]_N$ (for the given C_P) individually for each wavelength, and finally,
5. insert the result of these steps into Eq. (11.1) and compute the resulting C_P .

The distribution of the retrieved C_P values resulting from steps (1)-(5) will provide the distribution of the C_P from which the standard deviation in C_P can be found, providing $(\Delta C_P/C_P)_{\text{Sensor}}$ resulting from both intrinsic and induced sensor noise. The results of this exercise is presented in Figure 11.2 for three levels of $NE\Delta\rho_t$ in the NIR, and the values of $NE\Delta[\rho_w]_N$ at 443 and 550 nm developed earlier. First, note that when $NE\Delta\rho_t(\text{NIR}) = 0$ we find that $(\Delta C_P/C_P)_{\text{Sensor}} \approx 0.1$ for both $C_P = 0.1$ and 0.6 mg/m^3 , as expected because that is how we arrived at the $NE\Delta[\rho_w]_N$ values that were used in the visible. Surprising perhaps is the fact that $(\Delta C_P/C_P)_{\text{Sensor}}$ actually reaches a pronounced minimum near $C_P = 0.3 \text{ mg/m}^3$. Increasing $NE\Delta\rho_t$ in the NIR to 1×10^{-4} and then to 2×10^{-4} influences $(\Delta C_P/C_P)_{\text{Sensor}}$ only near the low and high C_P limits, with little effect near $C_P = 0.5 \text{ mg/m}^3$. This suggests that to keep $(\Delta C_P/C_P)_{\text{Sensor}} \approx 0.1$ near $C_P \approx 0.1 \text{ mg/m}^3$, may require a somewhat smaller $NE\Delta[\rho_w]_N$ in the visible to allow for some noise in the NIR.¹¹ Here is where the scientist must discuss the situation with the engineer

¹¹It may be of interest to consider the influence of noise in the NIR by itself, i.e., with $NE\Delta\rho_t = 0$ in the visible. For $C_P = 0.3 \text{ mg/m}^3$, $(\Delta C_P/C_P)_{\text{Sensor}} \approx 0.014, 0.028, 0.065, 0.114$, and 0.163 for $NE\Delta\rho_t(\text{NIR}) = 1 \times 10^{-4}, 2 \times 10^{-4}, 4 \times 10^{-4}, 6 \times 10^{-4}$, and 8×10^{-4} , respectively. This suggests that $NE\Delta\rho_t(\text{NIR})$ cannot be much larger than 2×10^{-4} and still keep error in the bio-optical algorithm as the primary cause of noise in the pigment retrievals. We note here that these estimates depend only weakly on the noise-free value of k .

and decide the best approach, i.e., what to trade off — less noise in the visible or less noise in the NIR — to maintain $(\Delta C_P/C_P)_{\text{Sensor}} \lesssim 0.1$ for $0 \leq C_P \leq 0.6 \text{ mg/m}^3$. We cannot answer this question, but with the tools provided in this section, an answer can be developed. For completeness, in Figure 11.3 we provide the distribution of the derived pigment concentrations for $NE\Delta\rho_t = 1 \times 10^{-4}$. Because of the nonlinear relationship between C_P and the reflectances, the distributions are not symmetric about their maxima, but usually have a pronounced “tail” toward larger C_P ; however, the mean value of C_P (averaged over 30,000 simulations) is usually excellent.

Finally, it should be remembered that the noise is usually related to the radiance itself, e.g., photon noise and shot noise, so the $NE\Delta\rho_t$ ’s or NER ’s should be specified at the reflectance (radiance) values that are typical of those encountered in the actual operation (Figure 10.4).

11.2.3 Signal Quantization and Saturation Radiance

The magnitude of the radiance (reflectance) measured by the sensor is always transmitted to Earth in digital format, while the radiance itself is measured as an analog signal. Thus, analog to digital (A to D) conversion must be carried out by the instrument electronics. The signal from the sensor is digitized in a N -bit binary number. For example, if $N = 8$ the radiance is divided into $2^8 = 256$ levels, i.e., ranging from the lowest, 0 to the highest, 255. This was the case with the CZCS. For SeaWiFS $N = 10$ and for MODIS $N = 12$. If the conversion is linear, i.e., $DN = K \times L$, where DN is the N -bit digital number (or digital count), L is the radiance and K is a constant, then the maximum radiance that can be transmitted to the ground is $(2^N - 1)/K$. This is called the *saturation radiance*. The difference in radiance (ΔL_1) corresponding to $\Delta DN = 1$ is $\Delta L_1 = 1/K$.¹² Thus, from its minimum to its maximum, the water-leaving radiance (reflectance) will consist of

$$\Delta DN_{L_w} = (L_w^{\text{Max}} - L_w^{\text{Min}}) \times K$$

digital counts. This is the portion of the DN range that contains information about the water. So, one task is to make sure that ΔDN_{L_w} is large enough, i.e., to provide the sensitivity to change in L_w that the science requires.¹³ This seems simple; however there is a tradeoff. If we make K larger this makes the saturation radiance smaller. But, if we

¹²There is an error inherent in any analog-to-digital conversion. If we are converting a radiance, then any radiance falling between DN and $DN + 1$ are encoded with the same DN value, i.e., there is an uncertainty in the radiance equal to 1 digital count. This leads to an RMS noise in radiance (quantization noise) of $1 \text{ count}/\sqrt{12}$ or $L_1/\sqrt{12}$. The SNR of the analog-to-digital conversion is $L/(L_1/\sqrt{12})$ or $\sqrt{12} \times DN$, where DN is the digital counts for radiance L .

¹³To reiterate, we want a sufficiently large number of digital counts between L_w^{Min} and L_w^{Max} to provide the sensitivity needed by the bio-optical algorithms.

make the saturation radiance too small, then too much of the imagery will be saturated in one or more bands and have to be discarded. We can increase N , but this will have an impact on the rate of transmission of data to the ground¹⁴ which usually has a fixed upper limit. One method that has been used to overcome some aspects of this tradeoff is to use a bilinear digitation scheme. This was used in SeaWiFS, and is shown for the 412 nm band in Figure 11.4. Bilinear analog-to-digital conversion allows a wide range of radiance to be recorded, but with higher sensitivity in certain radiance ranges — here lower radiance values.¹⁵ In the case of SeaWiFS, the saturation radiance corresponded to a reflectance of about 1.1, which allowed low-sensitivity imaging of clouds and land as well as high sensitivity imaging of water bodies. In contrast, the CZCS used a single linear analog-to-digital conversion curve and, as such, had to have a much lower saturation reflectance than SeaWiFS. Table 11.1 compares the radiometric characteristics of CZCS and SeaWiFS. As one might expect, the difference in the values of $\Delta\rho/\Delta DN$ between the two is almost completely the result of the increase in the analog-to-digital conversion from 8-bits to 10-bits (factor of 4). Comparing the SeaWiFS values of $\Delta\rho/\Delta DN$ with the $NE\Delta\rho_t$'s of the previous section suggests that the $NE\Delta\rho_t$'s correspond to roughly two digital counts.¹⁶ Recalling that at 443 nm the range in $[\rho_w]_N$ is approximately 0.005 – 0.035, $\Delta\rho/\Delta DN$ for CZCS shows that this corresponds to 7 to 47 counts while for SeaWiFS it corresponds to 28 to 200 counts. Similarly, at 550 nm, where the range is 0.005 to 0.010 the count range in $[\rho_w]_N$ is 12 to 24 for CZCS and 51 to 102 for SeaWiFS. It is instructive to define what might be called the *usable* signal-to-noise ratio as $[\rho_w]_N/NE\Delta\rho_t$. For CZCS in the blue this ranges from about 3.5 to 23 at 443 nm to about 6 – 12 at 550 nm. For SeaWiFS, the corresponding usable SNR's are 14 – 100 and 25 – 50.

11.3 General Principles of Imaging with Space-Borne Radiometers

There are some principles that are common to all ocean color radiometers. They all form an image of the water surface, albeit in different ways. The size of the spot on the water surface composing one element of the image (a *pixel* for picture element) is determined by the height of the sensor and its angular field of view. Usually the size of the pixel varies during the scan. The image that is formed is done so in several spectral bands; typically four or more. Careful analysis relating the height, field of view, speed of the spacecraft in orbit is required to ensure that the pixels are contiguous, i.e., that there are no gaps in

¹⁴Other factors that influence the data rate include the number of bands, the number of pixels per scan line and the number of scan lines recorded per second. Also, typically there will be additional sensors on a satellite which will affect the data rate as well.

¹⁵Bilinear encoding also requires that two linear segments be calibrated as opposed to one.

¹⁶This implies that averaging over several samples can increase the signal-to-noise ratio.

Table 11.1: Comparison of the sensitivities and saturation reflectances of CZCS and SeaWiFS. The sensitivities($\Delta\rho/\Delta DN$) refer to the lowest electronic gain, and for SeaWiFS refer to the more sensitive portion of the analog-to-digital conversion curve in Figure 11.4. Note that the CZCS data were 8-bit digitized, while SeaWiFS data were 10-bit digitized.

	CZCS	SeaWiFS	CZCS	SeaWiFS
Bands (nm)	$\Delta\rho/\Delta DN$ ($\times 10^{-4}$)	$\Delta\rho/\Delta DN$ ($\times 10^{-4}$)	ρ_{Sat}	ρ_{Sat}
412	—	2.01	—	1.15
443	7.42	1.74	0.19	1.15
490	—	1.31	—	1.13
510	4.96	1.19	0.13	1.14
555	4.12	0.98	0.11	1.11
670	2.31	0.67	0.06	1.11
765	—	0.59	—	1.11
865	—	0.52	—	1.08

the image. We shall begin by showing how a scanning radiometer could be made using a simple Gershun tube as the radiometer. We then indicate ways that the scanning system could be modified to obtain spectral images. Finally, we show some concepts for replacing the Gershun tube by more realistic optics.

11.3.1 Obtaining an Image of the Water-Leaving Radiance

How can we obtain an image of ρ_w over a large area of the ocean's surface from a satellite-borne radiometer? We begin with the simplest kind of radiometer we know — a Gershun tube equipped with a detector and spectral filter (Chapter 2). Aimed toward the water surface, i.e., pointed in a direction $-\hat{\xi}$, this arrangement allows measurement of the total radiance L_t exiting the atmosphere in a direction $\hat{\xi}$ (Figure 11.5). Some of $L_t(\hat{\xi}, \lambda)$ was backscattered out of the water, and that is what we really want. Given the extraterrestrial solar irradiance (instantaneous) $F_0(\lambda)$ we can compute $\rho_t(\hat{\xi}, \lambda)$, from which $\rho_w(\hat{\xi}, \lambda)$ is found. If the Gershun tube in Figure 11.5 is rotated around an axis parallel to the velocity vector \vec{v} of the spacecraft, and the electrical current from the detector is sampled at a uniform rate, then for one rotation of the tube the radiance appears to originate from the water surface along a string of positions as shown in Figure 11.6. The individual spots are the “*pixels*” and their area (size) depends on the properties of the radiometer. Most ocean color sensors image a spot on the water surface that has a size of about 1 km when the radiometer is pointed in the nadir direction ($\hat{\xi}$ normal to the water surface). By size we mean the diameter of a circular spot or the length of the side of a square spot. This is

also called the *spatial resolution*, the *footprint*, or the *instantaneous field of view* (IFOV). If the desired size is 1 km, then the angular field of view (α in radians) times the height above the water surface equals the spot size. For example, if the radiometer is a Gershun tube that is 1 m long and 1 mm in diameter, then the angular field of view is $\alpha = 1 \text{ mm}/1 \text{ m} = 0.001 \text{ rad}$, and from a height of 1000 km, this gives a spot size of 1 km at nadir. An individual pixel is a fraction $\alpha/2\pi$ of one rotation of the Gershun tube, i.e., there are $2\pi/\alpha$ divisions of increment α in each revolution. If \mathcal{R} is the rotation rate in revolutions per second (rps), then one must sample the current from the detector at a rate $2\pi\mathcal{R}/\alpha$ samples per second in order for the pixels to form a contiguous line as shown in Figure 11.6. If \vec{v} is toward the north, then the line of pixels can extend from the eastern horizon to the western horizon. If the tube continues to sample after completing one revolution, then a second line of pixels is obtained. If the rotation rate and the spacecraft velocity are in just the right relationship, there will be no gap between the two sets of samples (Figure 11.7). This relationship is easily seen to be $\mathcal{R} = v_g/s$, where v_g is the ground speed of the satellite, and s is the pixel size, e.g., its diameter.

At this point we need to consider the orbital dynamics of a spacecraft. If the satellite is at a distance r from the center of the earth, and the orbit is circular, Newton's laws¹⁷ tell us that $mv^2/r = mM_E G/r^2$, where m is the mass of the satellite, M_E is the mass of Earth, and v is the spacecraft velocity. To avoid having to use M_E , we Apply Newton's law to a falling object at the surface of the Earth. In this case we have $mg = mGM_E/R_E^2$, where R_E is the radius of the Earth ($\sim 6380 \text{ km}$) and g is the acceleration due to gravity (9.81 m s^{-2}), so $GM_E = gR_E^2$. Thus, $v = R_E\sqrt{g/r} = R_E\sqrt{g/(R_E + H)}$, where H is the height of the satellite above the surface. The spot on the surface of the Earth at nadir directly beneath the satellite moves at a speed $v_g = vR_E/(R_E + H) = R_E^2\sqrt{g}/(R_E + H)^{3/2}$. The period of the orbit (τ , the time for one revolution) is $\tau = 2\pi r/v = 2\pi(R_E + H)^{3/2}/R_E\sqrt{g}$. For a sensor at $H = 1000 \text{ km}$, these give $v_g = 6.36 \text{ km/s}$ and $\tau = 6303.8 \text{ s}$ or 105.05 min . For the individual scan lines to be contiguous for such a scanner, i.e., one with $\alpha = 0.001 \text{ rad}$, we find $\mathcal{R} = 6.36 \text{ rps}$.

Summarizing, for a Gershun tube 1 m long with a diameter of 1 mm at an altitude of 1000 km ($\alpha = 10^{-3}$ radians), rotating at a rate $\mathcal{R} = 6.36 \text{ rps}$ in a plane normal to the spacecraft's velocity vector, sampling at a uniform rate of $6.36 \times 2\pi/10^{-3} = 4 \times 10^4$ samples per second will image the surface with contiguous pixels that have a size of 1 km at nadir.

¹⁷The forms of Newton's laws we use here are from elementary college physics: (1) Newton's Second Law $\vec{F} = m\vec{a}$, the relationship between the force \vec{F} acting on a object and its acceleration \vec{a} ; and (2) Newton's Law of Universal Gravitation $|\vec{F}| = m_1m_2G/r^2$, where $|\vec{F}|$ is the magnitude of the (attractive) force individual masses m_1 and m_2 separated by a distance r exert on one another. The quantity G is called the *gravitational constant* and has the value $6.674 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$. In addition, we need to use the fact from elementary physics that if a particle is moving in a circle of radius r with speed v , its acceleration is v^2/r towards the center of the circle.

The reader will notice that in the above discussion, the pixel size s was always referred to as the size at nadir. This is because the pixel size increases with the viewing angle $\theta_v^{(c)}$ (Figure 11.8). Pixels at large $\theta_v^{(c)}$ are larger than those at nadir for two reasons. First, the pixels are farther from the sensor ($D > H$) so for a constant α they must be larger. Second, in the along-scan direction, the surface is inclined to the sensor by an angle $\theta_v^{(E)}$, which elongates the pixel in the along-scan direction compared to the along-track direction. Let's see how this works. Referring to Figure 11.8, the pixel size s at nadir is αH , while at a viewing angle of $\theta_v^{(c)}$, in the along-track direction the size is $s_{\perp} = \alpha D$ and in the along-scan direction it is $s_{\parallel} = \alpha D / \cos \theta_v^{(E)} = s_{\perp} / \cos \theta_v^{(E)}$, where “ \perp ” and “ \parallel ” mean “perpendicular” and “parallel” to the scan direction, respectively. Thus, we need to find D in the figure. This is easy to do by applying the law of sines to various sides and angles of the triangle SCP.

$$\begin{aligned} \sin(\pi - \theta_v^{(E)}) &= \sin \theta_v^{(E)} = \frac{R_E + H}{R_E} \sin \theta_v^{(c)} && \text{gives } \theta_v^{(E)}, \\ \theta &= \theta_v^{(E)} - \theta_v^{(c)} && \text{gives } \theta, \\ \sin \theta &= \frac{D}{R_E} \sin \theta_v^{(c)} && \text{gives } D. \end{aligned}$$

Figure 11.9 provides the ratio of s at $\theta_v^{(c)}$ to s at nadir for both the along-scan and along-track directions for a sensor at an altitude of 1000 km. Note that for such a sensor, the horizon is at $\sin \theta_v^{(E)} = 1$ or $\sin \theta_v^{(c)} = R_E / (R_E + H)$, which for $H = 1000$ km, gives $\theta_v^{(c)} = 59.83^\circ$. Also, the distance along the surface from nadir to the pixel is θR_E . Clearly, pixels at large scan angles are elongated in the scan direction and can be very much larger than pixels near nadir. This is shown schematically in Figure 11.10. When more scans are combined as in Figure 11.7, the pixels from adjacent scan lines will overlap as one proceeds from nadir to the edge, even if they do not overlap at nadir.¹⁸

Thus far, we have discussed the image formation as if the instrument records the reflectance of a pixel instantaneously and then moves on to the location of the next pixel and does the same, etc. However, the sampling is not instantaneous. The scan of an imager is smooth and continuous. As the scan progresses, the IFOV moves along the water surface and the current in the detector continuously responds to the radiance (reflectance) of the scene. How then are the individual pixels formed? They are formed by averaging the current over a specific length of time Δt_s — usually called the *integration* time or *sampling* time. We need to know how the sampling time affects the image formation; in particular, the spatial resolution. An analysis follows in which we discuss the imaging using an actual target — a bar chart.

¹⁸It is important to realize that for scan lines that are 1000-2000 km long on the surface, the values of $\theta_v^{(c)}$, ϕ_v , θ_0 and ϕ_0 are *different* for each pixel. This must be remembered when applying the various processing algorithms, e.g., atmospheric correction or normalizing the water-leaving reflectance.

Let the IFOV of the sensor have size s at nadir. At nadir the observation point at the surface moves at a speed $2\pi\mathcal{R}H$ radians/s, so in time $t_s = s/(2\pi\mathcal{R}H)$ the IFOV on the surface will have moved a distance equal to its size s . For the hypothetical sensor discussed above, with $H = 1000$ km and an angular field of view of 10^{-3} radians, for which $s = 1$ km, we find $t_s = 25 \mu\text{s}$.

Let's try to understand how varying the sample time relative to t_s affects the spatial resolution. Spatial resolution of an optical system is most frequently described by the contrast in the image when the object is a target as shown in Figure 11.11.¹⁹ The bar pattern in this target has a spatial period Λ or a spatial frequency $f_\Lambda = 1/\Lambda$. The scan direction is perpendicular to the bars. Assume the pattern has a reflectance of unity in the white areas and zero in the black areas. We then define the modulation M as

$$M = \frac{\rho_{\text{Max}} - \rho_{\text{Min}}}{\rho_{\text{Max}} + \rho_{\text{Min}}},$$

where ρ is the measured reflectance. For the bar target itself in Figure 11.11, $\rho_{\text{Min}} = 0$, so $M = 1$. Also, for the purposes of this discussion, we assume $\rho_{\text{Max}} = 1$. Below the bar graph on the left of the figure are IFOVs of various size measured in units of Λ (here, only the horizontal size of the IFOV is relevant, the height of the IFOV is not). Assume that the detection system measures the average reflectance of the portion of the scene that is within the IFOV, i.e., if the IFOV completely encompasses one white and one black area the measured reflectance is $1/2$, etc. To the right of each IFOV we have the reflectance (the red line) that would be measured as a function of the position of the *center* of the IFOV as it moves across the bar graph. For example, for the $\Lambda/2$ IFOV, when it is in the first position shown after the arrow under the target it will measure a reflectance of unity. If it is then moved a distance $\Lambda/4$ it will record $\rho = 1/2$, and when it has moved another $\Lambda/4$ it will record $\rho = 0$. On examination of the performance (red lines) of the four IFOVs in faithfully imaging the surface, there are several observations that can be made. First, the IFOV of size Λ records $\rho = 1/2$ everywhere in the scene and sees no variation whatsoever ($M = 0$). Second, smaller sized IFOVs will record the existence of a pattern, and the accuracy of the recording will increase as the size of the IFOV decreases. When the size is reduced to $\Lambda/2$ the modulation reaches $M = 1$. For IFOVs between Λ and $\Lambda/2$, the modulation follows $0 \leq M \leq 1$. Further reduction of the IFOV below $\Lambda/2$ provides a measured reflectance that more and more faithfully reproduces the reflectance distribution of the bar graph. Finally, for IFOVs larger than Λ , e.g., the $3\Lambda/2$ IFOV, displays variation with the correct periodicity and $M = 1/3$; however, it is 180° out of phase with the bar graph's variation. This phase shift is indicated by writing $M = -1/3$

¹⁹More generally, the target might be a sinusoidal function, e.g., $(1/2) + (1/2)\sin 2\pi f_\Lambda$. Then, the full power of Fourier analysis can be applied directly to the operation of the optical system. This is a particularly useful when diffraction affects are important.

rather than $M = 1/3$.²⁰ For larger IFOVs the situation should be clear: for 2Λ , $M = 0$, for $5\Lambda/2$, $M = 1/5$ and there is no phase shift, etc. Thus, for a faithful representation of a target with periodic variation Λ , we need an IFOV size that is less than (or equal to) $\Lambda/2$.

But wait, what is the effect of the integration time? Consider the case where the IFOV is $\Lambda/2$. This has $M = 1$ and a surface reflectance ρ given by the second curve from the bottom in Figure 11.11. Assume that $\Delta t_s = t_s$, i.e., we integrate to find the average reflectance as the scan moves a distance of one IFOV. A little thought shows that the result is identical to the third curve from the bottom in Figure 11.11, i.e., ρ is always $1/2$ and $M = 0$, and the pattern of periodicity of Λ is not detected. In contrast, let $\Delta t_s = t_s/2$ and it is easy to show that $M = 1/2$ and the pattern is similar to the bottom curve in Figure 11.11, but $\rho_{\text{Max}} = 3/4$ and $\rho_{\text{Min}} = 1/4$. Thus, if $\Delta t_s = t_s$, only variations with $\Lambda > 2s$ can be detected. For our hypothetical sensor, a sampling time of $25 \mu\text{s}$ would allow detection of periodic patterns with $\Lambda > 2 \text{ km}$. If $\Delta t_s < t_s$ there will be some overlap of pixels but the reflectance variations will be delineated with more contrast (higher M). Clearly, it is desirable to make $\Delta t_s/t_s$ as small as feasible; however, one must keep in mind that the $NE\Delta\rho_t$ will generally *increase* as Δt_s decreases — another engineering tradeoff.

A final note before moving on concerns the need to maneuver the spacecraft in orbit to produce a faithful image of the surface. Consider Figure 11.12, which shows the sensor viewing the nadir point at two different times in the orbit. In order to do this, the spacecraft had to rotate through an angle θ . Thus the spacecraft must rotate 360° per orbit to maintain the correct attitude to be able to view the surface at the nadir.

11.3.2 Producing a Spectrum of the Water-Leaving Radiance

The next question regarding sensors is that of obtaining a spectrum at each pixel, i.e., imaging the water surface in different spectral bands. One way to accomplish this is to have several rotating Gershun tubes as shown in Figure 11.13. If each tube is equipped with a separate spectral filter, then for one rotation of the array of tubes an image of the surface similar to Figure 11.14 is obtained. The four separate lines corresponding to different spectral bands can be combined to one line with four spectral bands through simple processing.

An alternative for achieving two spectral bands is to use a grating spectrometer in place of the multiple tubes. This is shown schematically in Figure 11.15. The detector at the end of the Gershun tube is replaced by a small aperture allowing the light to enter the

²⁰In optics the modulation is actually represented by a complex number. In the situation here, $M = (1/3)\exp[i\pi]$ corresponding to a phase shift of π .

spectrometer chamber and fall on a concave diffraction grating G. The grating separates the light into a spectrum as well as imaging the aperture of the Gershun tube as shown. Separate portions the spectrum are imaged onto detectors D₁ and D₂. In this diagram, the size of the detectors would determine the spectral bandwidths of the two bands. This concept was used by the first ocean color scanner — CZCS.

A second alternative for achieving spectral separation without multiple Gershun tubes is shown in Figure 11.16. As in Figure 11.15 the detector is replaced by a small aperture, which now is followed by a lens L₁ to render the light parallel. The radiation then encounters a dichroic beamsplitter “Di”. This optical element has the property that it transmits the red and reflects the blue. An example of such a dichroic filter is presented in the insert on the figure. In the example, radiation with wavelength between about 350 and 550 nm is reflected with close to 100% efficiency and likewise for $550 < \lambda < 825$ nm the radiation is transmitted with close to 100% efficiency. The spectrum is further isolated with narrow-band filters F₁ and F₂. (The element M is just a plane mirror to bend the beam through 90° and make a nicer drawing.) By using three dichroics and eight filters SeaWiFS used an arrangement similar to this to obtain eight narrow spectral bands covering the range 412 to 865 nm, and with three dichroics and thirty-six filters, MODIS obtained thirty-six narrow spectral bands covering the range 412 nm to 14.2 μ m.

11.3.3 Some Scan Mechanisms

Scanning with a 1 meter-long Gershun tube is obviously somewhat unwieldy. How is scanning accomplished with actual instruments? Figure 11.17 shows a schematic of the scanning mechanism used by the Coastal Zone Color Scanner (CZCS). In the figure, a plane mirror is mounted on an axis that is parallel to the velocity of the spacecraft. The mirror is inclined at an angle of 45° to the axis. In the position shown the mirror is reflecting light incident from the nadir direction and directing it to a collimating lens. A spot on the water surface (the “Object”) is imaged in the focal plane of the lens (the “Image”). The axis is then rotated causing the mirror angle with respect to the nadir to change imaging another spot on the water surface.²¹ Thus as the mirror rotates a line of spots similar to those in Figure 11.6. In the CZCS, this optical system replaces the Gershun tube in Figure 11.15, with a small aperture determining the IFOV placed at the “Image” position.

A second scanning mechanism is used by the Moderate Resolution Imaging Spectro-

²¹For optical situations such as this, the author believes that it is often easier to visualize what is happening by imagining the light to go through the system backwards. That is, in the position marked “Image,” imagine a small bright light source. The optics then images this source as a bright spot on the water surface. As the mirror rotates the spot moves across the water surface, in much the same way that a searchlight beam swings around a lighthouse.

diometer (MODIS), with an equivalent, but more complex mechanism used in SeaWiFS. A schematic of the MODIS scanning system is presented in Figure 11.18. The optical system as shown is identical to that in Figure 11.17, except now the mirror is rotated around a horizontal (parallel to the Earth surface at nadir) axis that is *into* the paper, but is still *parallel to the velocity vector* of the spacecraft. As the mirror rotates, the second “Object” to be imaged is just to the right of the first. Again, a line of pixels is imaged on the water surface as the mirror rotates. As with CZCS, in MODIS, this optical system replaces the Gershun tube in Figure 11.16, with a small aperture placed at the “Image” position.

What are the differences between these two scanning mechanisms? First, the image rotates in Figure 11.17 as the scan takes place. To explain this, note that because of the symmetry of the system, the scan is the same as if the whole drawing were rotated about the dashed axis. Thus as the mirror rotates, viewing objects in front of and behind the paper, the image will rotate with the mirror, i.e., the image arrow will always lie in a plane formed by the dashed axis and the normal to the mirror surface. That is, the object arrows all point in the same direction on the water surface, but the image arrows rotate around the rotational axis. In contrast, in Figure 11.18 the image is fixed (i.e., does not rotate) as the scan takes place. In that figure the image arrow always points in the same direction. Because of the optical design of some detection systems, image rotation with scan angle must be avoided. This is true for MODIS, but for CZCS image rotation was irrelevant. A second difference is that the angle of incidence of the light on the scan mirror is constant in Figure 11.17 but varies in Figure 11.18. This variation of incidence angle with scan angle is one of the additional properties of a MODIS-like scanner that must be *characterized* (Chapter 12) prior to launch.

Finally, with the development of systems in which a large number of small detectors can be placed in a linear array, a new kind of scanning with no moving parts becomes possible. The concept of this kind of scanner – a “push broom” scanner – is shown in Figure 11.19, where the array has only five detectors. Here the array of detectors is aligned parallel to the Earth’s surface, but perpendicular to the velocity vector. The footprints (large squares at the bottom) are imaged by the optics (lens) onto the array (small squares at the top). As the spacecraft moves a new line of pixels come into view. One could also place a dispersing element (diffraction grating or prism) after the lens and use a rectangular array. In this manner one could arrange to have the spectrum displayed in the along-track direction with the cross-track direction providing the spatial information. This is the scheme used in many *hyperspectral* remote sensing systems. This concept is shown schematically in Figure 11.20. In this case a diffraction grating has been placed behind the lens to disperse the spectrum, and a rectangular array rather than a linear array is used. However, now the aperture of the instrument must be a line one pixel wide, perpendicular to the flight line otherwise the next scan line and the spectrum would be intermixed. A disadvantage of such a system is that the (large number of) individual detectors in the arrays must *all* be

radiometrically calibrated.

11.4 Examples of Ocean Color Scanners

Now we will examine in some detail two ocean color instruments: the CZCS; and the MODIS. Rather than discussing sensors in precise detail, our purpose will be just to show how some of the ideas in the last section are actually put into practice. We will also examine some of the complications resulting from the design.

11.4.1 The CZCS

The actual optical arrangement of the CZCS is shown in Figure 11.21. In this drawing the scan mirror is labeled M-1. Its size was roughly elliptical with dimensions $20\text{ cm} \times 28\text{ cm}$ (Figure 11.22). It rotated at a rate of 8.08 revolutions per second on an axis parallel to the velocity vector of the spacecraft. In the position shown in the diagram it is oriented to reflect light from the nadir. M-1 directs the light into a Cassegrain telescope consisting of mirrors M-2 and M-3. This telescope functions in the same manner as the lens in Figure 11.17. The light then proceeds through a dichroic beam splitter that separates the visible from the far infrared (here $\sim 10\text{ }\mu\text{m}$). The far infrared is reflected off the dichroic and then reflects from a small mirror M-4 into its detection optics. The visible and NIR pass through the dichroic and a field stop that determines the size of an individual pixel. This beam is then rendered parallel by mirrors M-6 and M7 after which it passes through the *depolarizer* to effectively render the light unpolarized.²² Mirrors M-8 and M-9 fold the beam and direct it to a concave grating in a *Wadsworth* mounting as shown schematically in Figure 11.15. These mirrors are configured in such a manner as to introduce no additional polarization into the beam. The light is then dispersed into a spectrum with the position and apertures of the individual detectors defining the spectral band detected. The “Channels” 1 through 4 constitute the visible bands (443, 520, 550, 670 nm, respectively) while 5 is the NIR (700-800 nm). The visible bands were 20 nm wide, e.g., Band 1 ranged from 433 to 453 nm. Bands 1-4 were designed to have sensitivity sufficient to record the radiance levels over the ocean, while Band 5 had a reduced sensitivity for the purpose of imaging the land surface and was not intended to be used for ocean observations. The output signal from

²²The incident radiance is partially polarized and is further polarized upon reflection from the scan mirror. Because the diffraction grating has a high polarization sensitivity (10's of percent) it is necessary to depolarize the light. In reality the light is not “depolarized,” rather its polarization is spatially scrambled, i.e. different parts of the beam across its spatial profile emerge from the depolarizer with different polarizations, with the net effect that if a measurement of the polarization of the beam as a whole were carried out, the conclusion would be that it is unpolarized.

each band was converted from analog to digital in 8-bits (256 digital steps from zero to the maximum radiance expected over the *ocean* in each band) transmitted to the ground (or recorded). The pixel size (IFOV on the surface, or spatial resolution) was 0.825 km at nadir. The satellite altitude was 955 km. The maximum scan angle $\theta_v^{(c)}$ in Figure 11.8 was 39.2° .

One should note that the mirror M-7 has a small hole in its center. This allowed light from a calibrated incandescent lamp to enter the aft optics. This could be used to monitor the stability of the various elements following M-7, but not the fore optics, i.e., the five optical elements in the visible path before M-7. Actually the calibration system was hardly used as the primary degradation seen in the sensitivity of the overall instrument with time in orbit was due to the degradation of the fore optics; probably the scan mirror — the calibration assembly *did not* record any significant degradation in sensitivity with time.

11.4.2 The MODIS

MODIS differs from CZCS in that it uses narrow-band filters to separate the spectrum into bands, and rather than placing a single detector at the focus of the system (i.e., an individual pixel as in Figure 11.16) each of four *focal planes* contained a number of detectors for each spectral band. The visible focal plane is shown schematically in Figure 11.23. In this figure, the scan direction is horizontal (perpendicular to the line of detectors) and the direction of the velocity vector of the satellite is vertical (parallel to the line of detectors). The band numbers are along the bottom. Bands 8, 9, 10, 11, and 12 are visible bands for ocean observations with a resolution of 1 km at nadir. The resolution is determined by the physical size of the individual detectors. Bands 3 and 4 are high resolution (0.5 km) visible bands for land observations. Each ocean band has ten individual detectors, labeled 1 through 10, covered by a narrow-band spectral filter. When the scan mirror is in a position to image at nadir, an approximately 21 km \times 11 km image of the water surface falls on this focal plane. In Figure 11.24 we have superimposed a simulated image of the water surface on the focal plane. This is what actually happens in the operation of the sensor. As the scan mirror rotates the image will move across the focal plane from left to right and is sampled by the individual detectors.²³ Thus a single scan of the scan mirror provides ten lines of pixels for each spectral band, e.g., something like Figure 11.7 but with

²³Recall that with a CZCS type scan mechanism, the image rotates with the scan. If such a scan mechanism were used for MODIS, the image shown on the focal plane in Figure 11.24 would rotate with scan angle, i.e., if the direction of North is vertical toward the top of the image at nadir, it will be rotated throughout the scan and cut diagonally across the focal plane at large scan angles. This would clearly result in some confusion and require geometrical correction even to form an image of the surface. To avoid such confusion the MODIS-type scanning mechanism (Figure 11.18), for which there is no image rotation, was used.

ten lines instead of four. The pixels for different wavelength are of course interleaved and a given pixel is never simultaneously viewed by different wavelengths. Rather, there is a small time (or spatial) delay between the various spectral bands; however, it is a simple matter to properly register all of the bands. We note that if the scan mirror has two sides as MODIS's does, then one complete rotation of the mirror will produce twenty scan lines so it need not rotate as fast as the CZCS scan mirror (the actual rotation rate is 0.338 revolutions per second²⁴). The detector outputs for MODIS were digitized in 12-bits (4096 discrete levels).

Table 11.2: Comparison of some of the properties of the various U.S. Ocean Color Scanners. H is the altitude of the satellite orbit, v_g is the ground speed of the sub-satellite point, τ is the period of the orbit, Δx is the spatial resolution, and "Quant." is the number of bits used in the digital quantization of the radiance. The spectral bands are listed in each column and the band width is in parenthesis.

Quantity	CZCS	SeaWiFS	MODIS	VIIRS
H (km)	955	705	705	829
v_g (km/s)	6.42	6.75	6.75	6.59
τ (min)	104.14	98.92	98.92	101.44
Δx (km)	0.825	1.1	1.0	0.75
Quant. (bits)	8	10	12	12
Bands (nm)				
	—	412 (20)	412 (20)	412 (20)
	443 (20)	443 (20)	443 (10)	445 (20)
	—	490 (20)	488 (10)	488 (20)
	520 (20)	510 (20)	531 (10)	—
	550 (20)	555 (20)	547 (10)	555 (20)
	670 (20)	670 (20)	667 (10)	672 (20)
	—	—	678 (10)	—
	—	765 (40)	748 (10)	746 (15)
	—	865 (40)	869 (15)	865 (39)
	—	—	1375 (30)	1378 (15)

The principal drawback to a system such a MODIS is that each of the detectors in one spectral band will have slightly different responses to radiant power and must be separately calibrated. This poses a particular difficulty when viewing the moon as each pixel views a different portion and some do not see the moon at all. MODIS uses an on-board solar diffuser to calibrate the radiance into reflectance directly, with a separate optical system designed to monitor the stability of the solar diffuser. Another complication is due to

²⁴MODIS is in a lower orbit than CZCS, 705 km compared to 955 km and has lower spatial resolution, 1 km compared to 0.825 km, so the rotation rate of its mirror is not twenty times slower than CZCS, but a little less.

the variation of the footprint²⁵ with $\theta_v^{(c)}$, e.g., Figures 11.9 and 11.10. As the scan angle increases, the pixels will grow and overlap. Thus, one-half a rotation of the scan mirror produces something like that shown in Figure 11.25. This is referred to as the “bowtie” effect.

For completeness, in Table 11.2 we provide some of the characteristics of the four U.S. ocean color instruments: CZCS, SeaWiFS, MODIS, and VIIRS. In terms of $NE\Delta\rho$, SeaWiFS’s were about half that of CZCS, and MODIS’s were 2 – 3 times lower than SeaWiFS. Note that MODIS and VIIRS have a band listed at 1375 nm. As seen in Chapter 4, this is coincident with a strong H₂O absorption band in the atmosphere. The surface can be seen at the wavelength only under extraordinarily dry conditions (rare over oceans). Thus, any reflectance in this band is usually due to micron-sized particles in the atmosphere *above* the water vapor layer. In particular, thin cirrus clouds are clearly visible in this band, so it can be used to screen or mask imagery contaminated by thin cirrus clouds.

11.5 Summary

In this chapter we have tried to acquaint the reader with various performance requirements and design issues that must be faced with ocean color sensors. Although we did not discuss all of the issues (others come up with each new sensor), our discussion should be a solid starting point. The radiometric requirements in Section 11.2 are of prime importance. The methods we used to develop radiometric performance requirements, based on the bio-optical and atmospheric correction algorithms to be used to process the data, should carry over to any proposed sensor. Section 11.3, about the various ways of forming a 2-d image of the water surface, should convince the reader that there are many ways to accomplish the goal.

11.6 Appendix : Comments Regarding Satellite Orbits

The sensors are on satellites that orbit the Earth. There are two orbit types that have been used for these: polar and geosynchronous. Environmental satellites, e.g., Landsat, EOS, etc., are usually in polar orbits, while meteorological satellites such as GOES, Meteosat, etc., are usually in geosynchronous orbits. Each orbit has its special attributes (positive and negative), e.g., the sun-synchronous orbit allows the sensor to view the Earth in direct

²⁵Again, think of each of the ten detectors as being a source and form its image on the water surface. This gives the “footprint” as a function of scan angle.

Sunlight half of every orbit. In the geosynchronous orbit, the pixel at nadir can be viewed continuously, but it is illuminated by the Sun only about half the time. In sun-synchronous orbit the sensor can view a given location on the Earth at most once per day, but can view almost every point on the Earth once per day or once every two days depending on the latitude. However, as the sun-synchronous orbit is “locked” to the Sun and therefore to the solar-induced tide, interpretation of the data is more difficult in coastal regions. In contrast, a sensor in geosynchronous orbit it can continuously “stare” at a single location and acquire data as long as it is in the sunlight (and resolve the tide affects in coastal regions), but such a sensor can “see” at most only half of the Earth — the other half can never be viewed. Here, we discuss the properties and special physical requirements of these orbits.

11.6.1 Polar Orbit

Most environmental satellites for Earth observation are in what are called *polar orbits*. However, the orbits do not go over the poles, i.e., neither the North or the South Poles are ever at nadir as seen from the satellite. Consider Figure 11.26 (left panel), which shows the motion of the Earth progressing around the Sun through the seasons. The thick black line represents the plane of the orbit of the satellite in polar orbit around the Earth. If the Earth were a perfect sphere, then the plane of the orbit would never change relative to the “fixed” stars, i.e., the thick line would always be oriented in the same direction, say horizontally in the figure. Then, in Spring and Fall the satellite would be over the Solar-illuminated side of the Earth half of the time, but in Summer and Winter it would spend all of its time with the Earth’s terminator in its view, i.e., it would always be viewing some of the *un-illuminated* surface. Thus, to maximize the amount of the illuminated side of the planet that is seen the orbit must precess as shown in the figure, i.e., it must turn through an angle of 360° in 365.25 days. How can this be accomplished? Actually, the shape of the Earth accomplishes the task. The equatorial radius of the Earth is greater than the polar radius as shown in Figure 11.26 (right panel). If the satellite were at the far left portion of its orbit, it would be more attracted to the point “W” than the point “N,” so the gravitational force on the satellite is *not* toward the center of the Earth. This results in a torque that causes the orbit to precess much like the precession of a gyroscope with one fixed point, e.g. a child’s “top.” The rate of precession of the point “P,” where the satellite crosses the Equator (called the node), toward “W” in degrees per orbit is given by

$$\Delta\Omega = -360^\circ \frac{J}{a^2} \frac{\cos\theta_i}{(1 - \epsilon^2)},$$

where ϵ is the eccentricity of the orbit (zero for a circular orbit), a is the semi-major axis of the orbit divided by R_E , and

$$J = \frac{2}{3} \left(\frac{I_z - I}{MR_E^2} \right) = 1.624 \times 10^{-3},$$

and I and I_z are the moments of inertia of the Earth about an axis lying in the Equatorial plane and an axis perpendicular to the Equatorial plane through the poles, respectively. As most “polar orbiting” satellites are in nearly circular orbits, $\epsilon \approx 1$ and $a \approx (R_E + H)/R_E$. For a sensor with $H = 1000$ km, the orbital period is 105.05 min, so there are 13.708 orbits per day. In order that the orbit precesses 360° in 365.25 days or 0.9856° per day, we need $\theta_i = 99.47^\circ$. The CZCS was in an orbit with $H = 955$ km with $\theta_i = 99.28^\circ$.

11.6.2 Geosynchronous Orbit

One ocean color scanner, the South Korean Geostationary Ocean Color Imager (GOCI), is in a geosynchronous orbit. This orbit circles the Earth over the Equator, but always remains over the same point on the Equator, i.e., the sub-satellite point is stationary relative to the Earth. Clearly for this to happen, the period of the orbit must be one day. Recall that the period of the orbit (τ the time for one revolution) is $\tau = 2\pi(R_E + H)^{3/2}/R_E\sqrt{g}$. For this period to be one day requires $H = 35,879$ km or 22,294 miles. Because this altitude is much larger than that for imagers in polar orbit, the angular field of view α must be much smaller. For such an imager to obtain a 1 km footprint, its angular field of view must be about 35 times smaller than a sensor at 1000 km,²⁶ with a reduction of approximately a factor of $35^2 = 1225$ in radiant power. Thus, longer integration times are required for such a sensor compared to those in polar orbits, in order to achieve the same spatial resolution and SNR. Note that a sensor in geosynchronous orbit will view the fully illuminated Earth only at one time during the day, i.e., Earth is viewed at “full phase” once per day near local noon at the sub-satellite point (ignoring the progression of seasons). The rest of the time it will be seen at partial phase or even in total darkness (local midnight).

11.7 Bibliographic Notes

11.1 Introduction

Specifications and requirements, as well as their evolution, for ocean color sensors now under development are discussed [McClain et al. \[2014\]](#).

²⁶GOCI is similar to MODIS, but with a footprint of 500 m so its α is about 70 times smaller than the MODIS angular field of view.

11.2.1 Spectral Bands

Figure 11.1 was originally prepared by the author and has appeared in several reports, e.g., [Hooker et al. \[1992\]](#).

11.2.2 Acceptable Levels of Sensor Noise

A good discussion of the various noise sources in optical sensors can be found in [Dereniak and Crowe \[1984\]](#). Estimation of the *NER*'s required for an ocean color scanner was first carried out in [Gordon \[1988\]](#) and later using stochastic methods in [Gordon \[1990\]](#).

11.2.3 Signal Quantization and Saturation Radiance

Most of the CZCS characteristics can be found in [Ball Aerospace Division \[1979\]](#). SeaWiFS characteristics are from [Barnes et al. \[1994a\]](#).

11.4.1 The CZCS

Figure 11.21 is taken from [Ball Aerospace Division \[1979\]](#). Figures 11.23 and 11.24 were taken from [Xiong et al. \[2013\]](#).

11.4.2 The MODIS

The use of the 1375 nm band to observe thin cirrus clouds is described in [Gao et al. \[1993\]](#).

11.6.1 Polar Orbit

The precession formula for the node of an inclined polar orbit is from [Goldstein \[1980\]](#).

11.8 Figures

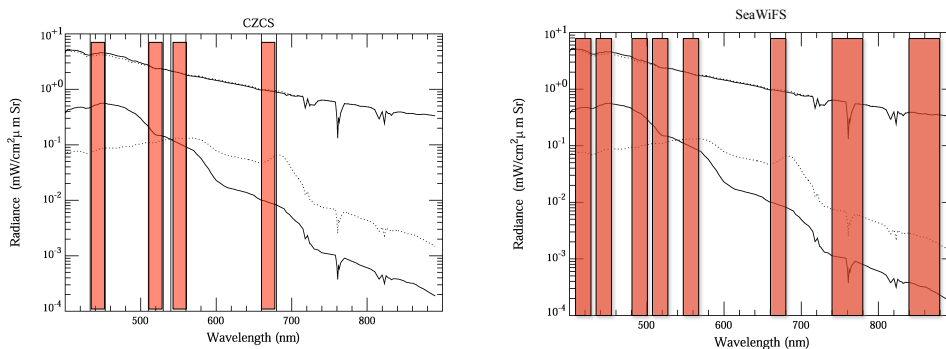


Figure 11.1: Left panel: spectral bands for the CZCS superimposed over the surface and top-of-atmosphere radiance spectrum for low (solid: $C_P = 0.01 \text{ mg/m}^3$) and high (dotted: $C_P = 10 \text{ mg/m}^3$) pigment concentrations. Right panel: same as the left panel but for SeaWiFS.

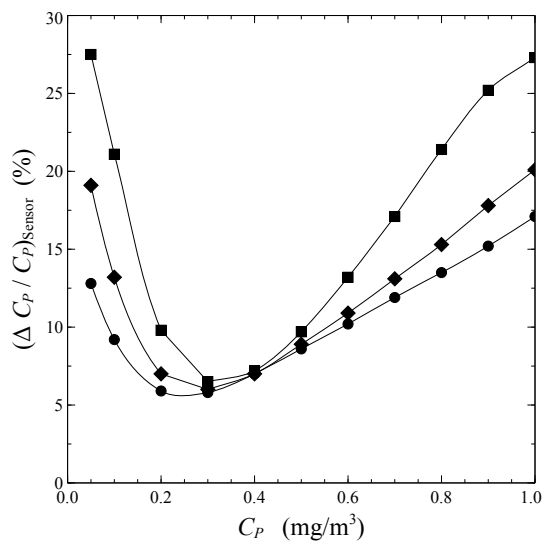


Figure 11.2: Result of simulations of sensor noise on the pigment concentration. Points represented as solid circles, diamonds, and squares are for $NE\Delta\rho_t(NIR) = 0, 1 \times 10^{-4}$ and 2×10^{-4} , respectively. For these simulations, $NE\Delta\rho$'s were 4.58×10^{-4} and 2.26×10^{-4} in $[\rho_w]_N$ at 443 and 555 nm, respectively, corresponding to $NE\Delta\rho$'s for the sensor ($NE\Delta\rho_t$) of 3.48×10^{-4} and 2.02×10^{-4} at 443 and 555 nm, respectively (nominal values $\theta_s = \theta_v = 30^\circ$ were used for computing the diffuse transmittances).

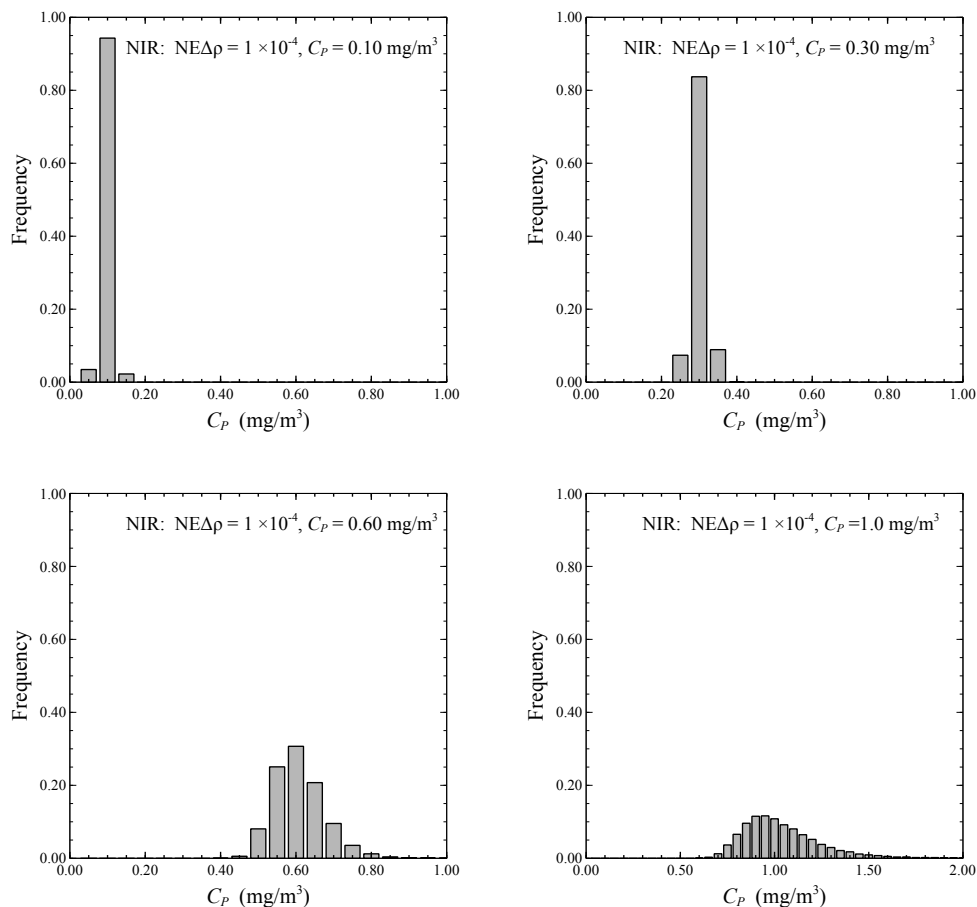


Figure 11.3: Histograms of the distribution of retrieved values of C_P for true values of $C_P = 0.1, 0.3, 0.6$, and 1.0 mg/m³. The mean values of the retrieved C_P 's were $0.099, 0.30, 0.61$, and 1.04 mg/m³, and the corresponding relative standard deviations ($\Delta C_P/C_P$) were $13.1, 6.0, 11.0$ and 20.2% . For these simulations, $NE\Delta\rho$'s were 4.58×10^{-4} and 2.26×10^{-4} in $[\rho_w]_N$ at 443 and 555 nm, respectively, corresponding to $NE\Delta\rho$'s for the sensor ($NE\Delta\rho_t$) of 3.48×10^{-4} and 2.02×10^{-4} at 443 and 555 nm, respectively (nominal values $\theta_s = \theta_v = 30^\circ$ were used for computing the diffuse transmittances). For the NIR, the $NE\Delta\rho$ for ρ_t was the same at both 765 and 865 nm with a value of 1×10^{-4} .

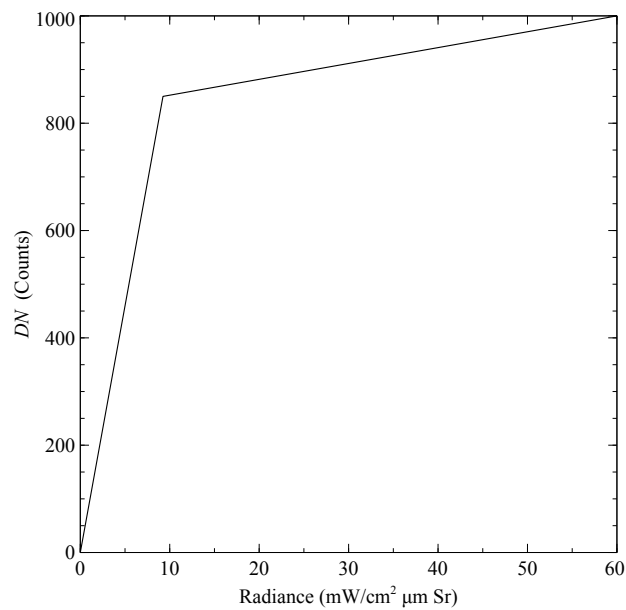


Figure 11.4: Bilinear analog-to-digital conversion curve for the 412 band on SeaWiFS. SeaWiFS digitization was 10 bit, i.e., 1024 radiance values.

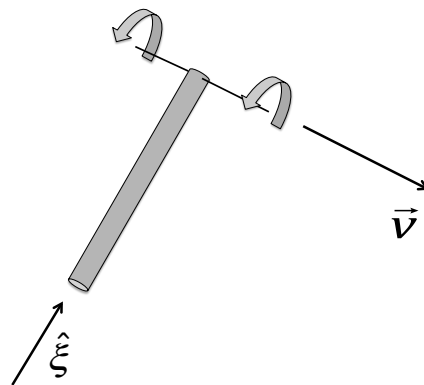


Figure 11.5: Schematic of a simplified scanning radiometer. The Gershun tube (plus detector element) rotates about an axis parallel to the velocity vector of the spacecraft.

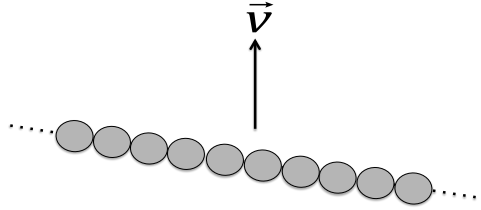


Figure 11.6: One complete rotation of the Gershun tube in Figure 11.5 produces a line of pixels slightly tilted to the velocity vector.

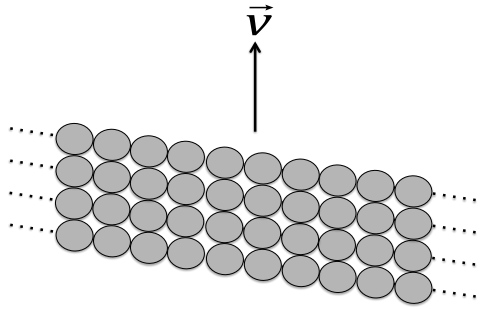


Figure 11.7: Four rotations of the Gershun tube results in four lines of pixels producing an image on the water surface. Here the pixel size s is related to the rotation rate \mathcal{R} by $s = v_g/\mathcal{R}$, where v_g is the ground speed of the satellite.

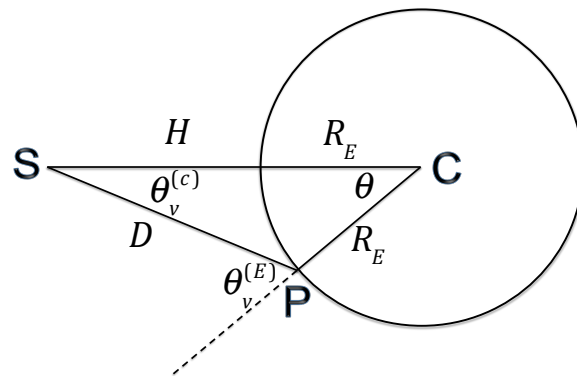


Figure 11.8: A sensor at “S” views a pixel at “P” on the surface of the Earth centered at “C”.

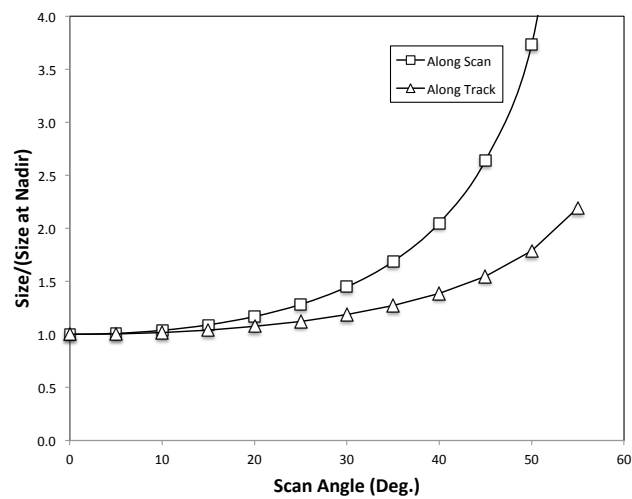


Figure 11.9: Ratio of s at a scan angle $\theta_v^{(c)}$ to that at nadir for a sensor at $H = 1000$ km.

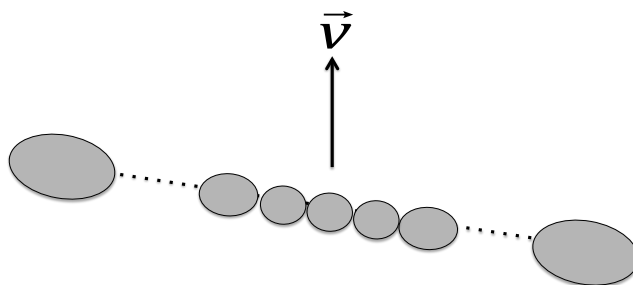


Figure 11.10: Schematic showing how the pixels increase in size (the “footprint” of the sensor) as the scan passes from nadir to the edge.

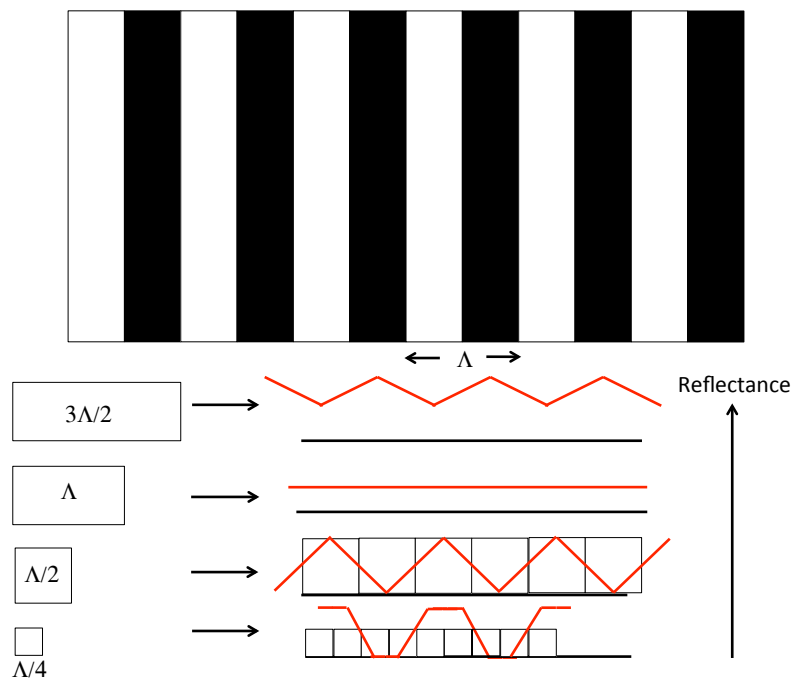


Figure 11.11: Bar chart and the result of imaging with several sized IFOVs. See text for details.

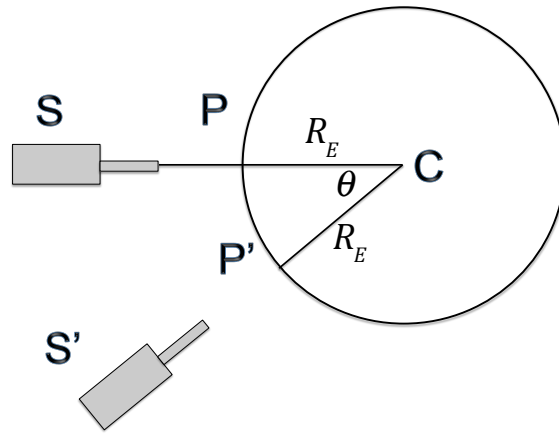


Figure 11.12: Schematic showing the sensor at S viewing a pixel P at nadir. Later, the sensor moves to S' and observes the nadir pixel P' . Clearly, in order to do this the spacecraft must have rotated through an angle θ .

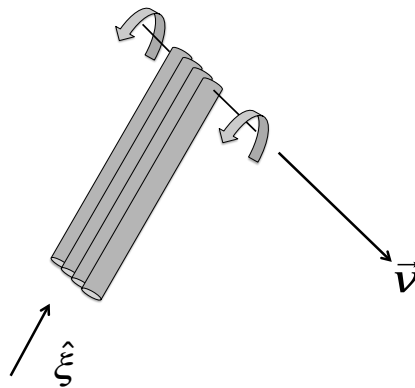


Figure 11.13: Obtaining a spectrum of the surface using multiple sensors.

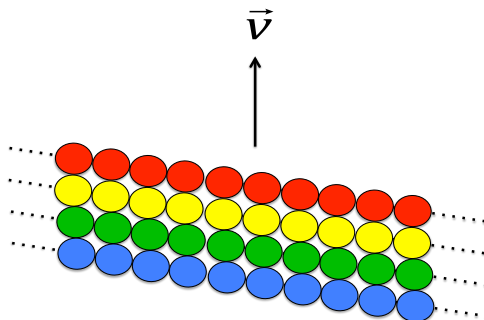


Figure 11.14: The sampling of the water surface using the arrangement in Figure 11.13.

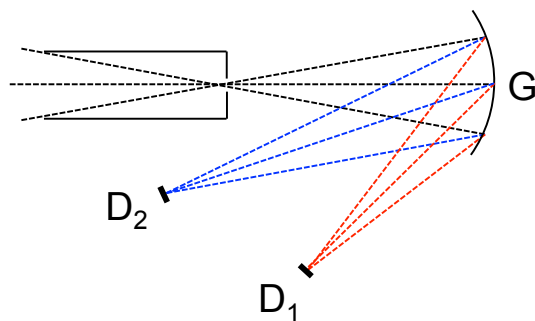


Figure 11.15: An alternative for achieving two spectral bands. The principal element here is a concave diffraction grating “G,” which separates the light into a spectrum as well as imaging the aperture of the Gershun tube. Separate portions the spectrum are imaged onto detectors D₁ and D₂. In this diagram, the size of the detectors would determine the spectral bandwidths of the two bands. This concept was used by the first ocean color scanner — CZCS.

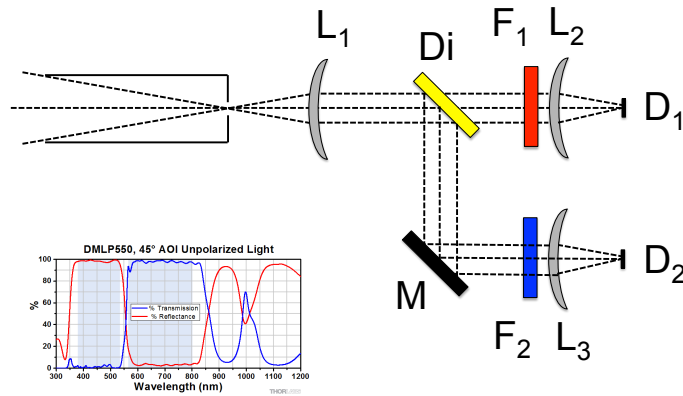


Figure 11.16: Alternative for achieving two spectral bands with a single Gershun tube scanner. The elements L_1 , L_2 and L_3 are focussing lenses, F_1 and F_2 are narrow band spectral filters, D_1 and D_2 are the detectors, M is a plane mirror and Di is a dichroic filter. The insert shows the transmission and reflective characteristics of a typical dichroic filter. It is from Thorlabs, Inc (www.thorlabs.com) for filter DMLP550, an example of a long-pass dichroic filter. An arrangement based on this concept was used in SeaWiFS.

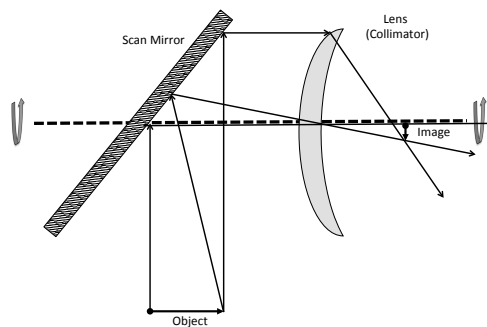


Figure 11.17: This is a design that would produce an image ("Image") of a pixel ("Object") on the water surface. The "Scan Mirror" rotates around an axis (dashed line) parallel to the velocity vector of the spacecraft.

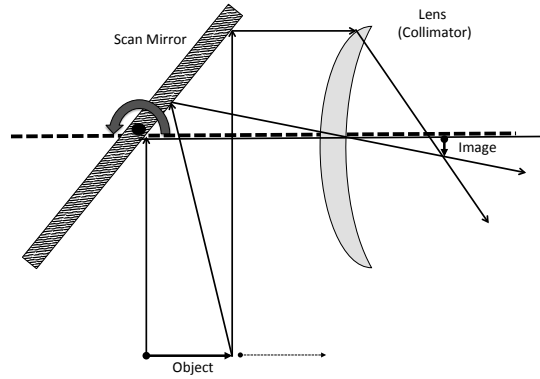


Figure 11.18: This is a design that would produce an image (“Image”) of a pixel (“Object”) on the water surface. The “Scan Mirror” rotates around an axis parallel to the velocity vector of the spacecraft, but unlike Figure 11.17 the velocity vector is now directed *into* the paper.

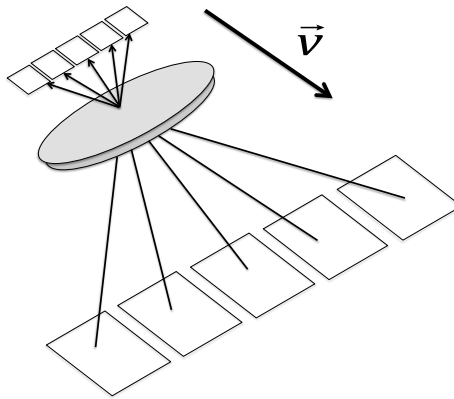


Figure 11.19: Concept of a push broom scanner. Instead of a single detector, there is a linear array of detectors aligned parallel to the Earth’s surface, but perpendicular to the velocity vector. The footprints (large squares at the bottom) are imaged by the optics (lens) onto the array (small squares at the top).

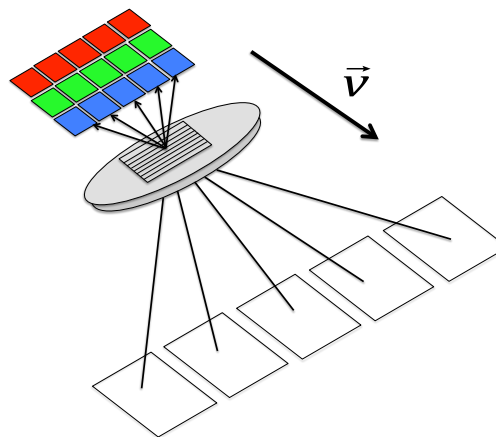


Figure 11.20: Concept of a hyperspectral push broom scanner. Here a diffraction grating has been placed behind the lens to disperse the spectrum and a rectangular array rather than a linear array is used. (Note that now the aperture of the instrument must be a line one pixel wide perpendicular to the flight line otherwise the next scan line and the spectrum would be intermixed.)

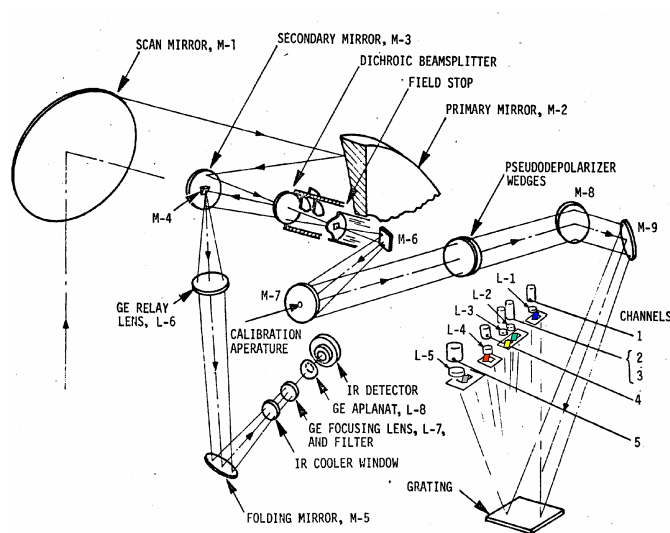


Figure 11.21: The optical layout of the CZCS. From [Ball Aerospace Division \[1979\]](#).

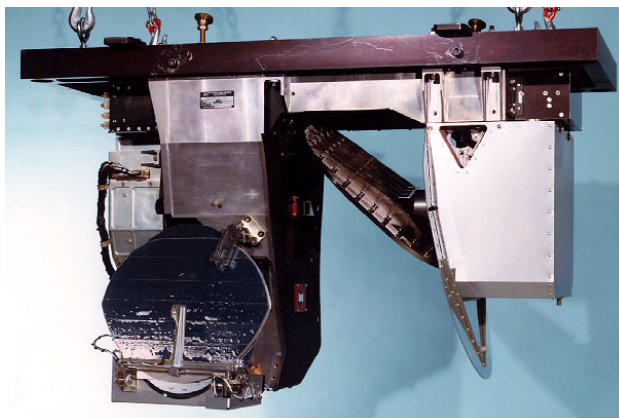


Figure 11.22: Photograph of the CZCS. The dimensions of the instrument are $81\text{ cm} \times 56\text{ cm} \times 38\text{ cm}$, and the total weight is 42 kg. The scan mirror can be seen at an angle of 45° to the horizontal. Its size is approximately $20\text{ cm} \times 28\text{ cm}$. The right box contains the scan motor assembly and the left box the spectrometer and electronics. The large shiny object at the lower left is the cover to the radiative cooler for the far-infrared detector.

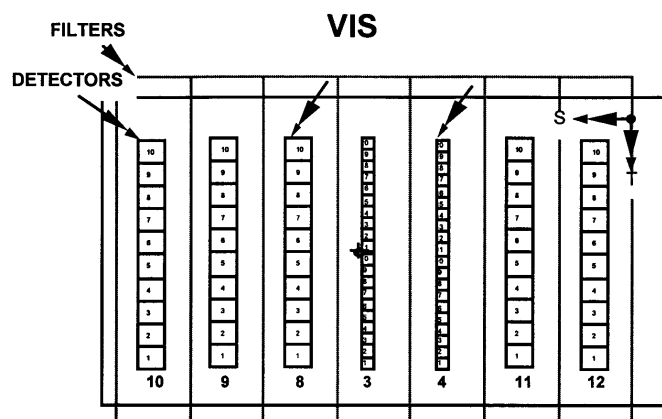


Figure 11.23: This is a schematic of the visible focal plane of the MODIS instrument. The scan direction is horizontal and the direction of the velocity vector of the satellite is vertical. The band numbers are along the bottom. Bands 8, 9, 10, 11, and 12 are visible bands for ocean observations with a resolution of 1 km at nadir. The resolution is determined by the physical size of the individual detectors. Bands 3 and 4 are high resolution (0.5 km) visible bands for land observations. Each ocean band has ten individual detectors labeled 1 through 10. When the scan mirror is in a position to image at nadir, an approximately $21\text{ km} \times 11\text{ km}$ image of the water surface falls on this focal plane.

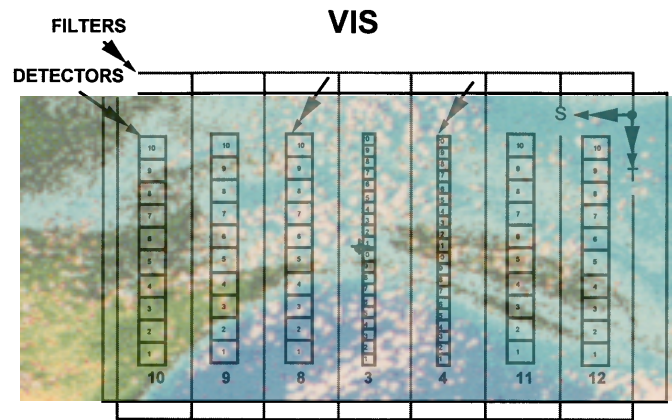


Figure 11.24: This is the same schematic as in Figure 11.23; however, a simulated image of the water surface has been superimposed on the focal plane. This is what happens in the operation of the sensor. As the scan mirror rotates the image will move across the focal plane from left to right. The image is then sampled by the individual detectors.

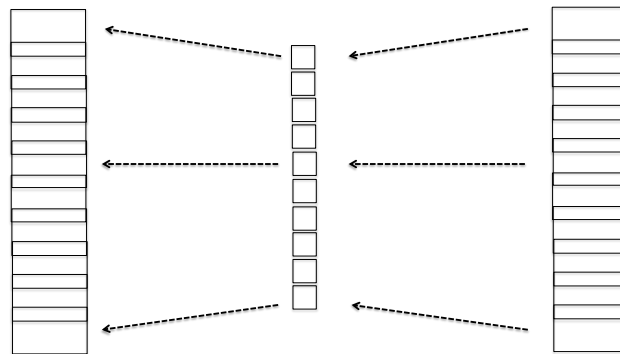


Figure 11.25: Schematic of the so-called “bowtie” effect in MODIS. The center line of pixels represents the footprint of the ten detectors of one spectral band at nadir. Note that they do not overlap. The pixels to the left and right are the footprints later and earlier in a single scan. They are larger and overlap. The overall footprint for one scan in shape resembles a bowtie.

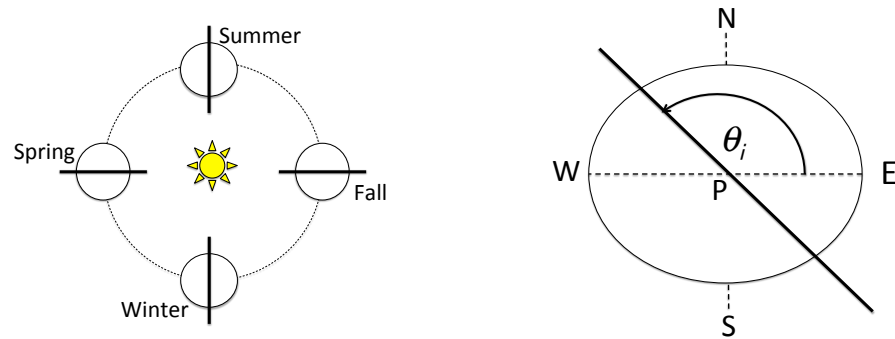


Figure 11.26: Left panel: the seasonal motion of the Earth around the Sun. The thick black line represents the plane of a polar-orbiting satellite's orbit. In order for the satellite to be over the illuminated portion of the Earth half of the time, the plane of the orbit must rotate through 360° in 365.25 days. Right panel: The plane of the satellite orbit around the Earth (thick line) inclined at an angle θ_i measured from the Equator at point "P" (the orbital node). This angle is adjusted for the plane of the orbit to carry out the motion shown in the left panel.

Chapter 12

Calibration and Characterization

12.1 Introduction

So far in our study of the remote sensing of water properties from measurements of reflectance ρ_t (or radiance) of the ocean-atmosphere system made from satellite altitudes, we defined the required (physical) quantity of interest, ρ_w , an apparent optical property, and related it to the inherent optical properties of the water and its constituents: a , b and β . We then used an example (the most important one, remote sensing the concentration of Chlorophyll a) to show how the IOPs relate to the particular constituent of interest. Finally, we developed a scheme to extract ρ_w from ρ_t — atmospheric correction. In all of our discussion we assumed that the reflectance ρ_t was accurately measured, i.e., error free (albeit not noise free). In this chapter we consider the consequences of error in ρ_t (a bias in the absence of noise), which we refer to as “calibration” error, and methods to minimize it.

In Chapters 9 and 10 we saw that in the blue ρ_w is at most 10-20% of ρ_t so, naively, it would seem that at first glance, a 5% error in ρ_t in the blue would lead to a 25-50% error in ρ_w . But wait, there is a somewhat complex atmospheric correction procedure employed to find ρ_w that involves two other spectral bands, each of which have their own calibration error. How does the error in these bands influence the error in the spectral band of interest? We actually need to know the effect of calibration errors on the error in the retrieved water-leaving reflectance *after* atmospheric correction. To estimate this, we will again use the single scattering atmospheric correction algorithm, which will certainly be adequate for the purpose. We start by writing the measured reflectance $\rho'_t(\lambda_i)$ (which

includes the error)¹ in terms of the true reflectance $\rho_t(\lambda_i)$ and the fractional error in $\rho_t(\lambda_i)$, which we denote by $\alpha(\lambda_i)$:²

$$\rho'_t(\lambda_i) = \rho_t(\lambda_i)(1 + \alpha(\lambda_i)). \quad (12.1)$$

These measured reflectances are then passed through the atmospheric correction algorithm to compute the estimated ρ'_w , i.e.,

$$t(\lambda_i)\rho'_w(\lambda_i) = \rho'_t(\lambda_i) - \rho_r(\lambda_i) - \varepsilon'(\lambda_i, \lambda_l)\rho'_{as}(\lambda_l),$$

where

$$\rho'_{as}(\lambda_l) = \rho'_t(\lambda_l) - \rho_r(\lambda_l)$$

and

$$\varepsilon'(\lambda_i, \lambda_l) = \frac{\rho'_{as}(\lambda_i)}{\rho'_{as}(\lambda_l)}.$$

But,

$$t(\lambda_i)\rho_w(\lambda_i) = \rho_t(\lambda_i) - \rho_r(\lambda_i) - \varepsilon(\lambda_i, \lambda_l)\rho_{as}(\lambda_l),$$

so

$$t(\lambda_i)(\rho'_w(\lambda_i) - \rho_w(\lambda_i)) \equiv t(\lambda_i)\Delta\rho_w(\lambda_i) = \alpha(\lambda_i)\rho_t(\lambda_i) + \varepsilon(\lambda_i, \lambda_l)\rho_{as}(\lambda_l) - \varepsilon'(\lambda_i, \lambda_l)\rho'_{as}(\lambda_l).$$

The last term can be written

$$\begin{aligned} \varepsilon'(\lambda_i, \lambda_l)\rho'_{as}(\lambda_l) &= (\varepsilon(\lambda_i, \lambda_l) + \Delta\varepsilon(\lambda_i, \lambda_l))(\rho'_t(\lambda_l) - \rho_r(\lambda_l)) \\ &= (\varepsilon(\lambda_i, \lambda_l) + \Delta\varepsilon(\lambda_i, \lambda_l))(\rho_{as}(\lambda_l) + \alpha(\lambda_l)\rho_t(\lambda_l)). \end{aligned}$$

Assuming the calibration errors are small quantities, i.e., $\alpha(\lambda_i) \ll 1$, and $\Delta\varepsilon(\lambda_i, \lambda_l)/\varepsilon(\lambda_i, \lambda_l) \ll 1$, upon multiplying this out, the term $\Delta\varepsilon'(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l)$ is second order in small quantities, so we'll ignore it compared to the others. Then

$$\varepsilon'(\lambda_i, \lambda_l)\rho'_{as}(\lambda_l) = \varepsilon'(\lambda_i, \lambda_l)\rho_{as}(\lambda_l) + \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l),$$

and

$$t(\lambda_i)\Delta\rho_w(\lambda_i) = (\varepsilon(\lambda_i, \lambda_l) - \varepsilon'(\lambda_i, \lambda_l))\rho_{as}(\lambda_l) + \alpha(\lambda_i)\rho_t(\lambda_i) - \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l). \quad (12.2)$$

¹Here all the quantities that include the calibration error are indicated by primes. The error-free quantities are unprimed.

²This is the error in ρ_t if the sensor is calibrated through a *reflectance-based* calibration (see Section 12.2.1 below). If a *radiance-based* calibration method is used, then the solar irradiance must be used to turn radiance into reflectance. In that case $\alpha = \alpha_L - \alpha_F$ (where the α 's are considered to be much less than unity), with $L'_t = (1 + \alpha_L)L_t$ and $F'_0 = (1 + \alpha_F)F_0$.

To complete the task, we need to evaluate $\varepsilon(\lambda_i, \lambda_l) - \varepsilon'(\lambda_i, \lambda_l)$. For this we need a model for $\varepsilon(\lambda_i, \lambda_l)$, and will assume, as in the single-scattering algorithm, that $\varepsilon(\lambda_i, \lambda_l) = \exp[k(\lambda_l - \lambda_i)]$. Then

$$\begin{aligned}
 \varepsilon(\lambda_i, \lambda_l) - \varepsilon'(\lambda_i, \lambda_l) &= \varepsilon(\lambda_i, \lambda_l) \left(1 - \frac{\varepsilon'(\lambda_i, \lambda_l)}{\varepsilon(\lambda_i, \lambda_l)} \right) \\
 &= \varepsilon(\lambda_i, \lambda_l) \left(1 - \exp[k'(\lambda_l - \lambda_i) - k(\lambda_l - \lambda_i)] \right) \\
 &= \varepsilon(\lambda_i, \lambda_l) \left(1 - \exp[\Delta k(\lambda_l - \lambda_i)] \right) \\
 &= \varepsilon(\lambda_i, \lambda_l) \left(-\Delta k(\lambda_l - \lambda_i) \right),
 \end{aligned} \tag{12.3}$$

where $\Delta k = k' - k$, and in the last equation we have dropped terms of order $(\Delta k)^2$ and higher. We can find Δk by using the band at λ_s , i.e.,

$$\Delta k = -\frac{1}{\lambda_l - \lambda_s} \left(\frac{\varepsilon(\lambda_s, \lambda_l) - \varepsilon'(\lambda_s, \lambda_l)}{\varepsilon(\lambda_s, \lambda_l)} \right). \tag{12.4}$$

But, we still need $\varepsilon(\lambda_s, \lambda_l) - \varepsilon'(\lambda_s, \lambda_l)$. Finding this is straightforward:

$$\begin{aligned}
 \varepsilon'(\lambda_s, \lambda_l) &= \frac{\rho'_t(\lambda_s) - \rho_r(\lambda_s)}{\rho'_t(\lambda_l) - \rho_r(\lambda_l)} \\
 &= \frac{\rho_t(\lambda_s) - \rho_r(\lambda_s) + \alpha(\lambda_s)\rho_t(\lambda_s)}{\rho_t(\lambda_l) - \rho_r(\lambda_l) + \alpha(\lambda_l)\rho_t(\lambda_l)} = \frac{\rho_{as}(\lambda_s) + \alpha(\lambda_s)\rho_t(\lambda_s)}{\rho_{as}(\lambda_l) + \alpha(\lambda_l)\rho_t(\lambda_l)} \\
 &= \left(\frac{\rho_{as}(\lambda_s)}{\rho_{as}(\lambda_l)} \right) \left(\frac{1 + \alpha(\lambda_s)\rho_t(\lambda_s)/\rho_{as}(\lambda_s)}{1 + \alpha(\lambda_l)\rho_t(\lambda_l)/\rho_{as}(\lambda_l)} \right) \\
 &= \left(\frac{\rho_{as}(\lambda_s)}{\rho_{as}(\lambda_l)} \right) \left(1 + \frac{\alpha(\lambda_s)\rho_t(\lambda_s)}{\rho_{as}(\lambda_s)} - \frac{\alpha(\lambda_l)\rho_t(\lambda_l)}{\rho_{as}(\lambda_l)} \right) \\
 &= \varepsilon(\lambda_s, \lambda_l) \left(1 + \frac{\alpha(\lambda_s)\rho_t(\lambda_s)}{\rho_{as}(\lambda_s)} - \frac{\alpha(\lambda_l)\rho_t(\lambda_l)}{\rho_{as}(\lambda_l)} \right),
 \end{aligned} \tag{12.5}$$

where the next to last step is valid because of the smallness of the α 's. This yields

$$\left(\frac{\varepsilon(\lambda_s, \lambda_l) - \varepsilon'(\lambda_s, \lambda_l)}{\varepsilon(\lambda_s, \lambda_l)} \right) = \left(-\frac{\alpha(\lambda_s)\rho_t(\lambda_s)}{\rho_{as}(\lambda_s)} + \frac{\alpha(\lambda_l)\rho_t(\lambda_l)}{\rho_{as}(\lambda_l)} \right),$$

so

$$\Delta k = \frac{1}{\lambda_l - \lambda_s} \left(\frac{\alpha(\lambda_s)\rho_t(\lambda_s)}{\rho_{as}(\lambda_s)} - \frac{\alpha(\lambda_l)\rho_t(\lambda_l)}{\rho_{as}(\lambda_l)} \right) \tag{12.6}$$

Substituting Eqs. (12.6) and (12.3) into Eq. (12.2) yields our final answer:

$$\begin{aligned}
 t(\lambda_i)\Delta\rho_w(\lambda_i) &= \alpha(\lambda_i)\rho_t(\lambda_i) - \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l) \\
 &\quad - \left(\frac{\lambda_l - \lambda_i}{\lambda_l - \lambda_s} \right) \left(\frac{\varepsilon(\lambda_i, \lambda_l)}{\varepsilon(\lambda_s, \lambda_l)} \alpha(\lambda_s)\rho_t(\lambda_s) - \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l) \right).
 \end{aligned} \tag{12.7}$$

The first term in Eq. (12.7) is the direct result of calibration error in the spectral band of interest at λ_i , while the second results from error at λ_l propagated to λ_i through the atmospheric correction process. The term on the second line is due to the error in $\varepsilon(\lambda_s, \lambda_l)$ generated by error at λ_s and λ_l and again propagated to λ_i through atmospheric correction. We note that if all of the calibration biases have the same sign there will be some cancelation of the terms in Eq. (12.7).

Let's use an example to gain some insight into the magnitude of the various terms. Recall that for SeaWiFS, $\lambda_s = 765$ nm and $\lambda_l = 865$ nm. From one of the stations in Figure 10.4 in Chapter 10 we obtain $\rho_t(443) \approx 0.16$, $\rho_t(765) \approx 0.022$ and $\rho_t(865) \approx 0.019$. Note that for Maritime aerosols all of the ε values are close to unity, so we will assume for the purpose of this example that they are unity. If we assume a calibration bias of +5% in each spectral band, then the magnitudes of the three terms in Eq. (12.7) are

$$\text{Term 1} = +8.00 \times 10^{-3}$$

$$\text{Term 2} = -9.50 \times 10^{-4}$$

$$\text{Term 3} = -6.33 \times 10^{-4}$$

Clearly, the most important of the three biases is $\alpha(443)$. This would need to be reduced to about 2% in order to achieve the desired accuracy in $\rho_w(443)$ of ± 0.002 (ignoring t of course). Consider a different scenario. Let $\alpha(865) = -0.01$ and $\alpha(765) = +0.01$, i.e., calibration biases that are equal but of *opposite* sign in the two NIR bands. In this case

$$\text{Term 2} + \text{Term 3} = -1.58 \times 10^{-3},$$

i.e., contributing as much to the error in $t\rho_w(443)$ as they did in the case where $\alpha(865) = \alpha(765) = +0.05$.³ This underscores the importance of the calibration errors in the two NIR bands having the same sign. Later in this chapter we will develop a method of vicarious calibration that guarantees that the calibration biases will all have the same sign.

Let us look at some sample results derived with the full multiple scattering algorithm. In these cases the multiple scattering algorithm was operated by inserting values of $\rho'_t(\lambda)$ rather than $\rho_t(\lambda)$. We assumed that the actual aerosol is M80 and use the original set of candidate aerosol models (M50, \dots , T99). The aerosol optical depth at 865 nm was 0.2, and viewing was at nadir. The error in $t\rho_w(443)$ is provided as a function of the the solar zenith angle for various bias scenarios in Figure 12.1. The various panels of Figure 12.1 are in quantitative agreement with the estimates above after allowance is made for the fact that unity ε 's were used in the single-scattering algorithm estimates and the ρ_t 's probably were not characteristic of M80. Assuming that all of the calibration biases are of the same

³If the calibration errors in the bands at 443, 765, and 865 nm were +5%, +5%, and -5%, respectively, then the error in $t\rho_w(443)$ would only be $+0.1 \times 10^{-3}$, an excellent result. But, remember that the calibration error is *unknown*, i.e., it is not at our disposal to adjust.

sign, the results above and in the figure suggests that most of the error is due to $\alpha(443)$ and that to achieve the desired accuracy, calibration uncertainty of only the order of 1% can be tolerated. This accuracy is difficult to achieve even in a laboratory setting, but we shall see that it can even be exceeded through methods of vicarious calibration described later in this chapter.

We start our discussion of calibration of ocean color sensors by carefully defining what we mean by “radiometric calibration,” and two ways to effect it. As this calibration is usually carried out pre-launch, there is uncertainty as to how well it still applies after launch. In fact, even in the laboratory it is difficult to achieve the necessary calibration uncertainty required for adequate retrieval of ρ_w . Thus, post-launch on-orbit calibration or “vicarious calibration” schemes have been developed for adjustment of the pre-launch calibration. We discuss two of these techniques, and show that they result in a considerable improvement in the radiometric calibration beyond the pre-launch effort. Then, realizing that each sensor has its own peculiarities and idiosyncrasies, we discuss some prelaunch measurements (“characterization”) that are required to account for their effects.

12.2 Radiometric Calibration

So far we have based our discussion on radiance, or reflectance derived from radiance. For review, recall that the radiance is operationally defined in the following manner. Start with a radiation detector of area ΔA_D . This has the property that it provides an electrical response, say a current $i(\mathcal{P})$, when its sensitive area is fully illuminated by a radiant power \mathcal{P} . We assume that the radiant power is a continuous function of wavelength, i.e., if $\Delta\mathcal{P}(\hat{\xi}, \lambda_0, \Delta\lambda)$ is the radiant power propagating in a direction $\hat{\xi}$ with wavelength in the interval between $\lambda_0 - \Delta\lambda/2$ and $\lambda_0 + \Delta\lambda/2$ falling on the detector, then as $\Delta\lambda \rightarrow 0$, $\Delta\mathcal{P}(\hat{\xi}, \lambda_0, \Delta\lambda)/\Delta\lambda$ converges to a well defined limit. Now, restrict the field of view of the detector by placing it at one end of a tube (Gershun tube), allowing the detector to receive light propagating only within a narrow range of angles around a given direction ($\hat{\xi}$) defined by the solid angle $\Delta\Omega(\hat{\xi})$.⁴ Arrange the surface of the detector to be perpendicular to the direction $\hat{\xi}$, and place a narrow band spectral filter in front of the detector so that only light within a spectral interval $\Delta\lambda$ centered on the wavelength λ_0 can reach the detector.⁵

⁴The open end of the tube is pointing in the direction $-\hat{\xi}$. The solid angle $\Delta\Omega(\hat{\xi})$ is the area of the open end of the tube divided by the square of its length.

⁵We assume the spectral filter is 100% transmitting for radiation with wavelengths within $\Delta\lambda$, and has no transmission for wavelengths outside $\Delta\lambda$. This assumption will be relaxed later in this chapter.

Under these conditions the detector records $\Delta^3\mathcal{P}(\hat{\xi}, \lambda_0)$ and the radiance is defined by

$$L(\hat{\xi}, \lambda_0) = \frac{\Delta^3\mathcal{P}(\hat{\xi}, \lambda_0)}{\Delta\Omega(\hat{\xi})\Delta A_D\Delta\lambda}$$

Because the $\Delta^3\mathcal{P}(\hat{\xi}, \lambda_0)$ is expected to be proportional to each of the quantities in the denominator, this ratio will approach a well defined limit for sufficiently small values of $\Delta\Omega(\hat{\xi})$, ΔA_D and $\Delta\lambda$.

In Chapter 2 we found that the radiant power $\Delta^3\mathcal{P}$ and the current from the detector i are proportional to one another, i.e.,

$$i(\lambda) \propto \Delta^3\mathcal{P}(\lambda) = L(\lambda)\Delta\lambda\Delta\Omega\Delta A_D.$$

Let $\kappa'(\lambda)$ be the proportionality constant, so

$$i(\lambda) = \kappa'(\lambda)L(\lambda)\Delta\lambda\Delta\Omega\Delta A_D.$$

If the filter has transmittance $T(\lambda)$, then the power reaching it is reduced the factor $T(\lambda)$ and in the above equation $L(\lambda)$ is replaced by $L(\lambda)T(\lambda)$. For a given radiometer $\Delta\Omega$ and ΔA_D are fixed and so can be incorporated into the proportionality constant. It is also convenient to incorporate $T(\lambda)$ into the constant as well and write $\kappa = \kappa'T\Delta\Omega\Delta A_D$. Then,

$$i(\lambda) = \kappa(\lambda)L(\lambda)\Delta\lambda. \quad (12.8)$$

(Note that $\kappa(\lambda)$ has dimensions of $i/L\Delta\lambda$, for example Amperes over Watts per meter squared Steradian.) The proportionality “constant” $\kappa(\lambda)$ is not only dependent on wavelength (because of T and also because radiation detectors have a response that is wavelength dependent) but on other factors such a temperature, etc. Where it is necessary to remind ourselves of this, we will write κ as $\kappa(\lambda, \dots)$, where \dots represents these additional dependencies.

If we measure $\kappa_B(\lambda)$ for all wavelengths for a particular spectral band B , by introducing quasi-monochromatic radiance⁶ into the radiometer, the resulting current will be

$$I_B = \int \kappa_B(\lambda)L(\lambda) d\lambda.$$

Now let

$$S_B(\lambda) \equiv \frac{\kappa_B(\lambda)}{\int \kappa_B(\lambda) d\lambda}.$$

⁶This radiance need not be monochromatic, but must have a spectral band width $\Delta\lambda$ that is small compared to the nominal spectral width of B .

$S_B(\lambda)$ is usually called the *relative spectral response* or *normalized spectral response* for band B . With the substitution,

$$I_B = \left(\int S_B(\lambda) L(\lambda) d\lambda \right) \left(\int \kappa_B(\lambda) d\lambda \right) = \langle L \rangle_B \int \kappa_B(\lambda) d\lambda,$$

where

$$\langle L \rangle_B \equiv \int S_B(\lambda) L(\lambda) d\lambda \quad (12.9)$$

is called the *band-integrated radiance* or just the *band radiance* in band B . It is convenient to redefine the integrated κ to be $1/K_B$, i.e.,

$$K_B = \frac{1}{\int \kappa_B(\lambda) d\lambda},$$

so

$$\langle L \rangle_B = K_B I_B, \quad (12.10)$$

and we shall refer to K_B as the *calibration constant* for band B .

By *radiometric calibration* we mean the quantitative relating of the electrical output of the detector (I_B) in a radiometer to the radiance of the scene $\langle L \rangle_B$ being viewed as in Eq. (12.10). This is usually accomplished by having the radiometer view a source of known radiance; a process called *radiance based* calibration. Usually the source of radiance is varied in magnitude to detect any nonlinearities in the current-radiance relationship. A second kind of calibration is known as *reflectance based*. In this case the radiometer is aimed at a target of known reflectance that is illuminated by a given source, and its output recorded. Then when the radiometer views a scene that is *illuminated by the same source* the radiometer's output determines the scene's "reflectance."⁷

12.2.1 Radiance Based Calibration

For radiance based calibration, we need to provide a known source of radiance. This can be accomplished by shining light onto a white diffusing material that has known reflecting properties. For example, the reflected radiance might be almost independent of the viewing angle and the illumination angle, as is nearly the case for reflection from a sheet of (non-glossy) white paper, i.e., a lambertian reflector. Such surfaces are available and are called

⁷In both of these methods the source is usually broad-band radiance, i.e., from an incandescent lamp rather than monochromatic radiation. Thus, we can find K_B , but in order to find $\kappa(\lambda)$ we need quasi-monochromatic radiation. Because producing quasi-monochromatic radiance at the required power levels that also fill the aperture of the sensor is very difficult, the calibration is usually carried out in two steps. First, K_B is determined using a broad-band source, then $\kappa(\lambda)$ is estimated using radiance from a monochromator and normalized to yield $S_B(\lambda)$.

reflectance plaques. Here, for simplicity, we will assume that the plaque is lambertian and has a known reflectance, $R_P(\lambda)$. The illuminated plaque must be large enough to fill the field of view of the radiometer, defined by $\Delta\Omega(\hat{\xi})$. The illumination of the plaque is from a standard lamp, which provides a known spectral irradiance $E_L(\lambda, D)$ at a fixed distance (D) from the plaque when operated at a specified current from a current-regulated power supply. The lamp irradiance is traceable to similar lamps calibrated at national standards laboratories. Because the source is an incandescent filament, its spectral irradiance will vary slowly with wavelength. Thus, illuminating the plaque in this manner provides a reflected irradiance of $R_P(\lambda)E_L(\lambda, D)$. Because the plaque is lambertian, the reflected radiance is $L(\lambda, D) = R_P(\lambda)E_L(\lambda, D)/\pi$. For band B ,

$$\frac{R_P(\lambda_B)E_L(\lambda_B, D)}{\pi} = \langle L \rangle_B = K_B I_B(D), \quad (12.11)$$

where λ_B is the nominal wavelength at the center of the band. Then, measuring $I_B(D)$, provides K_B . One then varies D noting that $E_L(\lambda, D)/E_L(\lambda, D') = D^2/D'^2$, allowing a determination of any nonlinearities in the detector (variation in K_B with the magnitude of the radiance measured).⁸ If the detector is linear, K_B is a constant, and $I_B(D)/I_B(D') = D^2/D'^2$.

For large radiometers, such as ocean color sensors, rather than using a flat reflectance plaque a *calibration sphere* is used. This is a hollow sphere painted with flat white paint inside and containing a large number (usually more than 10) standard calibration lamps. A hole is cut in the sphere and the radiometer to be calibrated is arranged to view into the hole. The lamps are arranged so that none are in the field of view of the radiometer. The radiance exiting the hole is totally diffuse, and known as a function of the number of lamps that are turned on.

12.2.2 Reflectance Based Calibration

In our development of the atmospheric correction algorithm we found it convenient to use the reflectance $\rho(\lambda)$ in place of the radiance $L(\lambda)$. The reflectance associated with a given radiance was defined by

$$\rho(\lambda) = \frac{\pi L(\lambda)}{F_0(\lambda) \cos \theta_0},$$

where $F_0(\lambda)$ is the instantaneous extraterrestrial solar irradiance (on a plane normal to the solar beam) and as usual, θ_0 is the solar zenith angle. For satellite radiometers it is

⁸The relationship $E_L(\lambda_0, D)/E_L(\lambda, D') = D^2/D'^2$ is only approximate since the lamp filament is not a point source. However, the actual relationship between $E_L(\lambda, D)$ and $E_L(\lambda, D')$ can be determined experimentally.

actually more practical to calibrate the instrument in terms of reflectance, as this can be accomplished on orbit.

All ocean color instruments are scanners, i.e., they have a mechanical mechanism that allows them to view in different directions (Chapter 11). This is equivalent to pointing a gershun tube in different directions. This is often accomplished via a rotating mirror. Figure 12.2 shows a schematic of how on-orbit reflectance-based calibration might be carried out. In the figure we see the radiometer viewing the scene in a nadir direction where the radiance is L_t . Then it points in a different direction, viewing a reflectance plaque that is also illuminated by the solar beam and records a radiance L_P as shown in the figure. The plaque has a reflectance $R_P(\hat{\xi}_P, \hat{\xi}_V)$ when light propagating in a direction $\hat{\xi}_P$ reflects off in a direction $\hat{\xi}_V$. The reflectance R_P (called the BRDF of the plaque) is determined for all $\hat{\xi}_P$ and $\hat{\xi}_V$ prior to launch. If the plaque is lambertian, R_P is independent of both $\hat{\xi}_P$ and $\hat{\xi}_V$. When viewing the scene the radiance is given by $\langle L_t \rangle_B = K_B I_{tB}$ and when viewing the plaque, $\langle L_P \rangle_B = K_B I_{PB}$. But

$$K_B I_{tB} = \langle L_t \rangle_B = \frac{F_0(\lambda_B) \cos \theta_0 \langle \rho_t \rangle_B}{\pi} \quad \text{and} \quad K_B I_{PB} = \langle L_P \rangle_B = \frac{F_0(\lambda_B) \cos \theta_0 \langle \rho_P \rangle_B}{\pi},$$

where, as in the last section, we have assumed that F_0 is a lowly varying function of λ . Finally, since $\langle \rho_P \rangle_B = \langle R_P \rangle_B$, we have

$$\langle \rho_t \rangle_B = \langle R_P \rangle_B \left[\frac{\cos \theta_P}{\cos \theta_0} \right] \left[\frac{I_{tB}}{I_{PB}} \right].$$

Note that F_0 and K_B do not appear in the final result.⁹ So the calibration constant K does not even have to be known, as long as the detection system is linear. Even if the detection system is not precisely linear, if the ratio I_{tB}/I_{PB} is close to unity, which requires $L_t(\lambda)/L_P(\lambda)$ to be close to unity, then the calibration constants are still not needed; however, this is difficult to achieve, as most plaques have a reflectance that is more or less spectrally neutral, but the scene L_t varies strongly from blue to the NIR.¹⁰

Experience shows that the reflectance of solar diffuser plaques in orbit degrade with time. If account is not taken of this degradation, the estimated ρ_t will be too large. This is remedied by monitoring the stability of R_P . To effect this monitoring, SeaWiFS uses the Moon (see Chapter 8), and MODIS uses an onboard monitoring device and occasional Lunar views.

⁹If there are any nonlinearities in the detection system then there will be a ratio of K 's at the two radiance levels, L_t and L_P , in the final result.

¹⁰The above analysis is correct as long as the spectral band B is relatively narrow compared to the visible spectrum. For SeaWiFS, the spectral bands were nominally 20 or 40 nm wide; however, some of them had significant response outside the nominal bands. This requires special treatment that is discussed in a later section.

12.3 Vicarious Calibration

The radiometric calibration of ocean color sensors to be flown in space is a particularly difficult task as the instruments are physically large and the calibration constants, $K_B(\dots)$, must be measured with the instrument in a vacuum chamber. This is expensive, but even if it is effected, how do we know that the same values of $K_B(\dots)$ apply after the violence of launch? This question makes the concept of *vicarious calibration* very attractive. In general vicarious calibration means calibrating the sensor in orbit by determining the radiance (or reflectance) that it *should* be recording when viewing a particular scene.¹¹ For ocean color sensors this requires that surface measurements of the water-leaving radiance as well as measurements to determine the state of the atmosphere, e.g., surface pressure, aerosol type and concentration, etc., are carried out in order to use radiative transfer to determine what we believe to be the actual radiance at the sensor.

In this section we describe two methods for carrying out vicarious calibration. The first provides a near-radiometric calibration and the second is capable of providing a calibration that in the blue is even *better* than could be achieved in a laboratory calibration. The latter method has been used extensively for vicariously calibrating the U.S. ocean color sensors SeaWiFS, MODIS Terra, and MODIS Aqua.

12.3.1 Direct — Using Surface Radiance Measurements

In the direct method, the basic idea is to use measurements of the sky radiance at the water surface in various directions, made with a well-calibrated radiometer (and atmospheric direct transmission for optical depth, if necessary), and deduce the radiance exiting the top of the atmosphere toward the sensor. Ideally one would like measurements of the sky radiance and satellite-sensed radiance to occur at the same time and location to minimize the effects of changes in the atmosphere. As a simplified example of such a procedure, consider a homogeneous atmosphere with a totally absorbing (nonreflecting) lower boundary. Radiance exits the top of the atmosphere (TOA) propagating in a direction $\hat{\xi}_T$ having been scattered from the solar beam propagating in the direction $\hat{\xi}_0$. Assume for the present that the optical thickness τ is sufficiently small that the single-scattering solution of the radiative transfer equation accurately provides the exiting radiances. The reflectance $\rho_T(\hat{\xi}_T)$ is

¹¹This is analogous to the methods in the previous section, where the radiance or reflectance was determined by reflectance plaques and standard lamps. The optical properties of these “tools” must be known or determined prior to the calibration. In this case the “tool” is the ocean-atmosphere system whose properties need to be determined prior to the calibration. In this sense the term “vicarious” is a misnomer, as the essential difference is the temporal stability of the tools: plaques and lamps are very stable, while the reflectance of the ocean-atmosphere system is not.

given by¹²

$$\rho_T(\hat{\xi}_T) = \frac{\omega_a \tau P(\Theta_T)}{|\cos \theta_T| \cos \theta_0}, \quad (12.12)$$

where, $\cos \Theta_T = \hat{\xi}_T \bullet \hat{\xi}_0$, with

$$\begin{aligned} \hat{\xi}_0 &= \hat{e}_x \cos \phi_0 \sin \theta_0 + \hat{e}_y \sin \phi_0 \sin \theta_0 + \hat{e}_z \cos \theta_0, \\ \hat{\xi}_T &= \hat{e}_x \cos \phi_T \sin \theta_T + \hat{e}_y \sin \phi_T \sin \theta_T + \hat{e}_z \cos \theta_T. \end{aligned}$$

Now, for the radiometer at the bottom of the atmosphere (BOA), the reflectance¹³ is

$$\rho_B(\hat{\xi}_B) = \frac{\omega_a \tau P(\Theta_B)}{\cos \theta_B \cos \theta_0}, \quad (12.13)$$

where, $\cos \Theta_B = \hat{\xi}_B \bullet \hat{\xi}_0$, and

$$\hat{\xi}_B = \hat{e}_x \cos \phi_B \sin \theta_B + \hat{e}_y \sin \phi_B \sin \theta_B + \hat{e}_z \cos \theta_B.$$

Now, if it can be arranged that $\Theta_B = \Theta_T$, we have

$$|\cos \theta_T| \rho_T(\hat{\xi}_T) = \cos \theta_B \rho_B(\hat{\xi}_B) \quad \text{or} \quad |\cos \theta_T| L_T(\hat{\xi}_T) = \cos \theta_B L_B(\hat{\xi}_B), \quad (12.14)$$

so measurement of L_B provides L_T directly. In addition the uncertainty in L_T is given directly as the uncertainty in the measurement by the BOA radiometer, i.e., L_B .

Is it always possible to arrange to have $\Theta_B = \Theta_T$? Actually for this to be possible usually requires a relatively large solar zenith angle. Given the sensing geometry, we can always find Θ_T . Then, if $\Theta_B = \Theta_T$,

$$\cos \Theta_T = \cos \Theta_B = \cos \phi_B \sin \theta_B \cos \phi_0 \sin \theta_0 + \sin \phi_B \sin \theta_B \sin \phi_0 \sin \theta_0 + \cos \theta_B \cos \theta_0$$

or

$$\cos \Theta_T = \sin \theta_B \sin \theta_0 \cos(\phi_B - \phi_0) + \cos \theta_B \cos \theta_0, \quad (12.15)$$

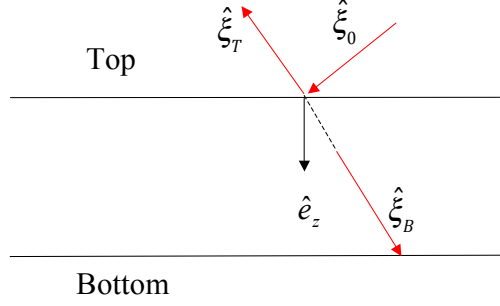
so

$$\cos(\phi_B - \phi_0) = \frac{\cos \Theta_T - \cos \theta_B \cos \theta_0}{\sin \theta_B \sin \theta_0}. \quad (12.16)$$

Consider a specific example. Let $\theta_0 = 45^\circ$ and $\phi_0 = 0^\circ$ along with $\theta_T = 135^\circ$ and $\phi_T = 0^\circ$. A simple drawing, or the equation following Eq. (12.12), shows that $\Theta_T = 90^\circ$. Clearly, if we choose $\theta_B = 45^\circ$ and $\phi_B = 180^\circ$, as shown below (where all of the $\hat{\xi}$ vectors are in the same plane, the principal plane), then $\Theta_B = \Theta_T = 90^\circ$. Equations (12.15) and

¹²Here, for simplicity, we will stop using the angle brackets to denote band averages, but understand that these *are* band averages.

¹³The term “reflectance” here is a misnomer as there is nothing “reflected.” A better term might be “normalized radiance.” In any event, as with all radiance-reflectance conversions, $\rho_B = \pi L_B / F_0 \cos \theta_0$.



(12.16) confirm that we are correct. Are there any other directions for $\hat{\xi}_B$ that also have $\Theta_B = \Theta_T = 90^\circ$? Clearly, if the plane formed by $\hat{\xi}_T$, $\hat{\xi}_B$, and $\hat{\xi}_0$ is rotated about $\hat{\xi}_0$, Θ_T and Θ_B remain equal to 90° . Does Eq. (12.16) confirm this? For a given θ_B Eq. (12.16) shows that

$$\cos(\phi_B - \phi_0) = \frac{-1}{\tan \theta_B \tan \theta_0}.$$

But for $\theta_0 = 45^\circ$ this gives $\cos(\phi_B - \phi_0) = -1/\tan \theta_B$. In order that $\phi_B - \phi_0$ be in its allowable range, i.e., $-1 \leq \cos(\phi_B - \phi_0) \leq +1$, we see that $\theta_B \geq 45^\circ$. For example, if $\theta_B = 90^\circ$, $\phi_B - \phi_0 = 90^\circ$. Thus, for any $45^\circ \leq \theta_B \leq 90^\circ$ there will always be a solution with $90^\circ \leq \phi_B - \phi_0 \leq 180^\circ$, so the answer is, yes. Clearly, Eq. (12.16) shows that as θ_0 decreases, the range of $\hat{\xi}_B$'s over which $\Theta_B = \Theta_T$, decreases, i.e., the range of θ_B for which $\phi_B - \phi_0 \leq 180^\circ$ are obtained, decreases. In practice, since we are using the single-scattering formulae, we would want to use the $\hat{\xi}_B$ with the smallest θ_B to minimize multiple-scattering effects.

An alternative method to using these equations is to employ measurements of ρ_B for various $\hat{\xi}_B$ in Eq. (12.13) to determine $\omega_a \tau P(\Theta)$ and then use these in Eq. (12.12) to determine ρ_T . The scattering angular range available to ρ_B is $0 \leq \Theta_B \leq \theta_0 + \pi/2$ (in the principal plane of the Sun $\phi_B - \phi_0 = 0$ and π). As long as Θ_T falls within this range for Θ_B the procedure will work and on-orbit calibration can be effected.

In the above discussion we have ignored any reflection by the water surface. The presence of a Fresnel-reflecting water surface adds significant complications. Consider first, the normalized radiance at the bottom of the atmosphere. We note that in the single-scattering approximation, the contributions of Rayleigh and aerosol scattering to the reflectance are completely separable. Furthermore, we can compute the Rayleigh contribution exactly, as the Rayleigh phase function and optical thickness are both known. Thus, we will assume in our discussion that this has been carried out and the Rayleigh contribution has been removed from ρ_B (and ρ_T), and confine our attention to the aerosol component. In the

single-single scattering approximation this component is given by

$$\rho_B(\theta_B) = \frac{\omega_a \tau_a}{4 \cos \theta_B \cos \theta_0} \left[P_a(\Theta_B) + r_f(\theta_0) P_a(\Theta_B^{(r)}) \right], \quad (12.17)$$

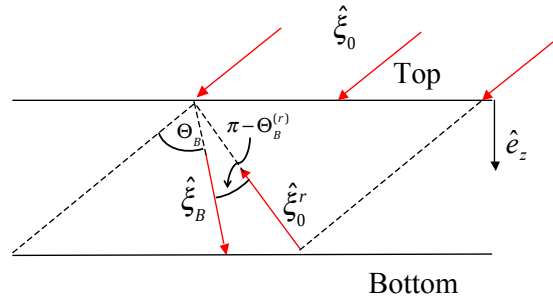
where $r_f(\theta_0)$ is the Fresnel reflectance for an incident angle θ_0 ,

$$\cos \Theta_B = \hat{\xi}_0 \bullet \hat{\xi}_B \quad \text{and} \quad \cos \Theta_B^{(r)} = \hat{\xi}_0^r \bullet \hat{\xi}_B,$$

with $\hat{\xi}_0^r$ being the direction of the solar beam *after* being reflected from the water surface (note the “ $-$ ” sign indicated in the equation below by the arrow):

$$\hat{\xi}_0^r = \cos \phi_0 \sin \theta_0 \hat{e}_x + \sin \phi_0 \sin \theta_0 \hat{e}_y - \cos \theta_0 \hat{e}_z.$$

These angles and the unit vectors are shown in the figure below, where for simplicity all of the $\hat{\xi}$'s are in the same plane, the principal plane. The first term in Eq. (12.17)



corresponds to direct scattering in the atmosphere toward the observer *without* interaction with the surface, while the second term corresponds to photons reflected from the surface and then scattered toward the observer. The second term is negligible when $\Theta_B \lesssim 60^\circ$, i.e., when the surface radiometer is aimed within 60° of the Sun, because of the strong forward scattering by the aerosol. However, there are configurations where the second term is comparable in magnitude with the first. For example, consider $\hat{\xi}_B = \hat{e}_z$, i.e., a surface radiometer viewing the zenith. In that case, $\cos \Theta_B = \cos \theta_0$ and $\cos \Theta_B^{(r)} = -\cos \theta_0$. For large θ_0 the phase functions for both aerosol and Rayleigh scattering are comparable at Θ_B and $\Theta_B^{(r)}$, and r_f can be significant, e.g., for $\theta_0 = 80^\circ$, $r_f \approx 0.35$. As ρ_B depends on the phase function at two angles, Θ_B and $\Theta_B^{(r)}$, it is not at all obvious that sufficiently accurate values of $\omega_a \tau_a P_a(\Theta)$ can be retrieved to effect a meaningful vicarious calibration.

Similarly, the reflectance of the water-atmosphere system is

$$\rho_T(\theta_T^{(c)}) = \frac{\omega_a \tau_a}{4 \cos \theta_T^{(c)} \cos \theta_0} \left[P_a(\Theta_T) + [r_f(\theta_T^{(c)}) + r_f(\theta_0)] P_a(\Theta_T^{(r)}) \right], \quad (12.18)$$

where $\theta_T^{(c)}$ is the *compliment* of θ_T , i.e., $\theta_T^{(c)} = \pi - \theta_T$. Also

$$\cos \Theta_T = \hat{\xi}_0 \bullet \hat{\xi}_T \quad \text{and} \quad \cos \Theta_T^{(r)} = \hat{\xi}_0^r \bullet \hat{\xi}_T,$$

where

$$\begin{aligned} \hat{\xi}_T &= \cos \phi_T \sin \theta_T \hat{e}_x + \sin \phi_T \sin \theta_T \hat{e}_y + \cos \theta_T \hat{e}_z \\ &= \cos \phi_T \sin \theta_T^{(c)} \hat{e}_x + \sin \phi_T \sin \theta_T^{(c)} \hat{e}_y - \cos \theta_T^{(c)} \hat{e}_z. \end{aligned}$$

From Eqs. (12.17) and (12.18) it is easy to see that, while we can still arrange to have $\Theta_T = \Theta_B$, the presence of the Fresnel reflection terms destroys the nice symmetry we had in Eq. (12.14), and the only way we can predict ρ_T from ρ_B is to use ρ_B to find $\omega_a \tau_a P_a(\Theta)$ using Eq. (12.17) and then introduce the result into Eq. (12.18). This requires measuring $\rho_B(\hat{\xi}_B)$ for a sufficient number of directions $\hat{\xi}_B$ to yield a reliable $\omega_a \tau_a P_a(\Theta)$. This is not as straightforward as it might seem: Θ_T corresponds mostly to scattering in the *backward* hemisphere, while Θ_B corresponds mostly to scattering in the *forward* hemisphere. An example of this is shown in Figure 12.3, which provides Θ_T and $\Theta_T^{(r)}$ as a function of $\theta_T^{(c)} = \pi - \theta_T$, for $\theta_0 = 60^\circ$. The curves are for three cases $\phi_T - \phi_0 = 60^\circ, 90^\circ$ and 120° . For most of the viewing angles, $\Theta_T > 90^\circ$. In contrast the $\Theta_T^{(r)}$'s are mostly in the forward hemisphere. Thus, in order to estimate ρ_T , the surface measurements must be carried out so as to incorporate as accurate an estimation of $\omega_a \tau_a P_a(\Theta)$ as possible for $\Theta > 90^\circ$. This generally requires large solar zenith angles.

We want the ρ_B measurements to be made at the same solar zenith angle extant at the satellite overpass. Ideally we also like the sensor to be viewing in the direction where the surface-measurement site is in or near its field of view, so there will be the same atmospheric state for both ρ_T and ρ_B . Remember, the whole idea is to measure ρ_B in various directions and use Eq. (12.17) to estimate $\omega_a \tau_a P_a(\Theta)$ at enough values of Θ to allow Eq. (12.18) to be used to estimate ρ_T at the sensor. This is accomplished by having the surface radiometer scan over various portions of the sky. There are two important kinds of sky scans that we will consider for ρ_B :¹⁴ the *almucantar* (A) scan for which $\theta_B = \theta_0$ and $\phi_B - \phi_0$ is varied through 0 to 360° ; and the *principal plane* (PP) scan for which θ_B is varied from 0 to 90° for *both* $\phi_B - \phi_0 = 0$ and $\phi_B - \phi_0 = 180^\circ$. For the almucantar scan the maximum Θ_B is $2\theta_0$, while for the PP scan it is $\theta_0 + \pi/2$. Figure 12.4 provides an example of $\Theta_B^{(r)}$ as a function of Θ_B for these two kinds of scans with $\theta_0 = 60^\circ$. Note that for a given Θ_B , $\Theta_B^{(r)}$ is always smaller in the almucantar scan than in the principal plane scan. Also, the only way to reach the largest values of Θ_B is with the PP scan, and for these $\Theta_B^{(r)}$ is near Θ_B but $P_a(\Theta_B^{(r)})$ can never be known for these angles because $\Theta_B^{(r)} > \Theta_B$ (except at the maximum Θ_B for which $\Theta_B^{(r)} = \Theta_B$).

¹⁴Cameras with fisheye lenses now exist that can be used to quickly obtain the entire sky radiance (and its polarization), so there is complete choice in what parts of the sky to use in the measurements. The cameras can be mounted on gimbals for operation on a ship.

For an example of recovering $\omega_a \tau_a P_a(\Theta_B)$, we will assume that single scattering is the correct physics (multiple scattering is considered later) and use it to simulate $\rho_B(\hat{\xi}_B)$ for the M99 aerosol model. This aerosol model has the most structure in the phase function in the backward hemisphere of all of the models considered in Chapter 4 (see Figure 12.5, left panel). Then we worked backward to estimate $\omega_a \tau_a P_a(\Theta_B)$ from ρ_B . As in Figures 12.3 and 12.4 the case of $\theta_0 = 60^\circ$ is considered. First, the PP scan was used to estimate $\omega_a \tau_a P_a(\Theta_B)$ over the entire range of available angles $0 \leq \Theta_B \leq 140^\circ$.¹⁵ We note that in this case the retrieval for $\Theta_B > 90^\circ$ requires knowledge of $\omega_a \tau_a P(\Theta_B)$ for scattering angles that are not available ($\Theta_B^{(r)} > 140^\circ$). This necessitated that an assumption regarding $P_a(\Theta)$ be made for the unavailable scattering angles. Here we assumed that $P_a(\Theta) = P_a(140^\circ)$ for $\Theta > 140^\circ$. The result of this exercise is provided in Figure 12.5, which gives the error in the retrieved $\omega_a \tau_a P_a(\Theta)$ as a function of Θ (dashed line for PP scan). The error for $\Theta \lesssim 75^\circ$ is usually quite small ($< 1\%$) because both of the required angles, Θ_B and $\Theta_B^{(r)}$, are in the accessible range, but for larger Θ , where the assumption about the large-angle phase function must be used, there is significantly more error. The error is maximum near 120° because of the large error in P_a near 180° . The error is small again near 140° , because the assumed phase function at 160° is close to correct.

Examination of Figure 12.4 shows that for the whole range $0 \leq \Theta_B \leq 105^\circ$, both of the required angles are accessible to ρ_B in the almucantar scan. This suggests that the almucantar scan should be better at retrieving $\omega_a \tau_a P_a(\Theta_B)$ for the range $75^\circ \leq \Theta_B \leq 105^\circ$ than the PP scan. However, the PP scan is still required for $\Theta_B > 120^\circ$. Figure 12.5 (right panel) shows the error in $\omega_a P_a(\Theta)$ obtained with the PP and A+PP scans. Clearly, the A+PP scan greatly reduces the error between 90° and 129° . The large error at 120° is again due to the large error in $P_a(180^\circ)$, and the small error at 90° is due to the error at 120° . The error in $\omega_a \tau_a P_a(\Theta_B)$ in the region $90^\circ \leq \Theta_B \leq 105^\circ$ is second order in $r_f(\theta_0)$ for the A+PP scans but first order in $r_f(\theta_0)$ for the PP scan. We note that for an aerosol phase function with less structured backscattering, e.g., the T50 model (Figure 12.5, left panel), the retrieval would be much better. Actually, for the T50 aerosol model the largest error for $\Theta \leq 120^\circ$ was 1.9% with this procedure, for $\theta_0 = 60^\circ$.

How do these errors translate to error in the computed ρ_T ? Consider the case in Figure 12.3 (where $\theta_0 = 60^\circ$) with $\phi_T - \phi_0 = 90^\circ$. Figure 12.6 shows the error in the computed value of ρ_T as a function of $\theta_T^{(c)}$ for both estimates of $\omega_a \tau_a P(\Theta_B)$. Clearly the A+PP scan provides a more accurate ρ_T than the PP scan alone. For the A+PP the maximum error in ρ_T is 2.5% for $\theta_T^{(c)} > 15^\circ$. Note that if $\phi_T - \phi_0 = 60^\circ$, then the angles at which the phase function must be known are smaller (Figure 12.3), which should result in an even better prediction of ρ_T .

¹⁵The possible range is really $0 \leq \Theta_B \leq 150^\circ$; however, considering the impossibility of measuring ρ_B exactly at the horizon, we limited the available measurements to 10° above the horizon.

In the above example we have assumed that the molecular scattering contribution has been removed (or there is no molecular scattering). If molecular scattering is included, then $\omega_a \tau_a P_a(\Theta_B) \rightarrow \omega_a \tau_a P_a(\Theta_B) + \tau_r P_r(\Theta_B)$. Because the Rayleigh terms are known precisely (given the surface atmospheric pressure) there is no error in the Rayleigh contribution. The same is true for the computation of ρ_T . In general this means that the error is *reduced* from what we see in Figure 12.6 by a factor $(1 + \rho_M/\rho_a)$, where ρ_M and ρ_a are respectively the molecular and aerosol contributions to ρ_T . However, note that as ρ_M increases, the viability of single scattering decreases.

We see that when single scattering correctly describes the radiative transfer process in the atmosphere, measurements of sky radiance at the bottom of the atmosphere can be used to radiometrically “calibrate” a sensor in orbit with an uncertainty of the order of 2-3%. It seems reasonable to ask; can the single scattering method still work reasonably well in a real atmosphere, where there is some multiple scattering? Starting with values of ρ_B derived from the exact solution to the radiative transfer equation for a two layer model atmosphere with aerosols (M99 with $\tau_a = 0.2$) in the lower layer and molecular scattering (for $\lambda = 865$ nm) in the upper layer, application of the single scattering formula (Eq. (12.17)) yielded a phase function that was in error by as much as a factor of 2 for the important range $80^\circ < \Theta < 130^\circ$, with $\theta_0 = 60^\circ$. So, the answer is no. Even when the aerosol concentration is low ($\tau_a = 0.2$), the fact that the solar zenith angle must be large, and that large values of $\theta_B^{(c)}$ are required to find the phase function in the backward hemisphere, means that we have to be looking in regions of the sky where multiple scattering is most important. Thus, we must consider multiple scattering in order to implement such a scheme for sensor calibration; however, as usual, we can use what we have learned from the single-scattering analysis to guide us.

Fortunately, a method of inverting sky radiance measurements to phase function and single scattering albedo in a multiple scattering atmosphere is available. The method is described in the appendix to this chapter. Applying it to the exact values for ρ_B at 865 nm (i.e., computed including multiple scattering for a two layer model atmosphere having aerosols described by the M99 model, with $\tau_a = 0.2$ in the lower layer, and all the molecular scattering in the upper layer) provided the retrieval of $\omega_a P_a(\Theta)$ shown in Figure 12.7. The same vertical structure was assumed in the retrieval as in the generation of the pseudo data for ρ_B . The correct value of τ_a was used in the retrieval which means that this quantity would have to be measured along with ρ_B . Note that the retrieval is excellent, with the error being approximately the same as when single scattering was assumed to be the correct physics. Figure 12.8 provides the resulting error in the top-of-atmosphere reflectance ρ_T computed by introducing the retrieved $\omega_a P_a(\Theta)$, along with τ_a , into the radiative transfer code that included multiple scattering. To simulate the effect of error in τ_a , the various curves in Figure 12.8 correspond to using values of 0.14, 0.16, 0.18, 0.20, 0.22, 0.24, and 0.26, as an estimate for τ_a in *both the retrieval and prediction codes*. The actual τ_a used

in generation of the pseudo data was 0.2. Noting that measurement of the sky radiance cannot be made looking directly at the sun ($\Theta_B = 0$), there will be a minimum value of Θ_B that can be reached. Here Θ_{Min} is the smallest value of Θ_B obtained in the almucantar scan, and in the figure it is 0.92° . The predicted ρ_T is excellent, with the error being only of the order of 1% when the correct value of τ_a is used, and no more than 2.5% in error when an incorrect value of τ_a of up to 25% is used. The fact that the error is quite insensitive to the value of τ_a used in the retrieval of $\omega_a P_a(\Theta)$ and the prediction of ρ_T actually owes to the property that when the value of τ_a that is used in the retrieval is too large, a smaller value of ω_a will be retrieved which partially compensates for the too-large τ_a , and vice versa. Although we have applied this technique here to $\lambda = 865$ nm, where ρ_w is typically negligible, a similar set of simulations shows that it works just as well at 443 nm, with the proviso that $\rho_w(443)$ is also accurately measured. However, it is envisioned that this method of direct vicarious calibration will be most useful in the NIR, where ρ_w is known to be negligible.

What are the effects of error in the measurement of ρ_B or in the assumed vertical structure of the atmosphere? We investigate the former by introducing a calibration error of $\pm 5\%$ in the radiometer used to measure ρ_B . We added this error to the values of ρ_B in the multiple scattering example above, and used it as pseudo data in the $\omega_a P_a(\Theta)$ -retrieval code and used its output in the ρ_T -prediction code. The result is provided in Figure 12.9 (left panel). This is a direct confirmation that what we discovered at the beginning of this subsection — that calibration error in the surface radiometer for measuring ρ_B directly translates to an equivalent error in the prediction of ρ_T . Here, unless the ρ_B error smaller than about 1-2% it will dominate the overall error in ρ_T .

An example of the effect of an incorrect assumption regarding the vertical structure of the atmosphere in the $\omega_a P_a(\Theta)$ -retrieval algorithm is provided in Figure 12.9. In this example, we prepared pseudo data for an atmosphere having an M99 aerosol in a 1 km thick layer above the surface, uniformly mixed with air, and a second layer containing a different aerosol (T50) and the remaining molecular scattering. The retrieval and prediction codes assumed all of the aerosol to be in a thin layer just above the surface and used the correct τ_a . The resulting error in ρ_T is shown in the right panel of Figure 12.9, and suggests that for this aerosol optical thickness (0.2) and wavelength (865 nm) that uncertainty introduced by an incorrect vertical structure is really not particularly significant.

There are many other assumptions made thus far that could affect the accuracy of this method of vicarious calibration. These include (1) the assumption of a flat water surface; (2) the use of a plane parallel atmosphere in the radiative transfer code rather than a spherical shell atmosphere, which would be more appropriate for situations with the required large values of θ_0 ; (3) the neglect of polarization in the radiative transfer code; and (4) the inability to make ρ_B measurements close to the Sun, i.e., in the solar aureole,

which results in a larger Θ_{Min} . These have all been studied through sensitivity analyses (references given in the Bibliographic Notes). The most important of these appears to be the neglect of polarization; however, this only increases the error by 1-2%. If needed, a multiple scattering inversion algorithm for retrieving the most important parts of the phase matrix $\omega_a \mathbf{P}_a(\Theta)$ is available, and a prediction of ρ_T in a manner similar to that above using vector radiative transfer can be effected with excellent accuracy. Operation of this inversion algorithm does, however, require measurement of the linear polarization components of the Stokes vector associated with ρ_B .

In summary, the direct technique of vicarious calibration appears to be a robust approach to estimating the top-of-atmosphere reflectance based on surface measurements of sky radiance (or normalized sky radiance).¹⁶ It is intended to be used in the NIR, where measurement of ρ_w is unnecessary. Because the direct term in ρ_T is determined by atmospheric scattering at Θ 's in the backward hemisphere, the surface measurements must be made at large values of θ_0 , which requires high latitude in summer and/or mid latitude in winter. Ideally, one would want the satellite sensor to be viewing the same portion of the atmosphere as the surface radiometer. Although it might seem difficult to arrange this, all that is required is that the surface measurement site be visible in the satellite image for a particular overpass. One could use stations on small islands and simply make measurements whenever the sensor can view the station. Probably the most difficult aspect of utilizing this method is having a clear (cloud-free) sky for PP and A+PP scans, coincident with an appropriate satellite overpass.

12.3.2 Full System Calibration

Some of the difficulties mentioned above can be circumvented by an alternate method of vicarious calibration, which we refer to as *system* calibration. Pragmatically, in this method, simultaneous measurements of $\rho_w(\lambda)$ and $\rho'_t(\lambda)$, the “uncalibrated” $\rho_t(\lambda)$, are carried out. In effect, ρ'_t is then inserted into the atmospheric correction algorithm and the sensor calibration coefficients, $K(\lambda, \dots)$, are varied (which in turn vary the value of ρ'_t) until agreement between the sensor-retrieved ρ_w and the surface-measured ρ_w is achieved. Simply speaking, the sensor calibration is adjusted so that the sensor *plus* algorithms provide the correct answer. This section describes how this can be done in a systematic manner.

As before, we start by assuming that single scattering is the correct physics for the atmospheric radiative transfer. We let, as in Eq. (12.1), the measured reflectance (including

¹⁶It is possible to calibrate a surface radiometer directly to reflectance, i.e., ρ_B , without having to introduce F_0 from external sources.

the calibration error) ρ'_t be related to the true reflectance ρ_t through

$$\rho'_t(\lambda_i) = \rho_t(\lambda_i)(1 + \alpha(\lambda_i)),$$

where the calibration error α is to be determined. In single-scattering physics,

$$\rho'_t(\lambda_i) = \rho_r(\lambda_i) + \rho_{as}(\lambda_i) + t(\lambda_i)\rho_w(\lambda_i).$$

Because $\rho_w(\lambda_i)$ is measured, we can compute the apparent value of $\varepsilon(\lambda_i, \lambda_l)$, i.e., when using ρ'_t instead of ρ_t :

$$\varepsilon'(\lambda_i, \lambda_l) = \frac{\rho_{as}(\lambda_i)}{\rho_{as}(\lambda_l)} = \frac{\rho'_t(\lambda_i) - \rho_r(\lambda_i) - t(\lambda_i)\rho_w(\lambda_i)}{\rho'_t(\lambda_l) - \rho_r(\lambda_l)}. \quad (12.19)$$

Inserting ρ'_t gives

$$\varepsilon'(\lambda_i, \lambda_l) = \frac{\alpha(\lambda_i)\rho_t(\lambda_i) + \rho_{as}(\lambda_i)}{\alpha(\lambda_l)\rho_t(\lambda_l) + \rho_{as}(\lambda_l)}. \quad (12.20)$$

If the radiometer is perfectly calibrated, the computed value, ε' , is equal to the true value, ε ; however, if the errors (α 's) are sufficiently large, ε' could be significantly different from ε , and likely may not even make *physical* sense. A numerical example may be helpful to understand this. Consider the data presented in Figure 10.4 in Chapter 10. Focus on the example from the station represented by the right pointing solid triangle and take $\lambda_l = 865$ nm. For this $\rho_t(412) \approx 0.2$, $\rho_t(865) \approx 0.012$, and $\rho_a(412) \approx \rho_a(865) \approx 0.005$. Approximate ρ_{as} by ρ_a . Then the true value of $\varepsilon(412, 865) \approx 1$, while the computed value of $\varepsilon'(412, 865)$ is dependent on $\alpha(412)$ and $\alpha(865)$:

$$\varepsilon'(412, 865) = \frac{0.20\alpha(412) + 0.005}{0.012\alpha(865) + 0.005}.$$

If $\alpha(412) = +0.05$ and $\alpha(865) = +0.05$, then $\varepsilon'(412, 865) \approx 2.7$, while if $\alpha(412) = -0.05$ and $\alpha(865) = +0.05$, then $\varepsilon'(412, 865) \approx -0.9$, which is most unlikely. Considering the location from which the satellite data were taken, both of these values would be completely inconsistent with the expectation of $\varepsilon \approx 1$, and would indicate a significant calibration error. Indeed, unrealistic ε values are a clear indicator of calibration error. Note also that the clearer the atmosphere, the more dramatic the ε behavior would be, e.g., as $\rho_{as}(\lambda_i) \rightarrow 0$,

$$\varepsilon'(\lambda_i, \lambda_l) \rightarrow \frac{\alpha(\lambda_i)[\rho_r(\lambda_i) + t\rho_w(\lambda_i)]}{\alpha(\lambda_l)\rho_r(\lambda_l)} \approx \frac{\alpha(\lambda_i)\rho_r(\lambda_i)}{\alpha(\lambda_l)\rho_r(\lambda_l)} \approx \frac{\alpha(\lambda_i)}{\alpha(\lambda_l)} \times \left[\frac{\lambda_l}{\lambda_i} \right]^4,$$

and ε' is *totally* determined by the calibration error.¹⁷ Thus, the clearer the atmosphere, the more evident the calibration error.

¹⁷Perhaps it is better to refer to the α 's as calibration *biases* rather than calibration errors. They are a *systematic* bias in the correct calibration slope $K(\lambda_i, \dots)$.

Let's assume that we know the correct value of $\varepsilon(\lambda_i, \lambda_l)$. Then we could adjust the sensor calibration, so that the *correct* value of $\varepsilon'(\lambda_i, \lambda_l)$ is obtained in Eq. (12.19). To effect this, multiply $\rho'_t(\lambda_i)$ by a factor $1 + \gamma(\lambda_i)$, so

$$\left(1 + \gamma(\lambda_i)\right)\rho'_t(\lambda_i) = \left(1 + \gamma(\lambda_i)\right)\left(1 + \alpha(\lambda_i)\right)\rho_t(\lambda_i) \approx \left(1 + \gamma(\lambda_i) + \alpha(\lambda_i)\right)\rho_t(\lambda_i),$$

where we have assumed $\gamma(\lambda_i)$ will be $\ll 1$, and have therefore ignored the product of $\gamma(\lambda_i)$ and $\alpha(\lambda_i)$ compared to the other terms. Note, that the fractional error in $\rho'_t(\lambda_i)$ is $\alpha(\lambda_i)$, but the fractional error in $\left(1 + \gamma(\lambda_i)\right)\rho'_t(\lambda_i)$ is $\gamma(\lambda_i) + \alpha(\lambda_i)$. Now, adjust $\gamma(\lambda_i)$ to bring ε' into agreement with the true value ε . For simplicity take $\gamma(\lambda_l) = 0$. This gives

$$\varepsilon(\lambda_i, \lambda_l) = \frac{\left(\gamma(\lambda_i) + \alpha(\lambda_i)\right)\rho_t(\lambda_i) + \rho_{as}(\lambda_i)}{\alpha(\lambda_l)\rho_t(\lambda_l) + \rho_{as}(\lambda_l)}, \quad (12.21)$$

and solving for $\gamma(\lambda_i)$:

$$\gamma(\lambda_i) + \alpha(\lambda_i) = \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\frac{\rho_t(\lambda_l)}{\rho_t(\lambda_i)}. \quad (12.22)$$

This tells us that if we know the true value of $\varepsilon(\lambda_i, \lambda_l)$ and if there is no calibration error at λ_l , i.e., $\alpha(\lambda_l) = 0$, then¹⁸

$$\left(1 + \gamma(\lambda_i)\right)\rho'_t(\lambda_i) = \rho_t(\lambda_i),$$

and the calibration error $\alpha(\lambda_i)$ has been completely removed. If $\alpha(\lambda_l) \neq 0$, the residual fractional error in $\rho_t(\lambda)$ is $\varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l)/\rho_t(\lambda_i)$. For the numerical example above,

$$\gamma(412) + \alpha(412) = 0.06\alpha(865),$$

and a 10% calibration error at 865 nm could be reduced to a 0.6% error in the total reflectance at 412 nm. There is of course no magic here, the improvement is simply due to the large Rayleigh component of $\rho_t(412)$ that can be computed with little or no error. A similar computation at 765 nm, where the Rayleigh component is significantly smaller, shows that a 10% calibration error at 865 nm is only reduced to about a 7% error at 765 nm.

What if we use this procedure, but have error in the $\varepsilon(\lambda_i, \lambda_l)$ estimation, i.e., we adjust $\gamma(\lambda_i)$ until $\varepsilon'(\lambda_i, \lambda_l)$ is replaced by $\varepsilon(\lambda_i, \lambda_l) + \Delta\varepsilon(\lambda_i, \lambda_l)$? A similar computation shows that in this case Eq. (12.22) is replaced by

$$\gamma(\lambda_i) + \alpha(\lambda_i) = \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\frac{\rho_t(\lambda_l)}{\rho_t(\lambda_i)} + \Delta\varepsilon(\lambda_i, \lambda_l)\left(\alpha(\lambda_l)\frac{\rho_t(\lambda_l)}{\rho_t(\lambda_i)} + \frac{\rho_{as}(\lambda_l)}{\rho_t(\lambda_i)}\right)$$

¹⁸Note: if we do not know the true value of $\varepsilon(\lambda_i, \lambda_l)$ then the $\rho_{as}(\lambda_i)$ and $\rho_{as}(\lambda_l)$ terms will not sum to zero.

and

$$(1 + \gamma(\lambda_i))\rho'_t(\lambda) = \rho_t(\lambda) + \varepsilon(\lambda_i, \lambda_l)\alpha(\lambda_l)\rho_t(\lambda_l) + \Delta\varepsilon(\lambda_i, \lambda_l)(\alpha(\lambda_l)\rho_t(\lambda_l) + \rho_{as}(\lambda_l)).$$

Again using the data from Figure 10.4, we find the relative error at 412 nm to be

$$\gamma(412) + \alpha(412) = 0.06\alpha(865) + \Delta\varepsilon(412, 865)(0.06\alpha(865) + 0.025).$$

Let's see what the result is if we make a terrible guess at ε . Consider Figure 10.5 in Chapter 10. Assume that the actual aerosol model is M98, but we incorrectly choose M50. The figure shows that $\Delta\varepsilon(412, 865) \approx 0.4$. Then for a 10% error at 865 nm, the residual error in the estimate of $\rho_t(412)$ would be about 1.85%, with two-thirds of the error due to the improper choice of $\Delta\varepsilon(412, 865)$. The error is still small, but this underscores the importance of making an informed choice of the aerosol model for the calibration site.

How can we estimate $\varepsilon(\lambda_i, \lambda_l)$ over and above just assuming it is unity because the procedure is carried out in a maritime environment? We cannot do it from the sensor itself as that would require a calibrated sensor, i.e., the goal of vicarious calibration. Therefore some surface measurements relating to $\varepsilon(\lambda_i, \lambda_l)$ are required. One possibility is provided in Figure 12.10, which shows that for a given geometry there is a close relationship between $\varepsilon(\lambda_i, \lambda_l)$ and $\tau_a(\lambda_i)/\tau_a(\lambda_l)$ (here, for $\lambda_i = 443$ nm and $\lambda_l = 865$ nm) for the various aerosol models used in Chapter 10. This suggests that, when geometry is taken into account, surface measurements of the spectral variation of the aerosol optical depth could be used to estimate $\varepsilon(\lambda_i, \lambda_l)$. Another method would be to use surface measurements to select a suitable aerosol model to use to provide an estimate for $\varepsilon(\lambda_i, \lambda_l)$.

We have seen here that the vicarious *system* calibration is effective and very straightforward in the single-scattering approximation, but can it be implemented in a real setting, i.e., in a multiple-scattering atmosphere? In what follows we provide a scheme for accomplishing this.

We will assume that the band at λ_l is calibrated correctly, i.e., $\rho'_t(\lambda_l) = \rho_t(\lambda_l)$. We also assume that in addition to $\rho_w(\lambda_i)$, atmospheric measurements are made that allow the determination of the most appropriate aerosol model among the candidate models. Recall from Chapter 10 that the atmospheric correction algorithm utilizes look up tables relating the following quantities:

$$\rho_A^{(j)}(\lambda) = \rho_A^{(j)}\{\rho_{as}^{(j)}(\lambda)\}, \quad \text{or the inverse} \quad \rho_{as}^{(j)}(\lambda) = \rho_{as}^{(j)}\{\rho_A^{(j)}(\lambda)\},$$

where the notation $Y = Y\{X\}$ means Y as a function of X , i.e., given X this relationship determines Y , and the superscript “ j ” is the index of the aerosol model, e.g., M50 ... T99. Thus, given $\rho_A(\lambda_l) = \rho_t(\lambda_l) - \rho_r(\lambda_l)$ and an aerosol model, we can find $\rho_{as}(\lambda_l)$, which

yields $\tau_a(\lambda_l)$. The model then also provides $\tau_a(\lambda_i)$, $\rho_{as}(\lambda_i)$, $t(\lambda_i)$ and $\rho_A(\lambda_i)$. Finally, $\rho_t(\lambda_i) = \rho_r(\lambda_i) + \rho_A(\lambda_i) + t(\lambda_i)\rho_w(\lambda_i)$. This progression can be summarized through

$$\begin{aligned} \rho_t(\lambda_l) &\xrightarrow{\rho_r(\lambda_l)} \rho_A(\lambda_l) \xrightarrow{\text{Model } j} \rho_{as}^{(j)}(\lambda_l) \xrightarrow{\text{Model } j} \tau_a^{(j)}(\lambda_l) \\ &\xrightarrow{\text{Model } j} \tau_a^{(j)}(\lambda_i) \xrightarrow{\text{Model } j} \rho_{as}^{(j)}(\lambda_i) \xrightarrow{\text{Model } j} \rho_A^{(j)}(\lambda_i) \end{aligned}$$

followed by

$$\rho_A^{(j)}(\lambda_i) \xrightarrow{t^{(j)}(\lambda_i)\rho_w(\lambda_i)+\rho_r(\lambda_i)} \rho_t^{(j)}(\lambda_i) \xrightarrow{\rho_t'(\lambda_i)} \gamma^{(j)}(\lambda_i),$$

where the superscript “ j ” is meant to remind the reader that the resulting quantity depends on the aerosol model. In the last step, the sensor calibration is varied to bring $\rho_t'(\lambda_i)$ into agreement with $\rho_t(\lambda_i)$. Upon reflection, one sees that this procedure is identical to (1) choosing an aerosol model, (2) applying the multiple-scattering atmospheric correction algorithm using this model, and (3) varying the sensor calibration to bring the retrieved $\rho_w(\lambda_i)$ into agreement with the measured value. The critical ingredients are the choice of the appropriate aerosol model and the measurement of the water-leaving reflectance. With this procedure, one can expect an improvement in calibration accuracy similar to that suggested by Eq. (12.22).

In principle this vicarious calibration procedure need only be applied once; however, in practice neither the aerosol model nor the ρ_w -measurement are error free. So in general one must make such measurements many times utilizing various Sun-viewing geometries. The ideal measurement site would be one with a maritime aerosol of low optical depth (to minimize the aerosol model error) and water with very low ρ_w (to minimize the ρ_w error). In addition, the site must be of low enough latitude to provide a range of Sun angles. In the case of the NIR, these requirements are easily met in oceanic central gyres, which are far from land.¹⁹ However, in the visible, these requirements are basically incompatible, so the requirement of low ρ_w is sacrificed in favor of the others; however, it is replaced by a requirement that the reflectance be relatively stable. The site chosen for calibration of U.S. color sensors in the visible is in the Hawaiian Islands off the island of Lanai. The ρ_w measurements are made continuously from a moored buoy and atmospheric measurements are made from the island. Vicarious calibration is effected whenever possible (given that the area was sufficiently free of cloud etc.) and the resulting K 's averaged. The details of the actual vicarious calibration of SeaWiFS are provided in Chapter 13 and papers referenced in the Bibliographic notes.

Comparing the two methods of vicarious calibration described this section, one notes that in the first, the ultimate accuracy is the accuracy of the surface radiometer used to measure the sky radiance, while in the second it is basically the accuracy in the chosen

¹⁹Note that there is no reason that the vicarious calibration of the visible and NIR have to be performed at the same location or time. In practice they are not.

aerosol model to represent the actual conditions at the measurement site. If measurement of ρ_w is required, as in the full system calibration method (in the visible), the error in ρ_w is not significant as long as it is not too large. For example if the error in $\rho_w(443)$ is 5%, it's contribution to the error in $\rho_t(443)$ will only be 0.5%. Contrast this with a 5% error in the sky radiance in the direct method which results in about a 5% error in ρ_t . The direct method is the only one available for calibration of the band at λ_l . However, it turns out that reduction of the error in this band has little effect on the accuracy of the retrieved ρ_w in the visible.²⁰

12.4 Characterization

The design and fabrication of all ocean color sensors is a sequence of compromises and trade offs. As such, no completed sensor behaves as envisaged. All instruments have idiosyncrasies that can only be addressed and corrected in orbit through a thorough understanding of their performance prior to launch. Realization of such understanding is termed instrument *characterization*. Here we provide three examples of instrument “misbehavior,” i.e., departure from the ideal, that must be understood prior to launch in order to effect an in-orbit correction: sensitivity of the instrument to the polarization state of the incident radiance; response of the instrument to radiance with wavelength significantly outside the nominal spectral bands; and light propagating within the instrument where it should not — stray light. Future instruments with newer designs will likely have similar, and new, idiosyncrasies.

12.4.1 Polarization Sensitivity

Most sophisticated radiometers have some sensitivity to the polarization state of the radiance they are intended to measure. Ocean color sensors are no exception. They all contain optical components such as mirrors, beamsplitters, diffraction gratings, interference filters, etc., that respond in a manner that depends on the polarization state of the radiant power. Corrections can be made for this polarization sensitivity, but only if the characteristics of the sensitivity is known, i.e., if it is characterized.

How do we specify the polarization sensitivity of an ocean color sensor? We note that

²⁰This is an important fact because, along with the technique of vicarious calibration, it shows that it is not productive to spend valuable resources to improve prelaunch radiometric calibration, which is very expensive to start with. At most prelaunch calibration need only be sufficiently accurate to allow setting of the sensor's saturation radiances (the radiance corresponding to the sensor's maximum number of digital counts for each spectral band), which does not require great precision.

the radiance from the atmosphere-ocean system into the sensor will possess some degree of linear polarization but negligible elliptical polarization. Thus, we need to determine only the linear polarization sensitivity of the instrument. As such, we introduce linearly polarized and monochromatic (λ) light into the instrument, i.e., the Stokes vector of the incident radiance is given by

$$\mathbf{I} = \begin{pmatrix} I \\ I \cos 2\chi \\ I \sin 2\chi \\ 0 \end{pmatrix}.$$

where \mathbf{I} propagates normal to the aperture of the instrument, χ is the angle with respect to a fixed direction in the plane of the aperture, and I is the radiance that would be measured by a sensor in the absence of polarization sensitivity. The response of the instrument can be described by providing the *measured* Stokes vector \mathbf{I}_m through the sensor's Mueller matrix \mathbf{M} : $\mathbf{I}_m = \mathbf{M}\mathbf{I}$. Because we are only interested in I given \mathbf{I}_m , we need only measure the top row of the Mueller matrix. Thus,

$$I_m = (M_{11} + M_{12} \cos 2\chi + M_{13} \sin 2\chi)I.$$

Now, let $m_{12} = M_{12}/M_{11}$ and $m_{13} = M_{13}/M_{11}$, and choose a and δ such that $m_{12} = a \cos 2\delta$ and $m_{13} = a \sin 2\delta$. Then

$$I_m = M_{11}I(1 + a \cos 2(\chi - \delta)).$$

Now, as the angle χ is varied from 0 to 2π , I_m will vary sinusoidal with a period of π . At its maximum value $\delta = \chi_{\text{Max}}$ and at its minimum value $\delta = \chi_{\text{Min}} \pm \pi/2$. Thus,

$$M_{11}I = \frac{1}{2}(I_{m_{\text{Max}}} + I_{m_{\text{Min}}}),$$

and

$$a = \frac{1}{2M_{11}I}(I_{m_{\text{Max}}} - I_{m_{\text{Min}}}).$$

This sequence of measurements determines a and δ or equivalently, m_{12} and m_{13} for the instrument and the Mueller matrix becomes

$$\mathbf{M} = M_{11} \begin{pmatrix} 1 & a \cos 2\delta & a \sin 2\delta & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{pmatrix},$$

where the entries indicated by \bullet are irrelevant for our purposes.²¹ Now, when partially polarized radiance enters the remote sensor, how does the sensor respond? Let the Stokes

²¹If the instrument is calibrated using *unpolarized* radiance, it will give the correct value of the radiance when the source is unpolarized. This means that the value of M_{11} must be unity.

vector of the incident radiance be written

$$\mathbf{I}_t = \begin{pmatrix} I_t \\ Q_t \\ U_t \\ 0 \end{pmatrix} = \begin{pmatrix} I_t(1 - P_t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} I_t P_t \\ Q_t \\ U_t \\ 0 \end{pmatrix},$$

where P_t is the degree of polarization: $P_t = \sqrt{Q_t^2 + U_t^2 + V_t^2}/I_t$. Now, for the polarized component we can write $Q_t = P_t I_t \cos \chi_t$ and $U_t = P_t I_t \sin \chi_t$.²² The radiance measured by the sensor is $\mathbf{I}_m = \mathbf{M}\mathbf{I}_t$, so

$$I_m = M_{11}I_t + M_{12}Q_t + M_{13}U_t.$$

Substituting for the M 's and Q_t and U_t , we find

$$I_m = M_{11}I_t((1 + aP_t \cos \delta \cos \chi_t + aP_t \sin \delta \sin \chi_t) = M_{11}I_t(1 + aP_t \cos(\delta - \chi_t)).$$

Since $M_{11} = 1$, this shows that, if the polarization sensitivity of the instrument is ignored, the error in the measured I_t , i.e., I_m could be as much as $\pm aP_t$.

How large is this error for a real sensor? Consider the MODIS sensor on the satellite TERRA. Figure 12.11 shows the prelaunch-measured polarization amplitude a as a function of the angle of incidence of the incident radiance on the scan mirror. At the center of the MODIS scan, the angle of incidence is 37.5° . The measurements are shown for two spectral bands: 412 and 865 nm. For each band there are two curves, which correspond to the two sides of the scan mirror.²³ So, if $P=1$, i.e., the radiance is 100% polarized, the error in the measured radiance at 412 nm could be as much as 6%. In reality the radiance is never fully polarized. Examples in Chapter 9 show that P can reach values as high as 60–70% at large viewing angles. This means that the maximum error could be as much as $0.7 \times 0.06 = 0.042$, or little over 4%. This translates to an error of the order of 40% in the retrieved value of ρ_w (in the blue). Clearly, one must take into account this polarization sensitivity to have any hope of retrieving accurate values of ρ_w . Given the polarization-sensitivity characteristics of the instrument, we still need to know P_t (and χ_t) in order to make a correction for the

²²One must be cautious here. The angle χ_t must be referenced to the same coordinate system as the angle δ . This is not always the case; however, here, if the scan plane is the reference for the measurement of a and δ , i.e., \hat{r} is normal to the scan plane and $\hat{\ell}$ is in the scan plane, then the reference for the two systems will be the same as long as the sensor scans in a plane normal to the water surface (such as MODIS). This is because the propagation direction $\hat{\xi}_t$ must be in the scan plane and the standard reference system for atmospheric radiative transfer when polarization is included is the plane containing $\hat{\xi}_t$ and the vertical. For a sensor that scans in a tilted plane (such as CZCS and SeaWiFS) a rotation of the reference plane must be effected in order to bring the two systems into confluence. This is accomplished using the rotation matrix $\mathbf{R}(\alpha)$ discussed in Chapter 1.

²³MODIS is a paddle-wheel scanner. The scan mirror's rotation axis is parallel to the satellite velocity vector, and both sides of the scan mirror are used.

polarization. This is usually effected by assuming that the aerosol contribution ρ_A and the water contribution $t\rho_w$ are totally unpolarized, so the polarization of \mathbf{I}_t is totally due to Rayleigh scattering by the air. Then we can write various Stokes vectors as

$$\begin{aligned}\mathbf{I}_t &= \mathbf{I}_r + \mathbf{I}_A + t\mathbf{I}_w \\ &= \begin{pmatrix} I_r \\ Q_r \\ U_r \end{pmatrix} + \begin{pmatrix} I_A + tI_w \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} I_r \\ Q_r \\ U_r \\ 0 \end{pmatrix} + \begin{pmatrix} I_t - I_r \\ 0 \\ 0 \\ 0 \end{pmatrix},\end{aligned}$$

where the I_t in $I_t - I_r$ on the right is the *uncorrected* value of I_t . In this manner a sufficiently accurate correction for polarization sensitivity can be effected.

12.4.2 Out-of-Band Response

It is difficult to build a radiometer as complex as an ocean color scanner with spectral bands that have large sensitivity within the spectral band and zero outside, i.e., $S_B = 1$ for λ within the band and $S_B = 0$ outside the band. For example, Figure 12.12 shows $S_B(\lambda)$ for the SeaWiFS band at 865 nm (Band 8). The response of the sensor to radiation outside the principal transmittance region (845-885 nm for Band 8) is called the *out-of-band response*. What is the significance of such out-of-band response? First, let's start with our basic equation for the radiance at the sensor,

$$L_t(\lambda) = L_r(\lambda) + L_A(\lambda) + t(\lambda)L_w(\lambda),$$

and form the band-averaged radiance for band i

$$\langle L_t(\lambda) \rangle_{S_i} = \langle L_r(\lambda) \rangle_{S_i} + \langle L_A(\lambda) \rangle_{S_i} + \langle t(\lambda)L_w(\lambda) \rangle_{S_i}, \quad (12.23)$$

where $\langle L(\lambda) \rangle_{S_i} = \int L(\lambda)S_i(\lambda) d\lambda$.²⁴ How important is the distinction between $L_t(\lambda_B)$ and $\langle L_t(\lambda) \rangle_{S_i}$, where λ_B is the nominal center of Band i ? Consider Band 8 for SeaWiFS shown in Figure 12.12. In the case of L_r , which varies with wavelength approximately as $F_0(\lambda)\lambda^{-4}$, direct computation shows that approximately 9% of the contribution to $\langle L_r(\lambda) \rangle_{S_8}$ comes from $\lambda < 600$ nm! In contrast for $L_A(\lambda)$ which varies approximately according to $F_0(\lambda)$ less than 2% comes from outside the band, and for a spectrally flat source only 0.7% comes from $\lambda < 600$ nm. Thus, the out-of-band effect is significantly different for each term in Eq. (12.23).

²⁴Note that as usual L_A stands for $L_a + L_{ra}$, and S_i is the *normalized* sensitivity: $\int S_i d\lambda = 1$.

How do we deal with this problem? A straightforward method is to appropriately modify the computation of the various terms in Eq. (12.23). For example, in the single scattering approximation

$$L_r(\lambda_i) = \frac{F_0(\lambda_i)\tau_r(\lambda_i)}{4\pi|\cos\theta_v|}(1 + \cos^2\Theta)\left(1 + r_f(\theta_v^{(c)}) + r_f(\theta_0)\right) \equiv F_0(\lambda_i)\tau_r(\lambda_i)G(\theta_0, \theta_v^{(c)}, \Theta), \quad (12.24)$$

where G is a wavelength-independent geometrical factor. Then computing the band averages,

$$\langle L_r(\lambda) \rangle_{S_i} = \langle F_0(\lambda)\tau_r(\lambda) \rangle_{S_i} G(\theta_0, \theta_v^{(c)}, \Theta).$$

Now introduce a new notation. Let

$$\langle \bullet \rangle_X \equiv \frac{\int \bullet X(\lambda) d\lambda}{\int X(\lambda) d\lambda},$$

where \bullet is *any* wavelength-dependent quantity. Then,

$$\begin{aligned} \langle F_0(\lambda)\tau_r(\lambda) \rangle_{S_i} &= \int F_0(\lambda)\tau_r(\lambda)S_i(\lambda) d\lambda \\ &= \frac{\int F_0(\lambda)\tau_r(\lambda)S_i(\lambda) d\lambda}{\int F_0(\lambda)S_i(\lambda) d\lambda} \int F_0(\lambda)S_i(\lambda) d\lambda \\ &= \langle \tau_r(\lambda) \rangle_{F_0S_i} \langle F_0(\lambda) \rangle_{S_i}, \end{aligned}$$

so

$$\langle L_r(\lambda) \rangle_{S_i} = \langle \tau_r(\lambda) \rangle_{F_0S_i} \langle F_0(\lambda) \rangle_{S_i} G(\theta_0, \theta_v^{(c)}, \Theta). \quad (12.25)$$

Thus, to calculate $\langle L_r(\lambda) \rangle_{S_i}$, use the same formula (Eq. (12.24)) but replace $\tau_r(\lambda_i)$ by $\tau_r(\lambda)$ averaged with the weighting function $F_0(\lambda)S_i(\lambda)$ and replace $F_0(\lambda_i)$ by $F_0(\lambda)$ averaged with the weighting function $S_i(\lambda)$. Examples of the difference between the original and averaged values of these quantities is provided in Table 12.1 for SeaWiFS. In the case of Band 8 the difference between $\langle \tau_r(\lambda) \rangle_{F_0S_8} \langle F_0(\lambda) \rangle_{S_8}$ and $F_0(\lambda_8)\tau_r(\lambda_8)$ is approximately 10%. But, this is for single scattering. What about multiple scattering? The procedure for multiple scattering is similar: use $\langle \tau_r(\lambda) \rangle_{F_0S_i}$ in the radiative transfer equation to compute the ratio $[L_r/F_0]$ and, in analogy to the single-scattering procedure, then form $\langle L_r(\lambda) \rangle_{S_i} = [L_r/F_0] \langle F_0(\lambda) \rangle_{S_i}$. This provides the correct $\langle L_r(\lambda) \rangle_{S_i}$ with an error no greater than 0.15%.

Now, we prefer to work with reflectance rather than radiance, so let's see how to band average the reflectance. Noting that $L(\lambda) = \rho(\lambda)F_0(\lambda)\cos\theta_0/\pi$,

$$\langle L(\lambda) \rangle_{S_i} = \frac{\cos\theta_0}{\pi} \langle \rho(\lambda)F_0(\lambda) \rangle_{S_i} = \frac{\cos\theta_0}{\pi} \langle F_0(\lambda) \rangle_{S_i} \langle \rho(\lambda) \rangle_{F_0S_i},$$

Table 12.1: Quantities needed to compute $\langle L_r(\lambda) \rangle_{S_i}$ and $L_r(\lambda_i)$ for the SeaWiFS bands.

Band (<i>i</i>)	λ_i nm	$\langle \tau_r(\lambda) \rangle_{F_0 S_i}$	$\tau_r(\lambda_i)$	$\langle F_0(\lambda) \rangle_{S_i}$ mW/cm ² μm sr	$F_0(\lambda_i)$ mW/cm ² μm sr
1	412	0.3132	0.3185	170.79	180.80
2	443	0.2336	0.2361	189.45	194.95
3	490	0.1547	0.1560	193.66	198.85
4	510	0.1330	0.1324	188.35	193.65
5	555	0.0947	0.0938	185.33	190.25
6	670	0.0446	0.0436	153.41	153.50
7	765	0.0256	0.0255	122.24	122.40
8	865	0.0169	0.0155	98.82	97.10

so

$$\frac{\pi}{\cos \theta_0} \frac{\langle L(\lambda) \rangle_{S_i}}{\langle F_0(\lambda) \rangle_{S_i}} = \langle \rho(\lambda) \rangle_{F_0 S_i},$$

and dividing $\langle L(\lambda) \rangle_{S_i}$ by $\langle F_0(\lambda) \rangle_{S_i}$ automatically yields a quantity proportional to $\langle \rho(\lambda) \rangle_{F_0 S_i}$. In the case of Rayleigh scattering, Eq. (12.25) shows that

$$\langle \rho_r(\lambda_i) \rangle_{F_0 S_i} = \frac{\cos \theta_0}{\pi} \langle \tau_r(\lambda_i) \rangle_{F_0 S_i} G(\theta_0, \theta_v^{(c)}, \Theta), \quad (12.26)$$

and, as Eq. (12.25) showed earlier, the relevant parameter is $\langle \tau_r(\lambda_i) \rangle_{F_0 S_i}$.

As for the other terms in Eq. (12.23), when in doubt as to what to do use single scattering as your guide (as we *always* do in these notes). For example, what do we do with $\langle L_A \rangle_{S_i}$? In single scattering, $L_A = L_{as}$ or $\rho_A = \rho_{as}$. At the core of the atmospheric correction algorithms is the spectral variation of the normalized single-scattered aerosol reflectance, i.e.,

$$\rho_{as}(\lambda_i) = \varepsilon(\lambda_i, \lambda_l) \rho_{as}(\lambda_l),$$

or, effecting the band averaging,²⁵

$$\langle \rho_{as}(\lambda) \rangle_{F_0 S_i} = \langle \varepsilon(\lambda, \lambda_l) \rangle_{F_0 S_i} \rho_{as}(\lambda_l).$$

This equation also applies to the band at λ_l , i.e.,

$$\langle \rho_{as}(\lambda) \rangle_{F_0 S_l} = \langle \varepsilon(\lambda, \lambda_l) \rangle_{F_0 S_l} \rho_{as}(\lambda_l),$$

²⁵Remember that λ_l is a fixed constant in the above equation.

so, eliminating $\rho_{as}(\lambda_l)$ yields

$$\langle \rho_{as}(\lambda) \rangle_{F_0 S_i} = \frac{\langle \varepsilon(\lambda, \lambda_l) \rangle_{F_0 S_i}}{\langle \varepsilon(\lambda, \lambda_l) \rangle_{F_0 S_l}} \langle \rho_{as}(\lambda) \rangle_{F_0 S_l}.$$

We can use the approximate values for ε (Chapter 10) to estimate the magnitude of the effect of band averaging, i.e.,

$$\varepsilon(\lambda_i, \lambda_l) \approx \exp[k(\lambda_l - \lambda_i)]. \quad (12.27)$$

The limits on k over the range 412 to 865 nm for the candidate models in Chapter 10 were $0 \lesssim k \lesssim 1.9 \times 10^{-3} \text{ nm}^{-1}$. Using $k = 2 \times 10^{-3} \text{ nm}^{-1}$ (the largest spectral variation), Table 12.2 provides $\langle \varepsilon(\lambda, 865) \rangle_{F_0 S_i}$ and $\varepsilon(\lambda_i, 865)$ for SeaWiFS Band i , and their % difference. We

Table 12.2: $\langle \varepsilon(\lambda, 865) \rangle_{F_0 S_i}$, $\varepsilon(\lambda_i, 865)$, and their % difference for $k = 2 \times 10^{-3} \text{ nm}^{-1}$.

Band	$\langle \varepsilon(\lambda, 865) \rangle_{F_0 S_i}$	$\varepsilon(\lambda_i, 865)$	% Diff.
1	2.4645	2.4744	-0.40
2	2.3192	2.3257	-0.28
3	2.1113	2.1170	-0.27
4	2.0350	2.0340	+0.05
5	1.8584	1.8590	+0.03
6	1.4842	1.4770	+0.49
7	1.2202	1.2214	-0.10
8	1.0131	1.0000	+1.31

note that, with the exception of Band 8, the effect of the out-of-band response is $\lesssim 0.5\%$ of the nominal $\varepsilon(\lambda_i, 865)$. Thus, with the exception of Band 8, $\langle \varepsilon(\lambda, 865) \rangle_{F_0 S_i}$ should follow Eq. (12.27) nearly as well as $\varepsilon(\lambda_i, 865)$, i.e., the spectral variations of $\langle \varepsilon(\lambda, 865) \rangle_{F_0 S_i}$, $i = 1$ to 7, and $\varepsilon(\lambda_i, 865)$ will be nearly identical, i.e., $\langle \varepsilon(\lambda, 865) \rangle_{F_0 S_i} \approx \varepsilon(\lambda_i, 865)$.

Finally, how do we deal with absorbing gases? A simple example is Ozone, which we model as consisting of an absorbing layer at the top of the atmosphere. This can be thought of as simply a modification of F_0 , i.e., $F_0(\lambda) \rightarrow F_0(\lambda) \exp[-M\tau_{Oz}(\lambda)]$, where τ_{Oz} is the Ozone optical depth and M is the two-way airmass ($1/\cos\theta_0 + 1/|\cos\theta_v|$). Noting that $\tau_{Oz}(\lambda) \ll 1$, $F_0(\lambda) \exp[-M\tau_{Oz}(\lambda)] \rightarrow F_0(\lambda)[1 - M\tau_{Oz}(\lambda)]$, so

$$\begin{aligned} \langle F_0(\lambda) \exp[-M\tau_{Oz}(\lambda)] \rangle_{S_i} &\approx \langle F_0(\lambda)[1 - M\tau_{Oz}(\lambda)] \rangle_{S_i} \\ &= \langle F_0(\lambda) \rangle_{S_i} \left[1 - M \langle \tau_{Oz}(\lambda) \rangle_{F_0 S_i} \right] \\ &\approx \langle F_0(\lambda) \rangle_{S_i} \exp \left[-M \langle \tau_{Oz}(\lambda) \rangle_{F_0 S_i} \right], \end{aligned}$$

and as in the case of τ_r earlier, τ_{Oz} must be averaged with weighting function $F_0 S_i$. This procedure works well for weakly absorbing gases, but for strongly absorbing gases like H_2O a detailed absorption line-by-absorption line analysis may be necessary. More detail is provided in the references.

We will not discuss the out-of-band effects on Eq. (12.23) further, as the pattern of dealing with these should now be clear. An important message to be taken from this is, given that $S_i(\lambda)$ is required for all of the computations in this section, it must be *measured*, i.e., the complete spectral properties of the individual spectral bands must be *characterized*.

It is interesting to note that in a sense in the quantity $\langle \rho_t(\lambda) \rangle_{F_0 S_i}$, the averaging over F_0 takes place automatically in reflectance-based calibration. From Section 12.2.2, the two currents (one viewing the scene and the other viewing the reflectance plaque) should actually be written

$$K_{S_i} I_{tS_i} = \langle L_t \rangle_{S_i} = \frac{\langle F_0(\lambda) \rho_t(\lambda) \rangle_{S_i} \cos \theta_0}{\pi} \quad \text{and} \quad K_{S_i} I_{PS_i} = \langle L_P \rangle_{S_i} = \frac{\langle F_0(\lambda) R_P(\lambda) \rangle_{S_i} \cos \theta_P}{\pi},$$

and noting that $\langle F_0(\lambda) \rho_t(\lambda) \rangle_{S_i} = \langle F_0(\lambda) \rangle_{S_i} \langle \rho_t(\lambda) \rangle_{F_0 S_i}$, etc., we have

$$\langle \rho_t \rangle_{F_0 S_i} = \langle R_P \rangle_{F_0 S_i} \left[\frac{\cos \theta_P}{\cos \theta_0} \right] \left[\frac{I_{tS_i}}{I_{PS_i}} \right],$$

so $\langle \rho_t \rangle_{F_0 S_i}$ is what we would actually measure in a setup as shown in Figure 12.2.

12.4.3 Stray Light

Stray light is simply light that ends up where it should not. Newer sensors have large focal planes containing linear or rectangular arrays of detectors. An example is the MODIS focal plane array shown schematically in Figure 12.13. The individual detectors are shown as small squares for the 36 bands. There are 10 detectors (in a vertical column) for each spectral band used to view the oceans. The velocity vector of the space craft is parallel to each column. The ocean bands are 8–16. Bands with more detectors are those with higher than the nominal 1 km spatial resolution for the “ocean bands.” As the scan mirror of MODIS rotates, the image is swept across the focal planes resulting in ten scan lines per mirror rotation.²⁶ The red object covering the lower right corner of each focal plane simulates the image of a cloud in the scene focussed by the scanner optics onto the focal plane. In the blue, the reflectance of a cloud is 4–5 times that of the ocean-atmosphere

²⁶Actually, as described in Chapter 11, the MODIS instrument uses a two-sided scan mirror, so a full rotation of the scan mirror results in twenty scan lines.

system, while in the NIR the associated factor is 30-40 times. The cloud moves across the focal planes as the instrument scans. One can imagine that some of the cloud's radiant energy can be scattered off the focal plane and end up in a detector that is far away from the cloud's image. It is necessary to characterize such effects prior to launch.

The seriousness of the stray light effect is dependent to a large part on the design of the instrument. Instruments with large focal planes (like MODIS) are particularly susceptible. However, an instrument designed like CZCS, which used a concave diffraction grating to focus the spectrum on single detectors placed in the focal plane, could also have stray light problems, even in the absence of clouds, as the instrument chamber could have scattered blue light falling on the detector placed at the position for red light, etc. In instrument design, every effort must be made to minimize the stray-light affects.

Sections 12.4.1–12.4.3 have provided examples of the kinds of phenomena that require characterization prior to launch. Failure to effect such characterization can seriously jeopardize the success of any satellite remote sensing mission.

12.5 Summary and Discussion

The “radiometric calibration” of the sensor is critical to the success of any ocean color mission. In this chapter we have reviewed the operational definition of radiance and the two kinds of radiometric calibration: radiance based and reflectance based. We determined the effect of error in the radiometric calibration and showed that, considering the accuracy required in the water-leaving reflectance, even error of the order of a few percent could be detrimental to an ocean color mission. We demonstrated that if the error has the same sign in all spectral bands, the propagation of the error through the processing algorithms is considerably reduced. Considering the difficulty of achieving the required calibration accuracy (even in a laboratory setting) and the stresses placed on the instrument during launch, on-orbit calibration (known as vicarious calibration) is a logical solution. We then looked at two different types of vicarious calibration. In the first, measurements of sky radiance are carried out with a well-calibrated radiometer. Simultaneously, the sensor views the water near the measurement site. The surface measurements allow estimation of the radiance the sensor “should” be viewing and the calibration is adjusted so that the sensor provides the indicated radiance. This method is most attractive in the NIR, where the water-leaving reflectance is essentially zero. The accuracy of this method is limited to the accuracy of the sky radiometer. In the second method, measurements of $\rho_w(\lambda)$ and atmospheric parameters (surface pressure, aerosol spectral optical depth, etc.) are made primarily to determine $\rho_r(\lambda)$ and the most likely of the atmospheric correction algorithm's candidate aerosol models. The algorithm is then operated with the chosen model and the

sensor calibration is adjusted to bring the measured and retrieved values of $\rho_w(\lambda)$ into confluence. This then is a vicarious calibration of the entire system; sensor *plus* algorithms. It is not radiometric, but it is shown that when employed, the calibration in the blue is uncertain to no more than a fraction of 1%. Any change in the atmospheric correction algorithm requires a revision of the calibration coefficients. In the final analysis, the vicarious calibration is a pragmatic exercise focussed on forcing the sensor *plus* algorithms to provide the correct result in a certain (restricted) setting in the hope that it will work equally well in other settings.

Finally, we discussed the need to *characterize* the sensor, and gave three examples of quantities that need to be assessed prior to launch: polarization sensitivity, out-of-band response, and stray light contamination.

The term “radiometric calibration” can take on different meanings depending on the context. For ocean color instruments, the author likes to think about it in a hierarchical manner: pre-launch to vicarious. First, the prelaunch radiometric calibration is really only necessary to set the saturation radiances of the sensor.²⁷ If the saturation radiance is set too low, then color data from some brighter features in the water or from some exceptionally hazy days may be lost. Conversely, if it is set too high, then the radiometric resolution available for the water-leaving radiance will be too low. Calibration accuracies of the order of 5-10% should be sufficient for this purpose.²⁸ More important than an accurate radiometric calibration is the short-term (\sim days) radiometric stability of the instrument in orbit, and the availability of a means of assessing it. Although the short-term stability of CZCS appeared to be quite good, there was no provision for assessing it. SeaWiFS had a solar diffuser that could be used on each orbit if necessary to assess short-term stability. The MODIS instrument actually had an on-board calibration sphere for this purpose.

Once in orbit, vicarious calibration can commence by examining the central gyres of the oceans where the water-leaving reflectance is known and relatively stable. In such regions, the aerosol is known to be mostly maritime and the surface atmospheric pressure can be estimated based on numerical weather models. Thus, an initial system calibration along the lines described in Section 12.3.2 can be carried out. Concurrently, a marine facility such as MOBY specifically designed to provide normalized water-leaving reflectance and some atmospheric measurements on a continuous basis can be utilized for providing the more accurate system calibration for each satellite overpass with the appropriate meteorological conditions and begin a time-series of calibration adjustments. In addition, views of the moon (if possible) can be used to assess the long-term stability. Only after a sufficient

²⁷The *saturation radiance* is the radiance associated with the maximum number of digital counts available for the analog-to-digital conversion of the current from the individual detectors.

²⁸Remember that radiometric calibration is very expensive. Halving the calibration uncertainty will likely increase the cost by much more than a factor of two.

number of calibration facility-sensor match ups are available (30 to 40) can the system calibration be considered complete. At this time the entire mission data set is reprocessed to provide a retrospective time series of imagery.²⁹ The “final” calibration uncertainty will be dependent on the uncertainty at λ_l , which can be reduced (or assessed) by the methods on Section 12.3.1. It is important to stress that this system calibration (sensor *plus* algorithms) must be reassessed if any changes are made to the algorithms, e.g., using a new set of candidate aerosol models or a revised extraterrestrial solar irradiance.

12.6 Appendix: Retrieval of ω_a and $P_a(\Theta)$ from $\rho_B(\hat{\xi}_B)$.

This appendix describes an algorithm that can be used to invert measurements of sky radiance to obtain the aerosol phase function and single scattering albedo, when the aerosol optical thickness is known. The retrieved ω_a and $P_a(\Theta)$ are defined to be those that, when inserted into the radiative transfer equation, result in a solution that reproduces the measured radiances. This is the multiple scattering algorithm that was employed in Section 12.3.1.

As in the atmospheric correction algorithm, we simplify the atmosphere as a two-layer system with molecules above the aerosols. Then we can write the total downward normalized radiance measured at the bottom of the atmosphere, ρ_B in a given direction as

$$\rho_B = \rho_r e^{-\tau_a/u_B} + \rho_a e^{-\tau_r/u_0} + \rho_{ra} + \rho_{\text{Sun}} e^{-(\tau_r+\tau_a)/u_0} \delta(u_B - u_0) \delta(\phi_B - \phi_0),$$

where ρ_r is the normalized radiance resulting from multiple scattering by air molecules (Rayleigh scattering) in the absence of aerosols and then transmitted through the aerosol layer, ρ_a is the normalized radiance resulting from multiple scattering by aerosols in the absence of the air but modified by the fact that the solar beam had to pass through the Rayleigh layer, ρ_{ra} is the interaction term between molecules and aerosols, i.e., the contribution to the normalized radiance from photons scattered by *both* molecules and aerosols, and ρ_{Sun} is the downward normalized radiance of the direct solar beam. The variables u_B and u_0 as usual are, respectively, the cosine of viewing and solar zenith angles, while ϕ_B and ϕ_0 are the viewing and solar azimuth angles. The ρ_{Sun} term in the above equation can be measured by a Sun photometer and is used to determine τ_a . For $u_B \neq u_0$ or $\phi_B \neq \phi_0$, we can write

$$\rho_B = \rho_r e^{-\tau_a/u_B} + \rho_a e^{-\tau_r/u_0} + \rho_{ra}.$$

²⁹In practice, the data set is likely to be reprocessed many times, and even before, the calibration is deemed reliable.

To relate the downward normalized radiance to the optical properties of aerosols, we rewrite the above equation as follows:

$$\rho_B - \rho_r e^{-\tau_a/u_B} = \rho_a e^{-\tau_r/u_0} + \rho_{ra}.$$

The dominant term in $\rho_B - \rho_r e^{-\tau_a/u_B}$ is the result of downward single scattering by the aerosol, ρ_{as} .³⁰ In the single scattering approximation, the downward normalized radiance is composed of two parts: direct scattering without reflection from the surface; and reflection of the direct solar beam from the sea surface followed by backscattering in the atmosphere. The second term, however, is an order of magnitude less than the first term. Therefore, for a specific viewing-sun geometry, in the single (aerosol) scattering approximation, it is legitimate to write

$$\rho_B - \rho_r e^{-\tau_a/u_B} \propto \rho_{as}, \quad (12.28)$$

where ρ_{as} is the direct downward single scattering normalized radiance without reflection from the sea surface, given by

$$\rho_{as} = \frac{\omega_a \tau_a P_a(\Theta)}{4 \cos \theta_B \cos \theta_0}, \quad (12.29)$$

with

$$\cos \Theta = \cos \theta_0 \cos \theta_B + \sin \theta_0 \sin \theta_B \cos(\phi - \phi_0).$$

We assume that ρ_r can be predicted to any desired accuracy, and τ_a is measured at the bottom of atmosphere, so the left-hand-side of Eq. (12.28) is known. We also assume that we are given ρ_B measurements in a large enough number of directions (u_B, ϕ_B) to provide as full a range of scattering angles Θ as possible. Then Eq. (12.29) provides $\omega_a P_a(\Theta)$ for that range of scattering angles.

We denote $\rho_B^{(m)}$ as the total downward normalized radiance measured (“ m ”) by the sensor, $\rho_B^{(c)}$ as the total downward normalized radiance computed (“ c ”) via the radiative transfer equation using estimated optical properties for the aerosol, $\omega_a P_a$, and the measured τ_a , along with distributing the aerosol according to the two-layer model. In the computation of $\rho_B^{(c)}$ the atmosphere is bounded by a flat Fresnel-reflecting ocean at the bottom. Radiance backscattered *out* of the water is ignored as we will be most interested in applying these results to the red and NIR, where ρ_w is small or negligible. In addition $\rho_r^{(c)}$ is the computed downward normalized radiance in the absence of aerosols. Then Eq. (12.28) becomes

$$\rho_B^{(m)} - \rho_r^{(c)} e^{-\tau_a/u_B} \propto \rho_{as}^{(m)} \quad \text{and} \quad \rho_B^{(c)} - \rho_r^{(c)} e^{-\tau_a/u_B} \propto \rho_{as}^{(c)}$$

where

$$\rho_{as}^{(m)} = \frac{\tau_a [\omega_a P_a(\Theta)]^{(t)}}{4 \cos \theta_B \cos \theta_0} \quad \text{and} \quad \rho_{as}^{(c)} = \frac{\tau_a [\omega_a P_a(\Theta)]^{(c)}}{4 \cos \theta_B \cos \theta_0}.$$

³⁰For sufficiently small τ_a , the right hand side is equal to $\rho_{as} \exp(-\tau_r/u_0)$.

The superscript “ t ” stands for the “true” value. Combining these equations, we have

$$\frac{\rho_B^{(c)} - \rho_B^{(m)}}{\rho_B^{(m)} - \rho_r^{(c)} e^{-\tau_a/u_B}} = \frac{\Delta[\omega_a P_a(\Theta)]}{[\omega_a P_a(\Theta)]^{(t)}},$$

where

$$\Delta[\omega_a P_a(\Theta)] = [\omega_a P_a(\Theta)]^{(c)} - [\omega_a P_a(\Theta)]^{(t)}.$$

Note we have assumed the same proportionality constant for the measured and calculated quantities. The error in $\omega_a P_a(\Theta)$ is then

$$\Delta[\omega_a P_a(\Theta)] = \frac{\rho_B^{(c)} - \rho_B^{(m)}}{\rho_B^{(m)} - \rho_r^{(c)} e^{-\tau_a/u_B}} [\omega_a P_a(\Theta)]^{(t)}. \quad (12.30)$$

Retrieval of $\omega_a P_a(\Theta)$ can be achieved through iteration on the above equation. Replacing the unknown $[\omega_a P_a(\Theta)]^{(t)}$ by $[\omega_a P_a(\Theta)]^{(c)}$ in the above equation, a new estimate of $[\omega_a P_a(\Theta)]$ can be written as

$$[\omega_a P_a(\Theta)]_{\text{new}} = [\omega_a P_a(\Theta)]_{\text{old}} - C \Delta[\omega_a P_a(\Theta)], \quad (12.31)$$

where C is a constant, the value of which is chosen so as to avoid instabilities in the iteration scheme when there are errors in the measured radiances (here we use $C = 1/2$). At the end, both $P_a(\Theta)$ and ω_a can be retrieved using the normalization of the phase function, i.e.,

$$\omega_a = \frac{1}{2} \int_0^\pi [\omega_a P_a(\Theta)] \sin \Theta d\Theta, \quad (12.32)$$

$$P_a(\Theta) = \frac{1}{\omega_a} [\omega_a P_a(\Theta)]. \quad (12.33)$$

Note that aerosol phase function $P_a(\Theta)$ is normalized to 4π . The basic idea then is to use Eqs. (12.30) and (12.31) to nudge $\omega_a P(\Theta)$ in a direction that will tend to decrease $\rho_B^{(c)} - \rho_B^{(m)}$. However, we need an initial guess for $\omega_a P(\Theta)$ to be able to solve the radiative transfer equation to provide the initial $\rho_B^{(c)}$.

The initial guess for $\omega_a P(\Theta)$ is provided by a series of simulations showing that $\rho_B - \rho_r e^{-\tau_a/u_B}$ is approximately related to ρ_{as} quadratically, i.e.,

$$\rho_B - \rho_r e^{-\tau_a/u_B} \approx a + b\rho_{as} + c\rho_{as}^2. \quad (12.34)$$

A set of such simulations for one particular geometry is provided in Figure 12.14. The figure is for $\theta_B = 60^\circ$, $\theta_0 = 60^\circ$ and $\phi_B - \phi_0 = 90^\circ$ at $\lambda = 865$ nm with three aerosol models: M99, M70, and T50. The computations for preparing the figure are for $\tau_a = 0.05, 0.10, 0.15, 0.20, 0.30$ and 0.40 . The line is computed using a least-squares determination

of a , b , and c in Eq. (12.34). The quality of the fit suggests that by knowing a , b , c , and $\rho_B^{(m)} - \rho_r^{(c)} e^{-\tau_a/u_B}$, ρ_{as} can be estimated from Eq. (12.34), with an error no greater than 20%. This will provide an estimate for $\omega_a P_a(\Theta)$ using Eq. (12.29). This estimate will be used *only* to *initiate* the algorithm. However, as we discussed in the text, the maximum scattering angle available is $\theta_0 + \pi/2$, so in order to use Eq. (12.32) and (12.33) to estimate ω_a and $P_a(\Theta)$, we have to make an assumption concerning $\omega_a P_a(\Theta)$ for $\Theta = \Theta_{\max}$ to 180° . The simplest is to assume P_a is simply constant for $\Theta > \Theta_{\max}$, i.e.,

$$\omega_a P_a(\Theta) = \omega_a P_a(\Theta_{\max}), \quad \text{if } \Theta > \Theta_{\max}.$$

The algorithm then consists of six steps delineated below.

1. Calculate a , b , and c of the quadratic fit of $\rho_B^{(c)} - \rho_r^{(c)} e^{-\tau_a/u_B}$ to ρ_{as} for each wavelength, wind speed (surface roughness), and the given geometry using a variety of aerosol models, e.g., the candidates of the atmospheric correction algorithm.
2. Use the measured value of $\rho_B^{(m)}$ and τ_a to compute ρ_{as} through Eq. (12.34), where a , b , and c are known from (1) for the given geometry. Then $\omega_a P_a(\Theta)$ can be initially estimated for the whole range Θ through Eqs. (12.29) and the assumption for $\omega_a P_a(\Theta)$ for $\Theta > \Theta_{\max}$.
3. Use the retrieved $\omega_a P_a(\Theta)$ and the radiative transfer equation to calculate the total downward radiances $\rho_B^{(c)}$.
4. The new value of $\omega_a P_a(\Theta)$ for $\Theta \leq \Theta_{\max}$ can then be retrieved through

$$\Delta[\omega_a P_a(\Theta)] = \frac{\rho_B^{(c)} - \rho_B^{(m)}}{\rho_B^{(m)} - \rho_r^{(c)} e^{-\tau_a/u_B}} [\omega_a P_a(\Theta)]_{\text{old}}$$

and

$$[\omega_a P_a(\Theta)]_{\text{new}} = [\omega_a P_a(\Theta)]_{\text{old}} - \frac{1}{2} \Delta[\omega_a P_a(\Theta)],$$

where $\Theta_{\max} = \theta_0 + \pi/2$, and the subscripts “new” and “old” indicate, respectively, the new and the previously retrieved value of $\omega_a P_a(\Theta)$.

5. From Eqs. (12.32) and (12.33), ω_a and $P_a(\Theta)$ can then be retrieved using the assumption of a constant $\omega_a P_a(\Theta)$ for $\Theta > \Theta_{\max}$.
6. Repeat steps (3) to (5) until convergence is achieved, i.e., $\rho_B^{(c)} - \rho_B^{(m)}$ is minimized over all of the measured ρ_B 's.

This algorithm was used in the preparation of Figures 12.7-12.9 in the text. The reader should note the similarity between this algorithm and that for retrieving the IOPs from measurement of the AOPs in Appendix 1 of Chapter 6.

12.7 Bibliographic Notes

12.1 Introduction

The influence of radiometric calibration error on the retrieved values of ρ_w follows the development in [Gordon \[1997\]](#).

12.2 Radiometric Calibration

Portions of the discussion here on radiometric calibration draw on the author's contribution to the [IOCCG \[2013\]](#) Report. An excellent discussion of radiometric calibration in general, including basic definitions, standards, and methods can be found in [Johnson et al. \[2014\]](#).

12.2.2 Reflectance Based

Reflectance-based calibration of satellite sensors is discussed in [Slater et al. \[1987\]](#).

12.3 Vicarious Calibration

The method of vicarious calibration was introduced by [Koepke \[1982\]](#) and first applied to METEOSAT. An excellent description of vicarious calibration applied to several different sensors is provided in [IOCCG \[2013\]](#).

12.3.1 Direct — Using Surface Radiance Measurements

The direct method of vicarious calibration and a sensitivity study of the influence of the particular assumptions is detailed in [Gordon and Zhang \[1996\]](#). The method of inverting sky radiance to obtain $\omega_a P(\Theta)$ was first developed by [Wang and Gordon \[1993\]](#) and later extended to the inversion of the surface-measured Stokes vector to obtain portions of the phase matrix $\omega_a P_a(\Theta)$ by [Zhang and Gordon \[1997\]](#). In anticipation of using the direct method from stations located on small islands, [Yang et al. \[1995\]](#) studied the island perturbation effects on the retrieval of $\omega_a P_a(\Theta)$. An application of the method in the NIR is provided in [Martiny et al. \[2005\]](#).

A method for reflectance-based calibration of surface radiometers is detailed in [Cattrall et al. \[2002\]](#).

12.3.2 Full System Calibration

The calibration requirements (and vicarious methodology) for ocean color sensors is

described in [Gordon \[1987\]](#). The full system calibration concept (for CZCS) was first discussed by [Gordon \[1981\]](#) (see also [Viollier \[1982\]](#)), and in considerable detail in [Gordon \[1998\]](#). [Wang and Gordon \[2002\]](#) showed that when applying such a calibration method as described in the text, the actual value of the calibration error at λ_l is relatively unimportant as long as it is not too large. They also suggest that it is unnecessary to try to reduce the prelaunch calibration uncertainties to very low values, as in-orbit calibration is sufficiently accurate. Instead it is more useful to use resources to improve the instrument's stability or to fully characterize the instrument, e.g, to determine the effect of the instrument's temperature on the measured calibration coefficients, $K(\lambda, \dots)$, even though they are not as accurate as they could be, etc.

The Marine Optical Buoy (MOBY) is described in [Clark et al. \[1997\]](#) and references therein. The application of MOBY to the full system vicarious calibration of SeaWiFS is detailed in [Franz et al. \[2007\]](#), where, interestingly, over a period of nine years there were 1450 contemporaneous MOBY measurements and satellite overpasses with only 150 of those passing their stringent quality control metrics. In addition, they show that it takes roughly 30 to 40 samples to reach a stable vicarious calibration, requiring 2-3 years to achieve. Thus, frequent reprocessing of the imagery starting from the first usable image is necessary.

12.4 Characterization

Portions of the discussion here on characterization draw on the author's contribution to the [IOCCG \[2013\]](#) Report.

12.4.1 Polarization Sensitivity

The material in the section is taken mostly from [Gordon et al. \[1997a\]](#), wherein the polarization sensitivity is discussed along with several examples of its possible affect on the retrieved ρ_t . In addition a method for correction of such effects is provided. An end-to-end example of the characterization and correction procedure as applied to MODIS (Aqua) is given in [Meister et al. \[2005\]](#).

12.4.2 Out-of-Band Response

A through discussion of the out-of-band problem for SeaWiFS is provided in [Gordon \[1995\]](#). [Wang et al. \[2001\]](#) used this as a basis for the analysis of the effect of the out-of-band response on SeaWiFS-derived normalized water-leaving radiances and concentration of Chlorophyll *a*. They emphasized the importance of pre-launch characterization of the spectral response, not only in understanding the performance of a given sensor, but also in merging data from different sensors, e.g., SeaWiFS and MODIS.

12.5 Appendix: Retrieval of ω_a and $P_a(\Theta)$ from $\rho_B(\hat{\xi}_B)$.

This was taken from [Wang and Gordon \[1993\]](#), wherein examples of several simulated retrievals are provided. A similar algorithm has been developed for measurements of the full Stokes vector of the sky radiance and using the vector radiative transfer equation to retrieve ω_a , $P_{11}(\Theta)$ and $P_{12}(\Theta)$ for aerosols [[Zhang and Gordon, 1997](#)].

12.8 Figures

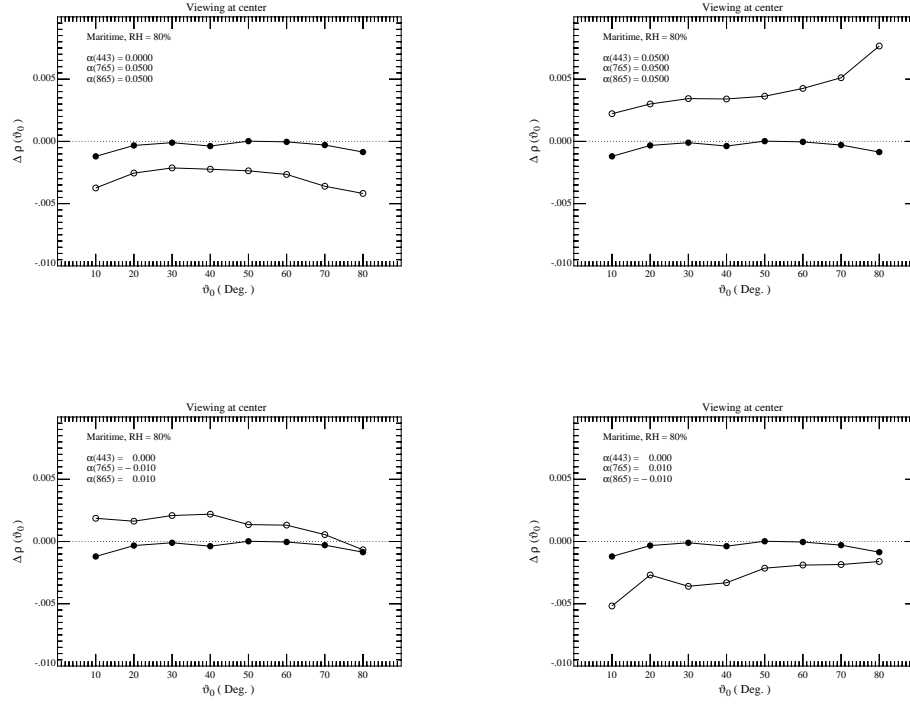


Figure 12.1: Error in the retrieved $t(443)\rho_w(443)$ for viewing at the center of the scan with a Maritime aerosol at RH = 80% as a function of the solar zenith angle with $\tau_a(865) = 0.2$ and calibration errors $\alpha(443)$, $\alpha(765)$, and $\alpha(865)$ in Eq. (12.1) (open circles). Solid circles are for $\alpha(\lambda_i) = 0$ for all λ_i . Redrawn from Gordon [1997].

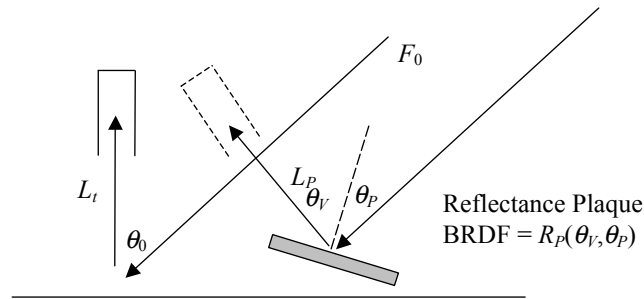


Figure 12.2: Schematic of the reflectance-based calibration of a sensor in orbit. From IOCCG [2013].

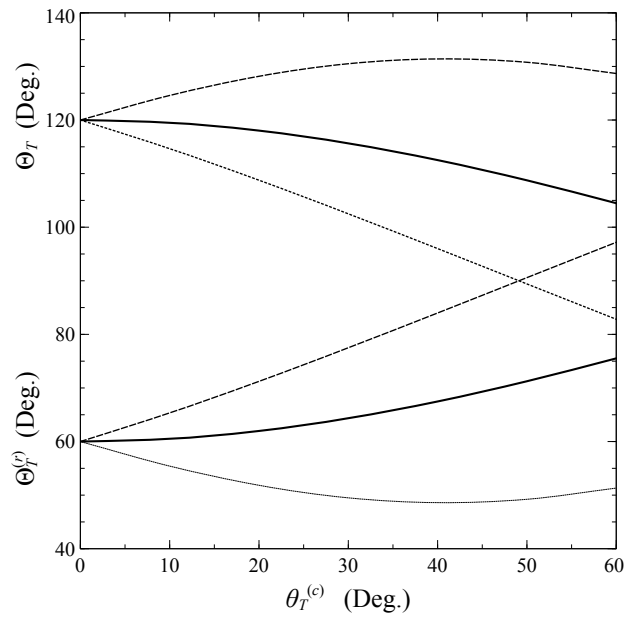


Figure 12.3: The quantities Θ_T (upper curves) and $\Theta_T^{(r)}$ (lower curves) as a function of $\theta_T^{(c)}$ for $\theta_0 = 60^\circ$. The individual curves for Θ_T and $\Theta_T^{(r)}$ are for $\phi_T - \phi_0 = 120^\circ$ (dashed), 90° (solid) and 60° (dotted).

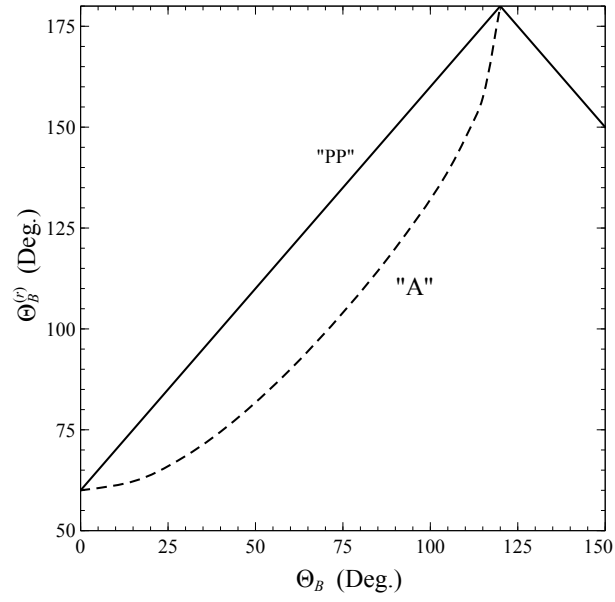


Figure 12.4: The quantity $\Theta_B^{(r)}$ as a function of Θ_B for principal plane “PP” and almucantar “A” scans with $\theta_0 = 60^\circ$.

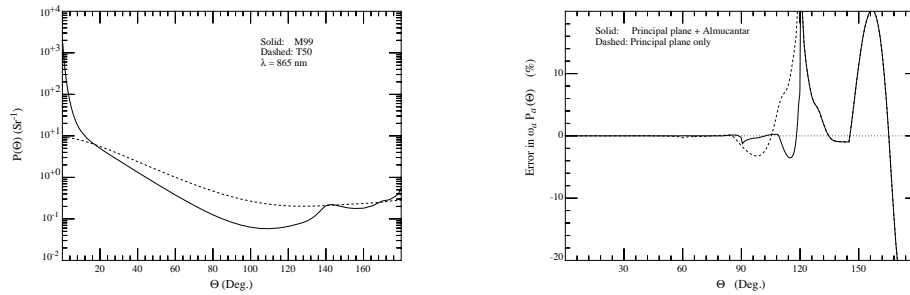


Figure 12.5: Phase functions for the M99 and T50 aerosol models. Error in the retrieved $\omega_a P_a(\Theta)$ for M99 assuming that single scattering is the correct physics. From [Gordon and Zhang \[1996\]](#).

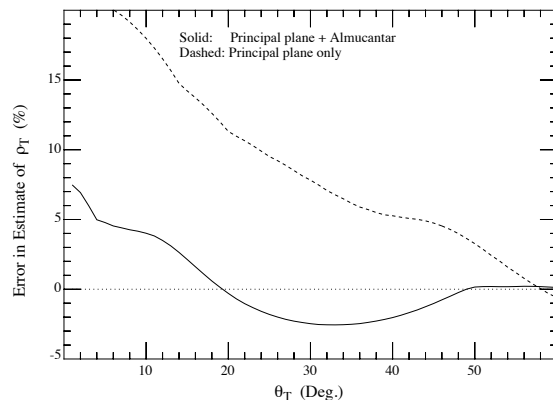


Figure 12.6: Error in the estimate of ρ_T as a function of $\theta_T^{(c)}$ (labeled “ θ_T ” on the figure) when single scattering is the correct physics. Here, $\theta_0 = 60^\circ$ and $\phi_T - \phi_0 = 90^\circ$. The curves correspond to using the PP scan alone or the PP+A scans to retrieve $\tau_a \omega_a P(\Theta)$. From [Gordon and Zhang \[1996\]](#).

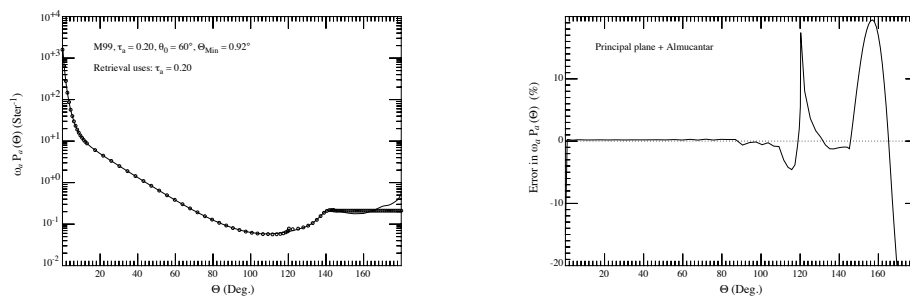


Figure 12.7: Retrieved $\omega_a P_a(\Theta)$ when multiple scattering is included in both the generation of the ρ_B pseudo data and the retrieval of $\omega_a P_a(\Theta)$ from ρ_B (left panel). Error in the corresponding retrieved $\omega_a P_a(\Theta)$ (right panel). From [Gordon and Zhang \[1996\]](#).

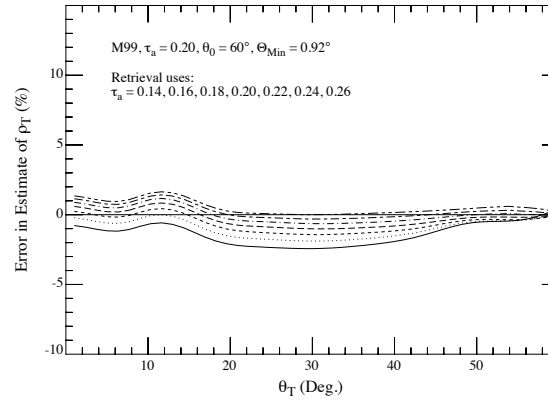


Figure 12.8: Error in ρ_T as a function of $\theta_T^{(c)}$ (labeled “ θ_T ” on the figure) when multiple scattering is included in both the generation of the ρ_B pseudo data and the retrieval of $\omega_a P_a(\Theta)$ from ρ_B . The various curves correspond from bottom to top to $\tau_a = 0.14, 0.16, 0.18, 0.20, 0.22, 0.24$, and 0.26 , in both the retrieval and prediction codes. The actual τ_a used in generation of the pseudo data was 0.2 . Θ_{Min} is the smallest value of Θ_B obtained in the almucantor scan. From [Gordon and Zhang \[1996\]](#).

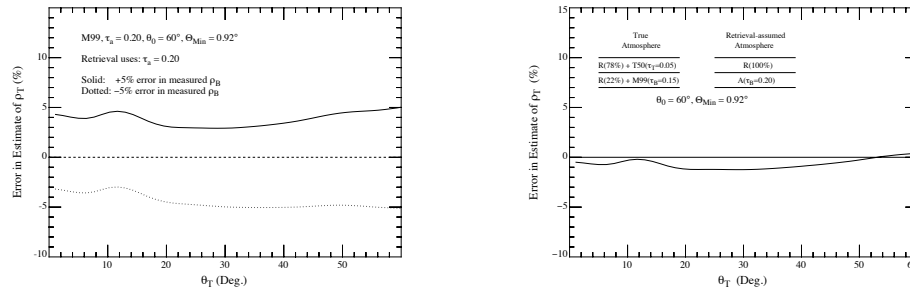


Figure 12.9: Error in ρ_T for a $\pm 5\%$ error in ρ_B for the same cases as Figures 12.7 and 12.8, when the correct value of τ_a is used (left panel). Error in ρ_T when an incorrect assumption is used about the vertical structure of the atmosphere (right panel). The pseudo data have M99 in a 1 km thick layer above the surface uniformly mixed with air, and a second layer containing a different aerosol (T50) and the rest of the molecular scattering. The retrieval and prediction codes assume all of the aerosol is in a thin layer just above the surface and uses the correct τ_a . From [Gordon and Zhang \[1996\]](#).

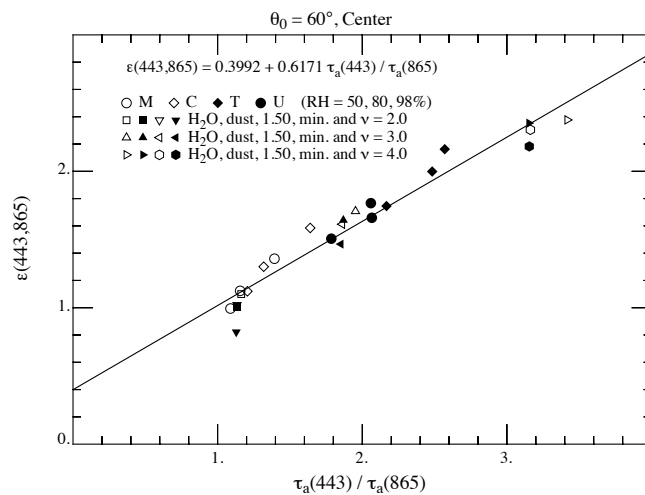


Figure 12.10: Relationship between $\tau_a(443)/\tau_a(865)$ and $\varepsilon(443,865)$ at the scan center with $\theta_0 = 60^\circ$ for several aerosol models. From [Gordon \[1997\]](#).

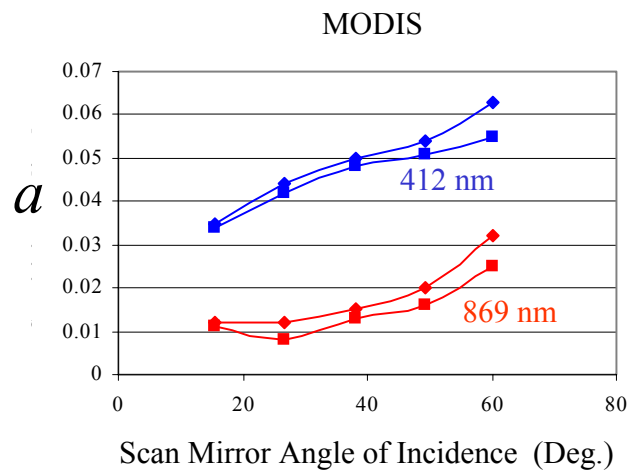


Figure 12.11: Polarization-sensitivity amplitude a for the Terra version of MODIS for Bands 8 (412 nm) and 16 (869 nm). For each wavelength the two curves are for the two sides of the MODIS scan mirror. From [IOCCG \[2013\]](#).

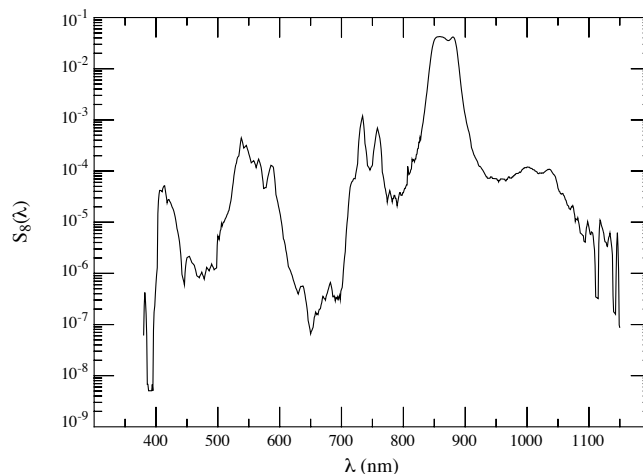


Figure 12.12: Spectral response of SeaWiFS band 8, normalized such that $\int S_8(\lambda) d\lambda = 1$. Data are taken from Barnes et al. [1994b] and graph is from Gordon [1995].

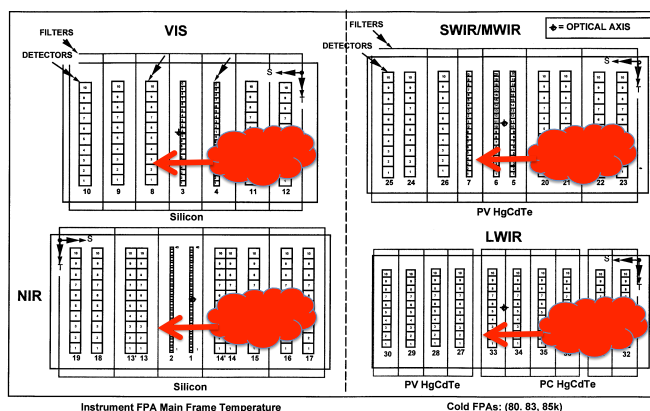


Figure 12.13: The four focal planes of the MODIS instruments. The individual detectors are shown as small squares for the 36 bands. There are 10 detectors (in a vertical column) for each spectral band for ocean applications. The ocean bands are 8–16. Bands with more detectors are those with higher than the nominal 1 km spatial resolution. The red object covering the lower right corner of each focal plane simulates the image of a cloud in the scene. The cloud moves across the focal planes as the instrument scans.

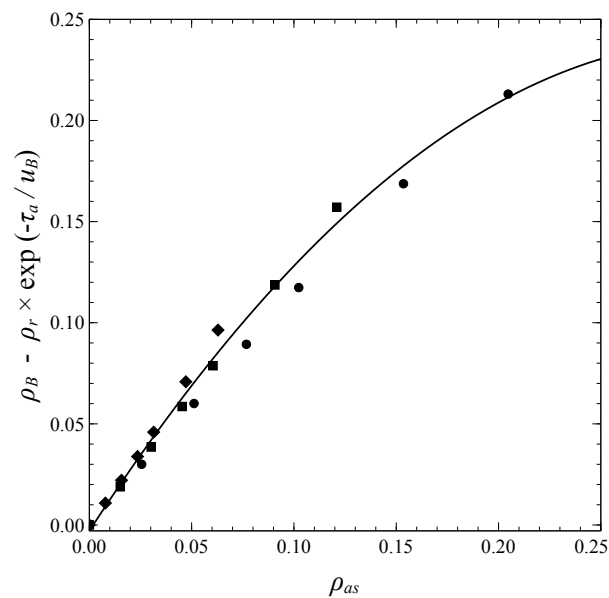


Figure 12.14: The quantity $\rho_B - \rho_r \exp[-\tau_a/u_B]$ as a function of ρ_{as} for $\theta_0 = 60^\circ$, $\phi_0 = 0$, $\theta_B = 60^\circ$, and $\phi_B = 90^\circ$ at $\lambda = 865$ nm. Diamonds: M99. Squares: M70. Circles: T50. The line is the least-squares fit to a quadratic, which reproduces the measured $\rho_B - \rho_r \exp[-\tau_a/u_B]$ with an average (unsigned) error of 9% and a maximum error of 18%.

Chapter 13

Vicarious Calibration and Validation — How Well Does It All Work?

13.1 Introduction

We have finally reached the point in our development when we begin to answer the question, how well does remote sensing of the ocean, through estimating the water-leaving radiance from measurements made by satellite-borne sensors, actually work? The process of determining its effectiveness, within a certain range of uncertainty, is called *validation*. Since we have already demonstrated that products of ecological and climatological importance in oceanography such as the concentrations of Chlorophyll *a* (*C*) or the pigment concentration (*C_P*) can be derived, with a roughly-known uncertainty, given the normalized water-leaving reflectances $[\rho_w(\lambda)]_N$, e.g., Figures 6.64 or 10.1 and Eq. (10.1), it is quite obvious that we should focus the validation effort mainly on $[\rho_w(\lambda)]_N$. However, validation of the principal *scientific* products, e.g., *C* or *C_P*, is still necessary, as an understanding of their accuracy is crucial in further applications.

Before we begin, it is useful to consider the differences between measurements at the surface, made perhaps from a ship, and measurements from a sensor in space. The ship-borne sensor can only effect measurement of a given quantity at a single location, or at most a small number of locations on the surface, or provide a time series of measurements at a single location. The space-borne sensor, through its capability of imaging large swaths of the water surface in a short period of time (minutes), and even global-scale images

over a period of \sim one week, can produce a snapshot of the large scale environment and provide the context in which any variation observed on the small scale takes place. Thus, the two kinds of measurements are complementary, but focus on different views of the same phenomenon — the spatial-temporal variation of the examined property. There is no reason to expect that these should have the same accuracy in the quantification of said property, and indeed, they need not in order for both to be useful. Considering the overall scope of the instrumentation and data processing required for the remote retrieval of most quantities of interest here, we should actually expect more uncertainty in the space-borne sensor. However, one can sacrifice some accuracy in the remote measurement to gain the perspective offered by the view from space. The purpose of validation is to understand the ability of the space-borne measurement to reproduce the surface measurement, and to quantify the uncertainty in the former, usually by taking the surface measurement as “truth,” or at least as having an uncertainty significantly less than the space measurement.

In the case of ocean color sensors and the role of the oceans in climate change, we have an additional goal — obtaining a long-term (decades or longer) data record having enough sensitivity to detect and document any long-term trends in the measured quantity, e.g., Chlorophyll *a*. For this, knowing the uncertainty in the data record is paramount, and raises an important point. Few measuring devices can maintain their original performance over a long period of time, e.g., their calibration. Space-borne ocean color sensors are no exception (neither are laboratory radiometers for that matter). Thus, repeat calibration and validation exercises must be carried out during the entirety of a mission, and there can be no single “validation” of the remotely-sensed product — it is a never-ending on-going process throughout the lifetime of the sensor. In this sense, fully-validated ocean color products must of necessity be retrospective and produced at the completion of the mission.

With the above comments in mind, we proceed here to explain in detail what must be done for an ocean color sensor to validate its products, principally $[\rho_w]_N$, but also C or C_P , and follow with two examples of validation: first the CZCS, where validation was only crudely carried out; and then MODIS, for which the detailed requirements for validation have been fully met.

13.2 How Do We Perform Validation?

Clearly, we need to compare sensor and surface measurements of $[\rho_w]_N$ to assess the fidelity of its remote measurement. We note that $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$ is measured at an extremely large number of $\sim 1 \text{ km} \times 1 \text{ km}$ (pixels) at a given instant of time t_v . However, unlike the sensor’s measurement, the surface measurement, e.g., from a ship, is never instantaneous and may in fact require hours (as in the case of the CZCS validation measurements). So, we are forced

to admit that simultaneous measurement of $[\rho_w]_N$ by the sensor and a surface observer is not (at this time) possible, and we are forced to be satisfied with contemporaneous measurements, e.g, measurements at the same place on the same day but maybe a few hours apart. In this section we will focus on what measurements are needed and how they must be processed in order to provide a meaningful comparison between surface and satellite measurements of $[\rho_w]_N$. At the end we briefly consider the validation of derived products.

13.2.1 Surface Measurements of $[\rho_w]_N$.

The goal of the validation exercise is to determine the accuracy with which $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$ (where we have included the full dependence on the viewing and illumination directions as well as the wavelength in the argument list) can be derived from the sensor, in a variety of situations. To effect this we will need to measure $[\rho_w]_N$ at the water surface and compare it to the estimate of the same quantity, made by applying all of the machinery developed in previous chapters, from the radiance L_t (or reflectance ρ_t) measured by the sensor in space.

Before starting, as we must determine $[\rho_w]_N$ at the water surface, we will review in some detail how it is defined, so we can delineate the procedure for its measurement. The water-leaving reflectance ρ_w is related to the water-leaving radiance L_w through

$$\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0) = \frac{\pi L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)}{F_0(\lambda) \cos \theta_0}.$$

In addition, the normalized water-leaving radiance $[L_w]_N$ is related to L_w through

$$[L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N = L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0) \frac{a_\oplus^2}{\cos \theta_0 t_E(\lambda, \hat{\xi}_0)},$$

where $t_E(\hat{\xi}_0)$ is the transmittance of *irradiance* from the top of the atmosphere to the water surface, and a_\oplus is the Earth-Sun distance, in astronomical units, at the time of the measurement. This leads to (see all of the steps in Chapter 7)

$$\frac{\pi [L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N}{\bar{F}_0(\lambda)} = \frac{\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)}{t_E(\hat{\xi}_0)},$$

with \bar{F}_0 the solar irradiance for $a_\oplus = 1$ AU. The left-hand-side in the above equation is *defined* to be the normalized water-leaving reflectance $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$, i.e.,

$$[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N \equiv \frac{\pi [L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N}{\bar{F}_0(\lambda)}, \quad (13.1)$$

so

$$[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N = \frac{\pi L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)}{F_0(\lambda) \cos \theta_0 t_E(\lambda, \hat{\xi}_0)} = \pi \frac{L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)}{E_d^+(\lambda, \hat{\xi}_0)} = \pi R_{rs}(\lambda, \hat{\xi}_v, \hat{\xi}_0), \quad (13.2)$$

where R_{rs} is the remote sensing reflectance.¹

Accordingly, from Eq. (13.2), the quantities that must be measured at the surface to determine $[\rho_w]_N$ are L_w and E_d^+ , for the appropriate Sun and viewing directions. In most cases, these are not measured. More often the upwelling radiance L_u and downwelling radiance E_d are measured (or extrapolated to) *just beneath* the water surface. In this case, these need to be transformed to $[\rho_w]_N$ in the viewing direction. To understand how this is done, recall from Chapter 7 that

$$R_{rs}(\lambda, \hat{\xi}_v, \hat{\xi}_0) = \Re(\lambda, \hat{\xi}_v, \hat{\xi}_0) \left(\frac{L_u(\lambda, \hat{\xi}'_v, \hat{\xi}'_0)}{E_d(\lambda, \hat{\xi}'_0)} \right) = \Re(\lambda, \hat{\xi}_v, \hat{\xi}_0) R_L(\lambda, \hat{\xi}'_v, \hat{\xi}'_0), \quad (13.3)$$

where $\hat{\xi}_0$ and $\hat{\xi}'_0$ as well as $\hat{\xi}_v$ and $\hat{\xi}'_v$ are related by Snell's law, and

$$\Re(\lambda, \hat{\xi}_v, \hat{\xi}_s) = \frac{t_f(\hat{\xi}_v, \hat{\xi}'_v) t_f(\hat{\xi}_0, \hat{\xi}'_0)}{m_w^2 (1 - r R(\lambda, \hat{\xi}'_0))} \quad (13.4)$$

can be computed for a *flat* air-water interface.² These lead to Eq. (7.38), which can be rewritten as the following relationship between normalized water-leaving reflectances viewed with $\hat{\xi}_{v1}$ and $\hat{\xi}_{v2}$ for solar propagation vectors $\hat{\xi}_{01}$ and $\hat{\xi}_{02}$, respectively:

$$\frac{[\rho_w(\lambda, \hat{\xi}_{v1}, \hat{\xi}_{01})]_N}{\Re(\lambda, \hat{\xi}_{v1}, \hat{\xi}_{01}) R_L(\lambda, \hat{\xi}'_{v1}, \hat{\xi}'_{01})} = \frac{[\rho_w(\lambda, \hat{\xi}_{v2}, \hat{\xi}_{02})]_N}{\Re(\lambda, \hat{\xi}_{v2}, \hat{\xi}_{02}) R_L(\lambda, \hat{\xi}'_{v2}, \hat{\xi}'_{02})}. \quad (13.5)$$

We will find it helpful to use a short-hand notation for the recurring fraction

$$\frac{\Re(\lambda, \hat{\xi}_{v1}, \hat{\xi}_{01}) R_L(\lambda, \hat{\xi}'_{v1}, \hat{\xi}'_{01})}{\Re(\lambda, \hat{\xi}_{v2}, \hat{\xi}_{02}) R_L(\lambda, \hat{\xi}'_{v2}, \hat{\xi}'_{02})} \equiv \mathcal{F}(\lambda; \hat{\xi}_{v1}, \hat{\xi}_{01}; \hat{\xi}_{v2}, \hat{\xi}_{02}) = \frac{1}{\mathcal{F}(\lambda; \hat{\xi}_{v2}, \hat{\xi}_{02}; \hat{\xi}_{v1}, \hat{\xi}_{01})}$$

so

$$[\rho_w(\lambda, \hat{\xi}_{v1}, \hat{\xi}_{01})]_N = \mathcal{F}(\lambda; \hat{\xi}_{v1}, \hat{\xi}_{01}; \hat{\xi}_{v2}, \hat{\xi}_{02}) [\rho_w(\lambda, \hat{\xi}_{v2}, \hat{\xi}_{02})]_N.$$

¹The variation of R_{rs} with $\hat{\xi}_v$ and $\hat{\xi}_0$ is usually termed the “BRDF effect” (see Chapter 2, Section 2.8.4).

²The unknown irradiance ratio R , which is usually much less or at least less than 0.1, can be determined by measuring E_u along with E_d , can be estimated by modeling, or can be estimated by πL_u , which assumes that L_u is totally diffuse.

It is easy to see that the following relationships among different \mathcal{F} 's are valid:

$$\begin{aligned} \frac{\mathcal{F}(\lambda; \hat{\xi}_{v1}, \hat{\xi}_{01}; \hat{\xi}_{v2}, \hat{\xi}_{02})}{\mathcal{F}(\lambda; \hat{\xi}_{v1}, \hat{\xi}_{01}; \hat{\xi}_{v3}, \hat{\xi}_{03})} &= \mathcal{F}(\lambda; \hat{\xi}_{v3}, \hat{\xi}_{03}; \hat{\xi}_{v2}, \hat{\xi}_{02}) = \frac{1}{\mathcal{F}(\lambda; \hat{\xi}_{v2}, \hat{\xi}_{02}; \hat{\xi}_{v3}, \hat{\xi}_{03})}; \quad \text{and} \\ \frac{\mathcal{F}(\lambda; \hat{\xi}_{v1}, \hat{\xi}_{01}; \hat{\xi}_{v3}, \hat{\xi}_{03})}{\mathcal{F}(\lambda; \hat{\xi}_{v2}, \hat{\xi}_{02}; \hat{\xi}_{v3}, \hat{\xi}_{03})} &= \mathcal{F}(\lambda; \hat{\xi}_{v1}, \hat{\xi}_{01}; \hat{\xi}_{v2}, \hat{\xi}_{02}) = \frac{1}{\mathcal{F}(\lambda; \hat{\xi}_{v2}, \hat{\xi}_{02}; \hat{\xi}_{v1}, \hat{\xi}_{01})} \end{aligned} \quad (13.6)$$

These will be useful in developing the validation procedures.

Let us now briefly consider the relevant radiometric measurements typically made above and/or below the water surface. There are very few instruments (probably fewer than 5 in 2018) capable of measuring $L_u(\lambda, \hat{\xi}'_v, \hat{\xi}'_0)$ for all $\hat{\xi}'_v$, so the full spectral radiance distribution is rarely measured. Rather, the most common radiance measurement is $L_u(\lambda, \hat{n}_S, \hat{\xi}'_0)$, where \hat{n}_S is the *upward* normal to the undisturbed water surface (radiometer aimed toward $-\hat{n}_S$, the nadir). The downward irradiance $E_d(\lambda, \hat{\xi}'_0)$ cannot be measured just beneath the water surface because of the wild fluctuations in its value (at the irradiance meter) due to surface waves. Thus, it is usually measured at depth (> 10 - 15 m) and extrapolated to determine its value just beneath the surface.³ These measurements and extrapolations then yield estimates of $R_{rs}(\lambda, \hat{n}_S, \hat{\xi}_0)$ or $[\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N$.

For radiometers attempting to measure $L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)$ from just above the surface using a down-looking radiometer to measure $L_{\text{Surf}}(\lambda, \hat{\xi}_v, \hat{\xi}_0)$, that portion of the measured radiance due to Sun glint and reflected skylight must be subtracted from the measured $L_{\text{Surf}}(\lambda, \hat{\xi}_v, \hat{\xi}_0)$ to obtain $L_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)$. However, since skylight and Sun glint can overwhelm L_w (as anyone looking down at the water in certain directions from a boat can attest), the viewing geometry must be such as to minimize the effects of these reflections. The protocols ("PC") for such measurements provide that the viewing direction $\hat{\xi}_{PC}$ should have components $(\theta_{PC}^{(c)}, \phi_{PC}) \approx (40^\circ \text{ to } 45^\circ, \phi_0 + 135^\circ)$ if possible, but always with $90^\circ \leq \phi_{PC} - \phi_0 \leq 180^\circ$. Therefore, the water-surface measurement can only provide $L_w(\lambda, \hat{\xi}_{PC}, \hat{\xi}_0)$. The measurement of $E_d^+(\lambda, \hat{\xi}_0)$ is relatively straightforward, needing only to be corrected for any perturbations (e.g., shadows, etc.) produced by structures near the irradiance meter, which must of course be designed and calibrated for operation in *air*. The surface or "ground truth" measurements then yield

$$[\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N \quad \text{or} \quad [\rho_w(\lambda, \hat{\xi}_{PC}, \hat{\xi}_0)]_N$$

for measurements made below and above the water surface, respectively, rather than the desired $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$.

³The irradiance E_d profile, i.e., $E_d(z)$, is measured and fit to $E_d(z) = E_d(z_{\text{Min}}) \exp[-K_d(z - z_{\text{Min}})]$ to determine K_d . Then $E_d(0) = E_d(z_{\text{Min}}) \exp(+K_d z_{\text{Min}})$, where z_{Min} is the minimum depth of reliable measurements.

13.2.2 How Do We Compare Surface and Sensor Measurements of $[\rho_w]_N$?

The space-borne estimate of the normalized water-leaving reflectance $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$ is made in a viewing direction that is almost always different from $\hat{\xi}_{PC}$ or \hat{n}_S . Thus, we must transform either the space-borne or the surface measurement of $[\rho_w]_N$ to the geometry of the other. Formally this can be done using Eq. (13.5). Let us transform the surface measurements to the space-view geometry.⁴ We have for the in-water measurement

$$\begin{aligned} N &= \left[\frac{\Re(\lambda, \hat{\xi}_v, \hat{\xi}_0)}{\Re(\lambda, \hat{n}_S, \hat{\xi}_0)} \right] \times \left\{ \frac{R_L(\lambda, \hat{\xi}'_v, \hat{\xi}'_0)}{R_L(\lambda, \hat{n}_S, \hat{\xi}'_0)} \right\} \times [\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N \\ &= \mathcal{F}(\lambda; \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, \hat{\xi}_0) \times [\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N, \end{aligned}$$

and for the above-water measurement

$$\begin{aligned} N &= \left[\frac{\Re(\lambda, \hat{\xi}_v, \hat{\xi}_0)}{\Re(\lambda, \hat{\xi}_{PC}, \hat{\xi}_0)} \right] \times \left\{ \frac{R_L(\lambda, \hat{\xi}'_v, \hat{\xi}'_0)}{R_L(\lambda, \hat{\xi}'_{PC}, \hat{\xi}'_0)} \right\} \times [\rho_w(\lambda, \hat{\xi}_{PC}, \hat{\xi}_0)]_N \\ &= \mathcal{F}(\lambda; \hat{\xi}_v, \hat{\xi}_0; \hat{\xi}_{PC}, \hat{\xi}_0) \times [\rho_w(\lambda, \hat{\xi}_{PC}, \hat{\xi}_0)]_N. \end{aligned}$$

The quantity on the left side of these equations is what needs to be compared to the estimate of $[\rho_w]_N$ by the sensor. The term in the square brackets in each equation is nearly constant and can be estimated with negligible error. In contrast, the term in the curly brackets can vary significantly with the viewing directions indicated. We can estimate the variation via the QSSA by recalling (Eq. (6.24))

$$R_L(\hat{\xi}'_v, \hat{\xi}'_0) = QSSA'_L(\hat{\xi}'_v, \hat{\xi}'_0) = \frac{\omega_0}{4\pi[1 - \omega_0(1 - B)]} \frac{P_r(\hat{\xi}'_0 \rightarrow \hat{\xi}'_v)}{u_{0w} + \mu'_v},$$

where, as in our usual notation, $u'_0 = \cos \theta_{0w}$ and $\mu'_v = \cos \theta_{vw}^c = -\cos(\pi - \theta_{vw})$.⁵ This gives, for example,

$$\left\{ \frac{R_L(\lambda, \hat{\xi}'_v, \hat{\xi}'_0)}{R_L(\lambda, \hat{n}_S, \hat{\xi}'_0)} \right\} = \frac{u_{0w} + 1}{u_{0w} + \mu'_v} \times \frac{P_r(\hat{\xi}'_0 \rightarrow \hat{\xi}'_v)}{P_r(\hat{\xi}'_0 \rightarrow \hat{n}_S)},$$

⁴So far we have assumed that the surface and the satellite measurement are made with the same $\hat{\xi}_0$ which will be the case for simultaneous measurements. However, for measurements that are separated in time by several hours, this will not be the case. Extension to contemporaneous measurements is immediate using Eq. (13.5) as long as one assumes that there is no change in the IOPs of the water during the time interval between the two measurements.

⁵The reader should recall that the subscript “(r)” on the scattering phase function in the QSSA stands for “remainder,” i.e., that remainder for the actual phase function of the medium after the forward peak is replaced by a Dirac delta function.

etc. This can give us an estimate of the variation of the term in the curly brackets with $\hat{\xi}_v$ and $\hat{\xi}_0$. As an example, assume the Sun is at the zenith, so $\hat{\xi}_0 = -\hat{n}_S$, then

$$\left\{ \frac{R_L(\lambda, \hat{\xi}_v', -\hat{n}_S)}{R_L(\lambda, \hat{n}_S, -\hat{n}_S)} \right\} = \frac{2}{1 + \mu_v'} \times \frac{P_r(-\hat{n}_S \rightarrow \hat{\xi}_v')}{P_r(-\hat{n}_S \rightarrow \hat{n}_S)}.$$

Figure 13.1 provides the quantity in the curly brackets as a function of the angle $\theta_v'^c = \pi - \theta_v'$ for the QSSA formula above. The phase function used in preparation of the figure is a combination of that of water and particles:

$$P_r = \frac{b_w P_w + b_p P_p}{b_w + b_p} = \frac{P_w + (b_p/b_w) P_p}{1 + (b_p/b_w)}.$$

Noting that $b_p/b_w = (B_w/B_p)[(b_b)_p/(b_b)_w]$, one can compute the BRDF effect as a function of the relative particle concentration by varying the ratio $(b_b)_p/(b_b)_w$, which for Case 1 waters is related to the Chlorophyll *a* concentration as provided in Table 13.1. The results in Figure 13.1, show a strong variation in the R_L ratio with both the viewing direction and with the contribution of particle backscattering relative to water. Clear ocean waters have $(b_b)_p/(b_b)_w$ less than unity and most mesotrophic waters have this ratio between 1 and 10. The variation with wavelength is caused largely by the strong spectral dependence of $(b_b)_w$. We expect that the QSSA results to be further modified by multiple scattering.

Table 13.1: The quantity $(b_b)_p/(b_b)_w$ as a function of λ (nm) and the concentration of Chlorophyll *a* (C) in mg/m^3 using the IOP model for $(b_b)_p$ in Chapter 5.

λ	$C = 0.01$	$C = 0.10$	$C = 1.00$	$C = 10.0$	$C = 100$
443	0.09	0.36	1.40	4.77	11.0
550	0.17	0.78	3.37	12.7	32.0
670	0.31	1.58	7.50	31.1	88.9

How much? Figure 13.2 provides the results of exact calculations for $(b_b)_p/(b_b)_w = 0.1$ (left panel) and for $(b_b)_p/(b_b)_w = 100$ (right panel), values that for the most part span the full range of this parameter in Figure 13.1. In the figure, the solid symbols refer to $\omega_0 = 0.1, 0.2, 0.4, 0.6, 0.8, 0.85$, and 0.9 . The open symbols at the very bottom are the result of the QSSA approximation from Figure 13.1. Several conclusions can be drawn from the Figures 13.1 and 13.2:

First, as the maximum value of $\theta_v'^c$ is about 48.6° for a flat surface, the results suggest that ignoring the BRDF effect, i.e., assuming $L_u(\hat{\xi}', \hat{\xi}_0')$ is independent of $\hat{\xi}'$, can result in an error in $[\rho_w(\lambda, \hat{\xi}_v', -\hat{n}_S)]_N$ computed from $[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N$ by as much as -14 to $+25\%$, which would severely limit the ability perform validation. (Recall, the original *goal* for the uncertainty

in $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$ in the blue was $\pm 5\%$ in waters with low Chlorophyll *a* concentration).

Second, if viewing angles θ_v^c are restricted to be $\lesssim 42^\circ$ ($\theta_v^c \leq 30^\circ$), the error in ignoring the BRDF effect altogether can be reduced to $\sim \pm 5\%$.

Third, for waters with $\omega_0 \lesssim 0.4$, the QSSA estimation can reduce this error to about 2%.

Fourth, to reduce the error to acceptable limits, i.e., to $\sim 1 - 2\%$, requires significant knowledge of the IOPs of the medium, so one could carry out detailed radiative transfer computations to properly assess the effect (e.g., of the kind leading to Figure 13.2). This is true even when the QSSA is appropriate for the assessment, because the scattering phase function of the water *plus* particles is required at the appropriate scattering angles — here, $\Theta = \arccos(\hat{\xi}_v' \bullet \hat{n}_S)$ and $\Theta = 180^\circ$.⁶

As the full sub-surface upwelling radiance distribution is rarely measured, we are led to the inescapable conclusion that the BRDF effect must be modeled to complete a validation that can reveal the true performance of the in-orbit sensor, and that the IOPs (or enough of the IOPs) at the validation site must be measured along with the water-leaving reflectance to carry out the modeling.⁷ In practice, there are practically no validation experiments with as comprehensive a suite of measurements as required to assess the BRDF effect. Therefore, the BRDF effect must be determined through *models* based on a combination of the surface and the satellite measurements. Such measurements usually provide incomplete estimates of the needed IOPs. The efficacy of such models will likely be a limiting factor in the accuracy of the validation process.

Considering the fact that in the above example we had to transform $[\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N$ to a different viewing angle, i.e., to form $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$, to carry out a validation, and that because such transformations are (at present) always necessary, one might wonder if there is some optimum set of transformations for carrying out such a validation. For example, SeaWiFS and MODIS report the radiometric output product as $[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N^{\text{Sat}}$.⁸ This

⁶In Chapter 6 Section 6.8 we examined two alternatives (beyond the QSSA) to detailed radiative transfer calculations that might be used to assess the BRDF effect (the Zaneveld approximation in Section 6.8 and the modified δ -Isotropic approximation in Section 6.9); however, both still require knowledge of the scattering phase function.

⁷The same comments apply to the effect of the BRDF in a vicarious calibration exercise. If a procedure such as this is not followed at the calibration site, then any calibration adjustment based in-water nadir radiance measurements will be accompanied by a larger uncertainty than necessary. Because the surface data requirements for calibration and validation are similar, such exercises are usually referred to as “cal/val” field experiments, meaning they can apply equally well to calibration or validation.

⁸The associated $[L_w(\hat{n}_S, -\hat{n}_S)]_N$ is the actual $L_w(\hat{n}_S, -\hat{n}_S)$ that would be extant if the atmosphere were

is derived from the retrieved quantity $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Sat}}$ through

$$[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N^{\text{Sat}} = \mathcal{F}(\lambda; \hat{n}_S, -\hat{n}_S; \hat{\xi}_v, \hat{\xi}_0) \times [\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Sat}},$$

They then compare this with $[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N^{\text{Surf}}$, which is derived from the measured quantity $[\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N^{\text{Surf}}$ by

$$[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N^{\text{Surf}} = \mathcal{F}(\lambda; \hat{n}_S, -\hat{n}_S; \hat{n}_S, \hat{\xi}_0) \times [\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N^{\text{Surf}},$$

So to compare the two, we consider

$$\frac{[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N^{\text{Sat}}}{[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N^{\text{Surf}}} = \frac{\mathcal{F}(\lambda; \hat{n}_S, -\hat{n}_S; \hat{\xi}_v, \hat{\xi}_0)}{\mathcal{F}(\lambda; \hat{n}_S, -\hat{n}_S; \hat{n}_S, \hat{\xi}_0)} \times \left[\frac{[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Sat}}}{[\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N^{\text{Surf}}} \right], \quad (13.7)$$

where the measured quantities are in the square brackets; one by the sensor in space (“Sat”) and one by the in-water sensor (“Surf”). What if instead, we transformed the surface measurement to the spacecraft’s geometry, i.e.,

$$[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Surf}} = \mathcal{F}(\lambda; \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, \hat{\xi}_0) \times [\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N^{\text{Surf}}.$$

Then

$$\frac{[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Sat}}}{[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Surf}}} = \frac{1}{\mathcal{F}(\lambda; \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, \hat{\xi}_0)} \times \left[\frac{[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N^{\text{Sat}}}{[\rho_w(\lambda, \hat{n}_S, \hat{\xi}_0)]_N^{\text{Surf}}} \right], \quad (13.8)$$

Equation (13.6) shows that the terms with \mathcal{F} factors are the same in both equations, which means that the same BRDF ratio is required for both comparisons. Thus, it is safe to conclude that there is no advantage of one set of $[\rho_w]_N$ comparisons over any other in terms of improving on the uncertainty introduced by the (unknown) BRDF.

13.2.3 Validation of Derived Products.

As mentioned earlier, it is important to validate the principal derived products (i.e., estimate the uncertainty associated with the determination from the space-borne system). For the CZCS this was C_P , while for the follow-on sensors SeaWiFS and MODIS, it was C . As with $[\rho_w]_N$, the idea is to compare surface measurements with satellite sensor measurements. This is usually done at a single location, where ship-borne determinations are carried out. One can also make continuous measurements along the ship’s track and compare these with the satellite measurement, understanding that there can be a significant time difference between the ship and satellite measurement. Usually, only surface measurements of C or C_P are required, as the depth of the mixed layer is usually below the

 absent and the Sun were at the zenith (Chapter 7).

depth of penetration (Chapter 6, Section 6.4.10). Because the retrieval involves the use of a bio-optical algorithm that relates the product to $[\rho_w]_N$, one must require that the protocols used in measuring the surface validation product are the same as those in place for producing the bio-optical algorithm, although this may not always be possible.

13.3 Validation of CZCS and MODIS

Here we describe the very different validation experiences for CZCS and MODIS. For the proof-of-concept mission, CZCS, the validation was more of a verification, i.e., showing that the concept worked and providing a path forward to future instruments. For the mature MODIS sensor, the validation was built on the SeaWiFS experience, for which extensive, carefully quality-controlled, surface data were acquired using well-developed measurement protocols. The amount of contemporaneous surface and sensor data for validating MODIS was close to two orders of magnitude more than for CZCS. As such, in the case of MODIS, the uncertainty in both $[\rho_w]_N$ and C potentially could be estimated for a variety of waters.

13.3.1 Verification of CZCS Concept

Before describing the CZCS verification procedure, it is useful to recall some details of the CZCS atmospheric correction. The atmospheric correction algorithm for CZCS was based on single scattering. Because there were no NIR bands on CZCS the algorithm could not be operated as described in Chapter 10; however, $\rho_w(670)$ can generally be taken to be zero, if the pigment concentration (C_P) is sufficiently low, which is valid in oligotrophic and mesotrophic ocean waters. Thus, the single scattering algorithm was typically operated with $\lambda_l = 670$ nm, using the assumption that $\rho_w(\lambda_l) = 0$. On CZCS, there was no shorter wavelength (λ_s) for which $\rho_w = 0$; however for $C_P \lesssim 0.25$ mg/m³, $[\rho_w(550)]_N$ is approximately a known constant (the “clear-water radiance concept,” Figure 9.20), so $t^*\rho_w(550)$ can be estimated and subtracted from $\rho_t(550) - \rho_r(550)$ to yield $\rho_{as}(550)$. This was used along with $\rho_{as}(670)$ to estimate $\varepsilon(550, 670)$ for “clear water” regions in a scene, i.e., regions for which $C_P \lesssim 0.25$ mg/m³, and provided a basis for extrapolation to 520 and 443 nm by assuming the ε varies in a known manner with λ , e.g., $\varepsilon(\lambda, \lambda_l) = (\lambda_l/\lambda)^n$.⁹ Then assuming that the resulting $\varepsilon(\lambda, \lambda_l)$ was valid for the entire image, retrieval of $[\rho_w(\lambda)]_N$ at 443, 520, and 550 could be effected for the entire image, even those areas that do not qualify

⁹In Chapter 10 an exponential variation of $\varepsilon(\lambda, \lambda_l)$ was found for the Shettle and Fenn aerosol models; however, prior to those models, aerosols were often characterized by Junge size distributions (Chapter 4) for which $\varepsilon(\lambda, \lambda_l) \approx (\lambda_l/\lambda)^n$. The exponent n is related to the Junge size distribution parameter ν through $n = \nu - 2$.

as “clear water.”¹⁰ Evidence that the latter assumption was viable was provided in scenes where the “aerosol radiance,” $L_a(670) \equiv L_t(670) - L_r(670)$ varied considerably, while the associated $\varepsilon(\lambda, 670)$ showed significantly less variation. An example is provided in Figures 13.3 and 13.4. Figure 13.3 shows $L_a(670)$ and $\varepsilon(520, 670)$ for a CZCS image obtained in the Gulf of Mexico off the Mississippi Delta. Note that there is significant variation in $L_a(670)$ but much less variation in $\varepsilon(520, 670)$.¹¹ Quantification of the variations is provided in Figure 13.4, suggesting that over the boxed area in Figure 13.3, where $L_a(670)$ varied by roughly $\pm 20\%$, $\varepsilon(520, 670)$ varied by only about $\pm 5\%$, and some of the $\varepsilon(520, 670)$ variation is the result of variations in $L_w(520)$, which were unaccounted for here. This atmospheric correction procedure was used in the verification studies.¹²

For CZCS there was very little surface data to effect either a vicarious calibration or a validation. The main reason for this was the existence of only a small number of radiometers ($\lesssim 5$) capable of measuring $L_u(\lambda)$. All that could be done with the small quantity of surface data was to verify that the system — sensor *plus* algorithms — performed as anticipated. This meant that with scenes for which surface measurements were available, the surface data were first used to effect a crude vicarious calibration at the ship location, and then the “calibrated” radiances were used to show, that when applied to the complete images, acceptable results were obtained over several days. This is what we mean by *verification*.

Based on the above discussion, the verification procedure was predicated on four assumptions:

- $L_w(670) = 0$ over the portion of interest in the image;
- $\varepsilon(\lambda, \lambda_l)$ is constant over an image;
- for the open ocean far from land, $\varepsilon(\lambda, \lambda_l) = (\lambda_l/\lambda)^n$; and
- the pre-launch CZCS calibration at 670 nm was perfect.

The last assumption was required, as at that time there was no way to effect vicarious calibration of the band at 670 nm, so in effect, the other bands had to be calibrated relative to the band at 670 nm.

¹⁰Even with these assumptions, it is still a requirement that $[\rho_w(670)]_N \approx 0$ for proper operation of the algorithm.

¹¹The $\varepsilon(520, 670)$ image is highly “stretched,” which magnifies any variation in $\varepsilon(520, 670)$ compared to that in $L_a(670)$.

¹²Actually, when $C_P \lesssim 0.25 \text{ mg/m}^3$, $[\rho_w(520)]_N$ is also approximately constant (Figure 9.20), allowing $\varepsilon(520, 670)$ to be estimated in the same manner. Taking $\varepsilon(\lambda, \lambda_l) = (\lambda_l/\lambda)^n$, the values of n at $\lambda = 520$ and 550 nm were averaged to provide n and determine $\varepsilon(443, 670)$.

A sequence of images was examined in the Northwestern Sargasso Sea near the Gulf Stream for which contemporaneous surface measurements of $[L_w]_N$ had been carried out.¹³ For much of the area of these images, the first assumption, i.e., $L_w(670) = 0$, was justified. A portion of one of these images is shown in Figure 10.3 in Chapter 10. The surface measurements were used to estimate $L_a(\lambda) = L_t(\lambda) - L_r(\lambda) - t^*[L_w(\lambda)]_N$ and then form $\varepsilon(\lambda, 670)$. The calibration constants of the sensor at 443, 520, and 550 nm were then adjusted by trial and error using sensor data over clear water areas of the images to yield values of $\varepsilon(\lambda, 670)$ which were stable from day to day and not unreasonable in their dependence on λ . For these, $n \approx 0.75$. Although one might expect $n \approx 0$ for a pure maritime aerosol, i.e., one that is locally generated, it was felt that $n \approx 0.75$ was not unreasonable, as the area is subjected to some fraction of continental aerosol from the U.S. East Coast. The factors by which the pre-launch calibration constants needed to be multiplied by to produce the desired results were 1.069, 0.993, and 0.955, respectively, for the spectral bands at 443, 520, and 550 nm. Using these calibration adjustments, the mean absolute errors in the retrieved $[L_w]_N$ for the three orbits used in the calibration adjustment (3157, 3171, and 3240) were 7.4, 6.1, and 11.8% at 443, 520, and 550 nm, respectively. For these, $[L_w]_N \approx 0.6, 0.5$, and $0.3 \text{ mW/cm}^2\mu\text{m Sr}$ or $[\rho_w]_N \approx 0.01, 0.008, 0.005$ at 443, 520, and 550 nm, respectively, and $L_a(670) \approx 0.2 \rightarrow 0.45 \text{ mW/cm}^2\mu\text{m Sr}$ or $\rho_a(670) \approx 0.004 \rightarrow 0.0093$. The values of $\rho_a(670)$ are similar to those shown in Figure 10.4, and those of $[\rho_w]_N$ similar to those in Figure 9.20 for low-Chlorophyll waters.

There were no other scenes in this time frame with which to test the quality of the derived calibration coefficients. However, during the field experiments, continuous measurements of C_P were carried out. Thus, we could examine the performance of the entire system by comparing the surface and CZCS measurements of C_P along the ship track. The ship track data examined covered a two-day period that spanned two CZCS images (Orbits 3151 and 3171). The track of the ship is shown as the white line on the left panel of Figure 13.5, which is an image of C_P from Orbit 3226 over the Northwestern Sargasso Sea and the Gulf Stream (unfortunately, the ship was in port at the time of this image). The image of C_P is in false color (blue \rightarrow green \rightarrow red in the direction of increasing C_P). The right panel in the figure compares the CZCS-retrieved values of C_P (thick line) along the ship track with the ship measurements (thin line) along the same track. The agreement between the two is excellent over most of the track, but degrades in the water over the Continental Shelf (however, it is still within a factor of 2). There are many reasons for the disagreement nearer the coast: (1) the assumption $L_w(670) = 0$ breaks down as one approaches the coast; (2) the bio-optical algorithm becomes less accurate as the water becomes more turbid; (3) the aerosol may have a more continental character resulting in a gradual change in $\varepsilon(\lambda, 670)$ as one approaches the coast; and (4) the advection of water features can change C_P considerably over the time period between the CZCS image acquisition and surface measurements. However, the

¹³These images were from Orbits 3157, 3171, 3226, 3240, and 3351.

comparisons do suggest that the system worked as advertised, and thus would constitute a verification of the CZCS concept.

Unfortunately, when the vicarious calibration above was used to process the few images with contemporaneous surface data obtained considerably earlier than those near Orbit 3200, it was found that the retrieved $[L_w]_N$ values at 443 nm were too high by many tens of percent. This was attributed to a change in sensitivity of the CZCS with time — the CZCS-determined radiances in the blue were too high compared to those in the Orbit 3200 time frame. Little or no additional error was detected in the green bands, although the use of “clear water” areas to determine $\varepsilon(\lambda, 670)$ would mask the effect of any decay in the sensitivity at these wavelengths.

In order to accurately assess any variation in the sensitivity of CZCS with time a significant quantity of surface data would be required, but there was little or no possibility that such would become available. This paucity of surface data was overcome by again using the “clear water radiance” concept. In the same area as shown in Figure 13.5, regions of sufficiently low C_P were identified, so $[\rho_w]_N$ could be accurately estimated. For these, a value of $C_P = 0.10 \text{ mg/m}^3$ was assumed to fix the value of $[\rho_w(443)]_N$ at 0.025 (Figure 9.20). For $0.05 \leq C_P \leq 0.25 \text{ mg/m}^3$ the values of $[\rho_w(520)]_N$ and $[\rho_w(550)]_N$ can be taken to be 0.008 and 0.005, respectively, and a precise value of C_P is not needed. For these areas the atmospheric correction algorithm was operated with values of $\varepsilon(\lambda, 670)$ fixed by choosing $n = 0.75$ and the calibration was again adjusted to result in the given “clear water” values for $[\rho_w(\lambda)]_N$. Specifically, let the reflectance of the ocean-atmosphere system for Orbit Number N , ρ_t^{Calc} , be the reflectance calculated using the adjustments determined near $N = 3200$. Then relate this to the true reflectance for Orbit Number N , ρ_t^{True} , through

$$\rho_t(\lambda)^{\text{Calc}} = f(N, \lambda) \rho_t(\lambda)^{\text{True}},$$

where $f(N, \lambda)$ is the time-varying sensitivity of the sensor. Then,

$$[f(N, \lambda)]^{-1} \rho_t(\lambda)^{\text{Calc}} = \rho_r(\lambda) + \varepsilon(\lambda, 670) \left(\rho_t(670) - \rho_r(670) \right) + t^*(\lambda) [\rho_w(\lambda)]_N.$$

Since ρ_r can be accurately computed, if ε and $[\rho_w]_N$ are provided, the only unknown in the above equation is f at each λ . Figure 13.6 provides the resulting $f(N, \lambda)$ estimated in the manner described above. As we have taken ρ_t^{True} to be the values computed using the vicarious calibration determined for Orbit Numbers around 3200, $f(3200, \lambda) = 1$. The figure shows a significant decay in the sensitivity of the blue band, with detectable, but considerably smaller, decays in the green and yellow bands.

In the light of the assumptions employed in estimating $f(N, \lambda)$, it is important to try and assess its uncertainty. Taking the atmospheric correction algorithm that was used to be exact, the contributors to the uncertainty in f are the uncertainties in C_P , which causes

uncertainty in $[\rho_w(443)]_N$, and the uncertainty in $\varepsilon(\lambda, 670)$ (or equivalently in n). Let Q be $t^*[\rho_w]_N$ or $\varepsilon(\lambda, 670)(\rho_t(670) - \rho_r(670))$. Then the error in f is related to the error in Q though, $\Delta f^{-1} = \Delta Q / \rho_t^{\text{Calc}}$ or $\Delta f = -f^2 \Delta Q / \rho_t^{\text{True}}$, where to the accuracy required we have replaced ρ_t^{Calc} by ρ_t^{True} . Thus,

$$\Delta f_{C_P} = f^2 t^* \frac{\Delta[\rho_w]_N}{\rho_t^{\text{True}}} \quad \text{and} \quad \Delta f_\varepsilon = f^2 \frac{\rho_a(670)}{\rho_t^{\text{True}}} \Delta \varepsilon,$$

where $\rho_a(670) = \rho_t(670) - \rho_r(670)$, for the error in f due to error in C_P and ε , respectively. For images with $0.05 \leq C_P \leq 0.25 \text{ mg/m}^3$, for which $C_P = 0.10 \text{ mg/m}^3$ was assumed, $\Delta[\rho_w(443)]_N \approx \pm 0.005$, while for the other bands the error is negligible (Figure 9.20). Minimum values of $\rho_t(443)$ are ~ 0.13 (Figure 10.4), and choosing nominal values for f and t^* (1 and 0.88, respectively), gives $\Delta f_{C_P}(443) \approx 0.034$, with negligible Δf_{C_P} in the other bands. The points for $f(N, 443)$ in Figure 13.6 labeled with a (*), which are for the summer months, are expected to have this error. For imagery acquired in the spring and labeled by (+) the uncertainty of C_P was much larger, i.e., $0.05 \leq C_P \leq 0.40 \text{ mg/m}^3$, so the assumption of $C_P = 0.10 \text{ mg/m}^3$ was likely too small and the associated value of $[\rho_w(443)]_N$ too large, perhaps by as much as 0.005. Thus, for these points, there is a bias due to the too small $[\rho_w(443)]_N$ and an increased uncertainty in $[\Delta \rho_w(443)]_N$ to about ± 0.007 . Thus, these points marked (+) should probably be moved upward by about 0.034 and have associated errors of the order of ± 0.05 .

In the case of error in the estimated ε , note that we chose $n = 0.75$, where for a pure maritime aerosol n might be assumed to be closer to zero (Chapter 4). Thus, we might legitimately assume that $\Delta \varepsilon(\lambda, 670) = \varepsilon(\lambda, 670) - 1$, i.e., we chose $n = 0.75$, but the correct value was $n = 0$. Imagery was specifically chosen with $\rho_a(670) < 0.008$, and taking $\rho_t(\lambda)$ from Figure 10.4, we find $\Delta f_\varepsilon(443) \approx -0.02$, $\Delta f_\varepsilon(520) \approx -0.02$ and $\Delta f_\varepsilon(550) \approx -0.025$. Therefore, error in ε might cause the points in Figure 13.6 to move vertically up by as much as 0.020-0.025, but probably not downward. Clearly, the uncertainties in f due to uncertainty in C_P and ε would not change the conclusion that the CZCS underwent a slow decay in radiometric sensitivity. When this decay equation was implemented into the data processing, water-leaving reflectances and C_P were retrieved with accuracies similar to those near Orbit 3200.

A complete study of the radiometric performance of the CZCS, as revealed by vicarious calibration effected as described above (but with $n = 0$), was carried out dividing the data into ten day intervals through the entire mission. That study indicated that the green and yellow bands begin to show random variations in sensitivity at the several digital count level starting at about Orbit 14,000 (Fall 1981). These variations were correlated between the two bands. There was no way to determine if the changes were also present in the blue and red bands.

Finally, although the uncertainty in the pre-launch radiometric calibration of CZCS was basically unknown,¹⁴ based on the results of this vicarious calibration exercise and extrapolating the sensor decay to $N = 0$, it appears that it was of the order of 5-16% (but note that it was assumed that there was *no* uncertainty in the 670 nm band). This suggests that ocean color sensors can be made to function well even in the absence of a pre-launch calibration with low uncertainties. This is a testimony to the viability of vicarious calibration.

In sum the comparisons (which might be called “bootstrap”¹⁵ comparisons) between surface data (radiance and C_P) suggested that, while the sensor was far from ideal for the application, the CZCS concept of studying the dynamics of phytoplankton biomass from space was viable. This led proponents to advocate the development of a new satellite ocean color program based the lessons learned from the CZCS experience. Let us delineate some of these lessons.

- The CZCS could only effect accurate atmospheric correction for waters for which $C_P \leq 0.25 \text{ mg/m}^3$, so the first, and most obvious lesson was the requirement for spectral bands in the NIR for atmospheric correction. In fact, for very turbid water, spectral bands in the atmospheric “windows” at 1200, 1600, and 2100 nm, the SWIR, are desirable. (The desirability of NIR bands was actually known before launch.)
- The fact that most of the radiance is due to backscattering from the atmosphere, a follow-on sensor should have lower noise and a higher digitization rate than CZCS (e.g., 10 bit or more instead of 8). (Note that at 550 nm, the entire range of variation in $[\rho_w(550)]_N$ shown in Figure 9.20 corresponded to only about 12 CZCS digital counts.).
- Because pre-launch calibration cannot be carried out with accuracy sufficient for proper operation of the algorithms, and even if it could, there is no guarantee that the calibration would be maintained through the violence of launch, vicarious calibration is *required*.
- Although high-quality pre-launch calibration is not essential, determi-

¹⁴The final report describing the design, fabrication and testing of CZCS [Ball Bros Report V2.] did not specify the calibration uncertainty. However, it did document the linearity of the output with input radiance, the noise characteristics of each spectral band as a function of the incident radiance and the spectral sensitivity $S(\lambda)$ (Chapter 12) of each spectral band.

¹⁵This refers to the old adage that one can lift oneself up by one’s bootstraps. Here it refers to the fact that the data used for vicarious calibration and for validation were essentially one and the same. In addition, these data were used in developing the bio-optical algorithms relating C_P to $[\rho_w]_N$, as well.

nation of the sensor characteristics such as polarization sensitivity, linearity to radiance (or characterizing non-linearity) of the detectors, signal-to-noise ratio, spectral response, etc., is vital, as these cannot be determined vicariously.

- Because the radiometric sensitivity of the CZCS was found to decay with time, vicarious calibration should be carried out frequently throughout the mission. In addition, in the design of future instruments, consideration should be given to ways to ensure and monitor the stability of the sensor. Stability is far more important than prelaunch calibration accuracy. Stability monitoring methods include viewing the Sun in diffuse reflection or providing a mechanism for Lunar views (or both).
- Vicarious calibration and validation both require a large quantity of surface radiance data and, for validation, derived product data as well (e.g., C_P and now C). The difficulty and expense of collecting such data means that it must be a community effort, not just that of a small number of investigators. As such, strict protocols for collection of surface data must be developed and a quality controlled database of such data maintained.

All of these lessons were in one way or another actually incorporated into the four follow-on U.S. ocean color missions — SeaWiFS, MODIS/Terra, MODIS/Aqua and VIIRS.

13.3.2 Validation of MODIS Performance

The validation of the MODIS sensor data provides an example of how accurate one can expect retrievals of $[\rho_w]_N$ and the concentration of Chlorophyll a with a well-designed ocean color system. With few exceptions, the procedures described here were also applied to the validation of SeaWiFS.

For SeaWiFS and MODIS a dedicated vicarious calibration site was established in the waters off Lanai in the Hawaiian Islands by the U.S. National Oceanic and Atmospheric Administration (NOAA). At this site a mooring with an attached scientific buoy — the Marine Optical Buoy or MOBY — was operated continually since July 1997 to the present (2019). Development of MOBY began in 1985 as a direct result of the lessons from the CZCS experience, and in anticipation of a follow-on sensor to CZCS. Figure 13.7 provides a schematic of the buoy. The principal sensors on MOBY measure the upwelling spectral radiance propagating toward the zenith, $L_u(\lambda, \hat{n}_s, \hat{\xi}'_0)$ and the downwelling irradiance

$E_d(\lambda, \hat{\xi}'_0)$ at three depths (1, 5, and 9 m), which provides the attenuation coefficients required for propagating the 1 m measurement to the water surface. The spectral resolution is of the order of 1 nm, allowing MOBY spectra to be combined with the sensor's spectral response function to produce the *in-band* normalized water-leaving reflectance. The total spectral range of MOBY, 340 to 955 nm, is covered by two spectrometers (called “MOS System” in Figure 13.7). The sensors are on arms that extend up to 3 m from the buoy's core to avoid shadowing by its surface structure. There is a sensor above the surface to measure the incident irradiance from the Sun and sky, $E_d^+(\lambda, \hat{\xi}_0)$. The radiometers are believed to have a calibration uncertainty of the order of 3%, derived through NIST-traceable procedures. MOBY provides surface vicarious calibration data for every sensor overpass of the Hawaiian Islands, except when clouds or Sun glint contaminate the sensor imagery.

In preparation for the launch of the CZCS follow-on sensors, NASA developed several programs at Goddard Space Flight Center (GSFC) in the area of data processing and quality control. The main activities were carried out by the Ocean Biology Processing Group (OBPG), which provided support for calibrating, validating, processing, archiving and distributing ocean color products, e.g., $[\rho_w]_N$, etc. As part of the OBPG effort was to develop a system for archiving data collected in support of the vicarious calibration and validation effort: the “SeaWiFS Bio-optical Archive and Storage System” (SeaBASS). In a parallel effort a NASA program, the “Sensor Intercomparison for Marine Biological and Interdisciplinary Ocean Studies (SIMBIOS),” developed a set of protocols for surface measurements made in support of vicarious calibration and validation. Following these protocols was required of investigators in order that the surface data they collected be included in the SeaBASS archive. This provided a high level of quality control. The OBPG also led the development of the “SeaWiFS Data Analysis System” (SeaDAS) for processing, display and analysis of ocean color data, i.e., a total end-to-end processing system, and the “NASA bio-Optical Marine Algorithm Dataset” (NOMAD) a high quality global in situ bio-optical data set for use in algorithm development and sensor validation. Thus, a complete structure for calibrating (vicariously), validating, and processing ocean color data was assembled for the CZCS successors. Next, we summarize the results of the vicarious calibration and validation with examples from MODIS and SeaWiFS.

First, we describe the vicarious calibration procedure that was used, as it is central to the validation effort. The familiar equation relating $\rho_t(\lambda, \hat{\xi}, \hat{\xi}_0)$ to $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$ is

$$\rho_t(\lambda, \hat{\xi}_v, \hat{\xi}_0) = \rho_r(\lambda, \hat{\xi}_v, \hat{\xi}_0) + \rho_A(\lambda, \hat{\xi}_v, \hat{\xi}_0) + t(\lambda, \hat{\xi}_v)[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N,$$

where we have used the usual abbreviation ρ_A for the combination $\rho_a + \rho_{ra}$ and assumed that proper account has been taken of the contributions of Sun glint and whitecaps. We take it as given that the band at λ_l in the NIR is perfectly calibrated. Then the band at λ_s in the NIR can be calibrated by choosing sites in the open ocean far from land where the aerosol is totally maritime and stable from clear day to clear day. For such, using

the presumably-known long-term aerosol properties at that site, the quantity $\varepsilon(\lambda_s, \lambda_l)$ is estimated. Then, the reflectance formed from

$$\rho_t^{\text{VC}}(\lambda_s, \hat{\xi}_v, \hat{\xi}_0) = \rho_r(\lambda_s, \hat{\xi}_v, \hat{\xi}_0) + \varepsilon(\lambda_s, \lambda_l)\rho_A(\lambda_l, \hat{\xi}_v, \hat{\xi}_0),$$

where the superscript “VC” stands for vicarious calibration, and $\rho_t^{\text{VC}}(\lambda_s, \hat{\xi}_v, \hat{\xi}_0)$ is taken to be the true reflectance. This value of $\rho_t^{\text{VC}}(\lambda_s, \hat{\xi}_v, \hat{\xi}_0)$ is then compared to that computed using the original sensor calibration (“OC”) combined with any long-term change as indicated by the on-board devices and/or Lunar views, $\rho_t^{\text{OC}}(\lambda_s, \hat{\xi}_v, \hat{\xi}_0)$, and the ratio

$$g(\lambda_s) = \frac{\rho_t^{\text{VC}}(\lambda_s, \hat{\xi}_v, \hat{\xi}_0)}{\rho_t^{\text{OC}}(\lambda_s, \hat{\xi}_v, \hat{\xi}_0)},$$

is formed. The quantity $g(\lambda_s)$ is called the “vicarious gain” at λ_s . It is the factor that the original calibrated reflectance ρ_t^{OC} must be multiplied by to produce what is believed to be the correct reflectance, ρ_t^{VC} . This process is carried out over a long time period (and wide range of $\hat{\xi}_v$ and $\hat{\xi}_0$) resulting in an average value $g(\lambda_s)$. The averaging process is important as the estimated values of ε , i.e., the aerosol model chosen for the site, may only be correct in an average sense. It is noteworthy that this process only requires an area for which $[\rho_w(\lambda_s, \hat{\xi}_v, \hat{\xi}_0)]_N$ is known to be negligible (oligotrophic waters) and the atmosphere guaranteed to be maritime (central ocean gyres in the Southern Hemisphere).

For the visible bands, the $[\rho_w(\lambda)]_N$ data from MOBY are used to estimate vicarious gain coefficients. Specifically, the atmospheric correction algorithm is operated at the MOBY site, utilizing the already-estimated the NIR vicarious gain $g(\lambda_s)$, to derive $\varepsilon(\lambda, \lambda_l)$ and then $\rho_A(\lambda, \hat{\xi}_v, \hat{\xi}_0) = \varepsilon(\lambda, \lambda_l)\rho_A(\lambda_l, \hat{\xi}_v, \hat{\xi}_0)$ yielding $\rho_t^{\text{VC}}(\lambda, \hat{\xi}_v, \hat{\xi}_0)$ given by

$$\begin{aligned} \rho_t^{\text{VC}}(\lambda, \hat{\xi}_v, \hat{\xi}_0) &= \rho_r(\lambda, \hat{\xi}_v, \hat{\xi}_0) + \varepsilon(\lambda, \lambda_l)\rho_A(\lambda_l, \hat{\xi}_v, \hat{\xi}_0) \\ &\quad + t(\lambda, \hat{\xi}_v)\mathcal{F}(\lambda, \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, \hat{\xi}_0)[\rho_w(\lambda, \hat{n}_S, \theta_0)]_N^{\text{MOBY}} \end{aligned} \quad (13.9)$$

where $t(\lambda, \hat{\xi}_v)$ is actually estimated by $t^*(\lambda, \hat{\xi}_v)$. The quantity $\mathcal{F}(\lambda, \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, \hat{\xi}_0)$ is estimated from look-up tables based on radiative transfer computations with the IOPs of the water modeled as a function of the concentration of Chlorophyll *a*. The whole process sounds simple, but it is not. Over a period of nine years, MOBY provided 1450 samples of contemporaneous data with SeaWiFS, which after quality control yielded 150 match ups, or roughly 15 per year. It took averaging of the order of 40 match ups before the vicarious gain factors $g(\lambda)$ became stable, i.e., about 2.5 years of data. This shows the time and effort (and money) that must go into what, on paper, seems to be a relatively straightforward procedure.

One must realize that the procedure described above is a vicarious calibration of the entire system — sensor *plus* atmospheric correction algorithm: the sensor is being *forced*

to provide the top-of-atmosphere reflectance that the atmospheric correction algorithm requires. If any changes in the algorithm are made, e.g., using a revised set of candidate aerosol models, a better approximation to t , or a different model to compute \mathcal{F} , the above procedure must be redone.

Upon completion of the vicarious calibration, validation can be carried out by using the vicarious gains to examine an independent (from MOBY) set of images for which surface truth is available, i.e.,

$$[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N = [t(\lambda, \hat{\xi}_v)]^{-1} \left(g(\lambda) \rho_t^{\text{OC}}(\lambda, \hat{\xi}_v, \hat{\xi}_0) - \rho_r(\lambda, \hat{\xi}_v, \hat{\xi}_0) - \varepsilon(\lambda, \lambda_l) \rho_A(\lambda_l, \hat{\xi}_v, \hat{\xi}_0) \right).$$

The available surface reflectance data are usually in the form of $[\rho_w(\lambda, \hat{n}_S, \hat{\xi}'_0)]_N$ or $[\rho_w(\lambda, \hat{\xi}_{PC}, \hat{\xi}'_0)]_N$, where we take the solar beam, $\hat{\xi}'_0$, for the surface measurement in a direction different from $\hat{\xi}_0$,¹⁶ as the surface and satellite data are usually contemporaneous rather than simultaneous. So, to effect a comparison requires estimation of $\mathcal{F}(\lambda, \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, \hat{\xi}'_0)$ or $\mathcal{F}(\lambda, \hat{\xi}_v, \hat{\xi}_0; \hat{\xi}_{PC}, \hat{\xi}'_0)$. These are determined using the same lookup tables based on C as in vicarious calibration. This is done with an iterative procedure as follows. The Chlorophyll concentration is usually determined using the reflectances $[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N$ in the blue and green spectral regions (see below); however, in the first iteration it is determined using $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$. This value of C is then used in the lookup tables to determine $\mathcal{F}(\lambda, \hat{\xi}_v, \hat{\xi}_0; \hat{n}_S, -\hat{n}_S)$, to find $[\rho_w(\lambda, \hat{n}_S, -\hat{n}_S)]_N$, and a new value of C , etc. Through this procedure the retrieved $[\rho_w(\lambda, \hat{\xi}_v, \hat{\xi}_0)]_N$ can be transformed to the Sun-viewing geometry of the surface data, or vice versa.

Figure 13.8 provides the remote sensing reflectance $R_{rs}(\lambda)$ match ups at 443 nm and 555 nm using the above procedure for MODIS/Aqua over a wide range of R_{rs} values.¹⁷ This was the result of 589 match ups at 443 nm and 429 at 555 nm. For these results, the surface and satellite data were considered contemporaneous if they were no more than ± 3 hours apart. There is no stratification of the data here between Case 1 and Case 2 waters; however, the three outliers (high in-situ but low MODIS) at 555 nm come from Case 2 coastal waters, for which adequate atmospheric correction was likely not achieved. For other coastal areas which one expects are Case 2, the correction appears adequate. The mean absolute error in R_{rs} at 443 and 555 nm was 0.00091 and 0.00064, respectively (or 0.0029 and 0.0020 in $[\rho_w]_N$). There are some retrieved values of $R_{rs}(443)$ that are less than zero. These are associated with already low values of R_{rs} and usually result from inadequate atmospheric correction due to the presence of absorbing aerosols (Chapter 10). The comparisons suggest that the error is somewhat higher than that expected from atmospheric correction (the largest source of uncertainty in the processing stream); however,

¹⁶Here, $\hat{\xi}'_0$ is not related to $\hat{\xi}_0$ by Snell's law as they were earlier. They are simply different directions of the solar beam propagating *above* the surface.

¹⁷Recall, $[\rho_w]_N = \pi R_{rs}$.

these data were not screened to ensure that all of the assumptions underlying the algorithm are valid, so the result is not surprising. In addition, there is uncertainty in the surface measurement that has not been included in these retrievals.

How close does MODIS come to meeting the performance goal of the mission — retrieving $[\rho_w]_N$ over oligotrophic waters in the blue region of the spectrum with an uncertainty of 5% or less? An overall mean absolute error in $[\rho_w]_N$ at 443 nm of 0.0029 is a relative error of around 10% or a little more. If we take the uncertainty in the surface measurements to be about 5%, then in an RMS sense the error in $[\rho_w]_N$ is about 8.6%. So in a sense, MODIS did not *meet* the performance goal, it only *approached* it; however, Figure 13.8 is for *all* of the validation data, i.e., the statistics represent all waters investigated, not just those for oligotrophic waters. With this caveat in mind, it is likely that MODIS came very close to meeting the goal.

Similar vicarious calibration and validation efforts have been carried out using above water measurements of $[\rho_w]_N$. These are not discussed here, but are referenced in the Bibliographic Notes.

Figure 13.9 compares the retrieved concentration of Chlorophyll *a* with contemporaneous surface measurements. The bio-optical algorithm relating remote sensing reflectance and C used in the retrieval of C from MODIS data is called “OC3M” and is given by¹⁸

$$\log_{10} C = a_0 + \sum_{n=1}^4 a_n \left(\frac{R_{rs}(\lambda)}{R_{rs}(551)} \right)^n \quad (13.10)$$

with

$$a_0 = 0.2424 \quad a_1 = -2.7423 \quad a_2 = 1.8017 \quad a_4 = 0.0015 \quad a_5 = -1.2280$$

and λ is the larger of $R_{rs}(443)$ and $R_{rs}(488)$ observed in the retrieved $R_{rs}(\lambda)$. Figure 13.9 provides the comparison of the retrieved C with the surface measurements. In the figure, the line is the one-to-one line. Statistics associated with the comparison are the mean absolute error (MAE) and the bias ($BIAS$) of the log-transformed concentrations. These are defined by

$$\log_{10}(MAE) = N^{-1} \sum_{i=1}^N \left| \log_{10} C_i^{\text{Sat.}} - \log_{10} C_i^{\text{Surf.}} \right|$$

¹⁸Unlike the algorithms described in Chapter 6, that used two separate bands, modern algorithms use three or four bands. These were developed when, after surface data for a wide range in C became available, it was realized that for the SeaWiFS band set, the ratio $R_{rs}(443)/R_{rs}(555)$ worked best for oligotrophic waters, $R_{rs}(490)/R_{rs}(555)$ for mesotrophic waters, and $R_{rs}(510)/R_{rs}(555)$ for eutrophic waters. Thus, an algorithm called OC4 that incorporated all of these ratios, was developed. In the case of MODIS, there is no band at 510 nm, so it uses two ratios (three bands); hence the designation “OC3M.” The “M” stands for MODIS.

and

$$\log_{10}(BIAS) = N^{-1} \sum_{i=1}^N \left(\log_{10} C_i^{\text{Sat.}} - \log_{10} C_i^{\text{Surf.}} \right).$$

The two metrics can be interpreted through a simple example as follows. Assume that the individual satellite retrievals are all a factor of x too high. Then $C_i^{\text{Sat.}} = x C_i^{\text{Surf.}}$ and $\log_{10}(MAE) = \log_{10}(x)$ or $MAE = x$, so MAE is a measure of the *multiplicative* error in $C_i^{\text{Sat.}}$. In this case the $BIAS$ would also be x : however, if half the retrievals were a factor of x too high and the other half a factor of x too low, i.e., for half $C_i^{\text{Sat.}} = x C_i^{\text{Surf.}}$ and for the other half $C_i^{\text{Sat.}} = C_i^{\text{Surf.}}/x$, then the MAE would be the same but the $BIAS$ would be unity. For the data in Figure 13.9, $MAE = 1.68$ and $BIAS = 1.18$, so the retrieved differ on the average from the measured by 68% and the bias is about 18%. In log transformed space, $\log_{10}(MAE) = 0.225$ (or about a quarter of a decade on the vertical axis of Figure 13.9) and $\log_{10}(BIAS) = 0.072$. Some of the error in the retrievals is due to error in the surface measurements, but a larger component is error in the OC3M algorithm itself. The algorithm developers did not provide MAE for the algorithm fit of C to R_{rs} ratios; however they did provide the $RMSE$ defined as

$$RSME = \sqrt{\left[\sum_{i=1}^N \frac{\left(\log_{10} C_i^{\text{Fit}} - \log_{10} C_i^{\text{Meas.}} \right)^2}{N} \right]},$$

where C_i^{Fit} is that retrieved by the algorithm, i.e., Eq. (13.10) applied to the reflectance data that was actually used in developing the algorithm, and $C_i^{\text{Meas.}}$ is the measured value. This is a measure of the inherent error in the algorithm itself, and the value obtained was about 0.25.¹⁹ The same quantity evaluated for the data in Figure 13.9 is 0.314. Thus, one could argue that a significant component of the error in the retrievals from satellite measurements is due to the inherent uncertainty in the bio-optical algorithm itself: about half in an RMS sense. Considering that the surface measurements cover more than three orders of magnitude in C , it is clear that the combination of MODIS (and SeaWiFS) along with the algorithms — atmospheric correction and bio-optical — can provide excellent global-scale estimations of C ; however, it is important to understand that the accuracy possible on regional scales with more regional-specific algorithms should be significantly better.

¹⁹Most of the various algorithms developed using the NOMAD data base also have $RSME \approx 0.25$, resulting from natural variability in the data themselves.

13.3.3 Effect and Robustness of Vicarious Calibration

We have spent a considerable amount of time discussing vicarious calibration, both theory and practice, but how essential is it? What if we ignore it altogether? What if we only perform part of it? What if there were no facility dedicated to calibration such a MOBY? Now we can address these questions using results from the SeaWiFS experience. Tables 13.2 and 13.3 provide performance results from the various stages of vicarious calibration. After

Table 13.2: Demonstration of the effect of vicarious calibration using MOBY data on the relative accuracy of SeaWiFS retrievals of $[\rho_w(\lambda)]_N$ and Chlorophyll a (C) compared to surface measurements. The numbers correspond to the *median absolute percent difference*.

λ	MOBY	Full VC	No VC	NIR VC
412	1.22	11.8	80.8	40.6
443	1.18	15.5	55.4	23.5
490	1.19	12.2	25.7	11.3
510	1.03	10.6	24.7	10.7
550	2.27	14.8	—	15.0
C	25.64	26.1	—	26.1

MOBY: Fit of retrieval of MOBY data after vicarious calibration, e.g, fit of the MOBY data itself.

Full VC: Relative error in the retrievals for the “deep water” validation data set after full (visible and NIR) vicarious calibration.

No VC: Relative error in the retrievals for the “deep water” validation data in the absence of vicarious calibration.

NIR VC: Relative error in the retrievals for the “deep water” validation data set if *only* the NIR is vicariously calibrated, i.e., the prelaunch calibration, corrected for long-term sensitivity changes, is used for the visible bands.

the full vicarious calibration was carried out, the $[\rho_w(\lambda)]_N$ retrieval algorithm was operated, using the “correct” $\rho_t^{\text{VC}}(\lambda)$ in place of the original $\rho_t^{\text{OC}}(\lambda)$, with the MOBY data used in the vicarious calibration, and $[\rho_w(\lambda)]_N$ recalculated for each of the vicarious calibration data points and compared to the original measured values. The median % difference between the

Table 13.3: Same as Table 13.2, except that the quantity given is the *median* ratio of satellite retrieval to surface measurement of $[\rho_w(\lambda)]_N$.

λ	MOBY	Full VC	No VC	NIR VC
412	1.002	1.002	0.245	0.595
443	1.003	0.950	0.447	0.779
490	1.001	0.942	0.760	1.002
510	1.003	0.957	0.753	0.964
550	0.998	0.968	—	0.965
C	—	0.994	—	0.984

originally measured and the recalculated values of $[\rho_w(\lambda)]_N$ is provided in the first column of Table 13.2 under the label “MOBY.” (Table 13.3 provides the *ratio* of the retrieved to the measured values of $[\rho_w]_N$.) With the exception of the band at 555 nm, the medians are all roughly 1% (2% at 555 nm). So, when the vicariously-calibrated SeaWiFS is applied to the MOBY data itself, Tables 13.2 and 13.3 show that almost perfect retrievals are obtained. Next, when the correct $\rho_t^{\text{VC}}(\lambda)$ are applied to an independent data set (from waters with bottom depth greater than 1000 m — a “deep water” data set that is a stand-in for Case 1 waters) the result in the second column under “Full VC” are obtained. Now the errors are mostly a little over 10% and the ratios mostly around 0.95. This is considerably poorer than that for the “MOBY” set, but the surface data are likely to be of poorer quality and have a much wider range of aerosol types. Next, we examine the case where there is no vicarious calibration at all. These results are in the third column of Tables 2 and 3. Note the large error at 412 nm ($\sim 80\%$) with a steadily decreasing error in the blue and blue-green. These results would be totally unacceptable, so omitting vicarious calibration is not an option. Finally, the NIR vicarious calibration did not require a dedicated calibration facility such as MOBY. What if the NIR vicarious calibration were the only one effected, with the pre-launch calibration used for the other bands? This question is answered in the forth column of Tables 2 and 3 labeled “NIR VC.” There, we see that excellent results are obtained for all of the bands except the two in the blue. These two bands clearly require MOBY. It seems that vicarious calibration is not needed in the blue-green and green bands; however, this is misleading. It results from the fact that the vicarious gains in these bands was very close to unity. (The gain $g(\lambda)$ values were 1.0377, 1.014, 0.9927, 0.9993, 1.000, 0.9738, and 0.9720 for 412, 443, 490, 510, 555, 670, and 765 nm, respectively.) The vicarious calibration of the 490, 510, and 555 nm bands simply showed that the original calibration was very close to being correct.

In carrying out the vicarious calibration it was taken as given that the band at λ_l (865 nm on SeaWiFS) was perfectly calibrated before launch. In addition, an aerosol model that was thought to be “appropriate” for the calibration site was used for estimating $\varepsilon(\lambda_s, \lambda_l)$. What if the pre-launch calibration of the λ_l band was not actually perfect, or an

incorrect aerosol model was chosen? Tables 13.4 and 13.5 examine these questions. The two columns labeled “ $\lambda_l : +4\%$ ” and “ $\lambda_l : -4\%$ ” provide the results of the *full* vicarious calibration carried out with the sensor-measured reflectances $\rho_t^{\text{OC}}(\lambda_l)$ at λ_l increased and decreased by 4%, respectively (i.e., the pre-launch calibrated value $\rho_t^{\text{OV}}(\lambda_l)$ was multiplied by 1.04 or 0.96). The results show only minor changes in the quality of the vicarious calibration. The last two columns in the tables labeled “M50” and “O99” provide the results obtained when the band at λ_l is assumed to be perfectly calibrated, but the aerosol model used for

Table 13.4: Demonstration of the effect of the scheme used in the NIR portion of the vicarious calibration using MOBY data on the relative accuracy of SeaWiFS retrievals of $[\rho_w(\lambda)]_N$ and Chlorophyll *a* (*C*) compared to surface measurements. The numbers correspond to the *median absolute* percent difference.

λ	Original	$\lambda_l : +4\%$	$\lambda_l : -4\%$	M50	O99
412	11.8	11.8	11.0	10.8	12.0
443	15.5	14.8	15.2	16.0	14.2
490	12.2	12.5	12.4	13.4	12.2
510	10.6	10.8	11.2	10.6	10.6
550	14.8	15.3	14.0	14.3	14.9
<i>C</i>	26.1	26.2	24.7	26.0	25.5

Original: Relative error in the retrievals for the “deep water” validation data set after the original full (visible and NIR) vicarious calibration.

$\lambda_l + 4\%$: Relative error in the retrievals for the “deep water” validation data set if the calibration of the band at $\lambda_l = 865$ nm is *increased* by 4%.

$\lambda_l - 4\%$: Relative error in the retrievals for the “deep water” validation data set if the calibration of the band at $\lambda_l = 865$ nm is *decreased* by 4%.

M50: Relative error in the retrievals for the “deep water” validation data set if the M50 aerosol model is used in the NIR vicarious calibration.

O99: Relative error in the retrievals for the “deep water” validation data set if the O99 aerosol model is used in the NIR vicarious calibration.

Table 13.5: Same as Table 13.4, except that the quantity given is the *median* ratio of satellite retrieval to surface measurement of $[\rho_w(\lambda)]_N$.

λ	Original	$\lambda_l : +4\%$	$\lambda_l : -4\%$	M50	O99
412	1.002	0.998	1.005	0.995	0.999
443	0.950	0.953	0.954	0.955	0.957
490	0.942	0.944	0.940	0.934	0.947
510	0.957	0.965	0.953	0.953	0.968
550	0.968	0.977	0.958	0.961	0.967
C	0.984	0.980	0.987	1.007	0.943

the NIR vicarious calibration was the Maritime model at 50% RH (M50) or the Oceanic model²⁰ at 99% RH, respectively from the pre-launch calibration. These are likely end-members of aerosol models applicable to the open ocean. Again, we see that the particular model used is unimportant for vicarious calibration.²¹ The statistics for the retrieved C are also provided in Tables 13.2-13.5, and show a similar lack of sensitivity to the assumptions for the NIR. The results recounted above for SeaWiFS data acquired over deep water (\sim Case 1) show conclusively that (1) vicarious calibration is absolutely necessary, (2) a calibration facility like MOBY is required for the blue bands and generally the green bands as well, (3) vicarious calibration is quite insensitive to the assumption of the absence of error at λ_l , and (4) the NIR vicarious calibration is insensitive to the aerosol model used for the estimation of $\varepsilon(\lambda_s, \lambda_l)$. These suggest that vicarious calibration is a very robust method for adjustment of sensor calibration post-launch.

It should be noted that in the case of SeaWiFS, the vicarious calibration was carried out on data that had already been corrected for long-term sensitivity changes by using the on-board solar diffuser and Lunar views. It is reasonable to wonder then whether the performance could be maintained even if these facilities for long-term sensitivity monitoring were not available. That is, could the retrieved ρ_w quality be maintained on the basis of vicarious calibration alone? In the case of SeaWiFS, the sensitivity of the band at 865 nm decreased about 20% over approximately ten years. Let's assume that excellent vicarious calibrations were performed over that period of time. In particular, consider an ideal situation where at the vicarious calibration site, the O99 aerosol was extant, and the aerosol optical depth was 0.05 (very clear). Further, assume that a very close aerosol model, M99, was used in the vicarious calibration. Simulations have been carried out (including multiple scattering) similar to those described in Chapter 12 (Section 12.3.2) to effect the

²⁰The Oceanic model is the large-particle component of the Maritime model (no small-particle component is present) and should be extant far from land, where the principal aerosol is that generated by breaking waves (see Chapter 4).

²¹This may seem confusing, but at the vicarious calibration sites chosen for the NIR, the aerosol optical depth is low, which minimized the influence of the aerosol on the top-of-atmosphere reflectance.

vicarious calibration as a function of the calibration error at 865 nm, $\alpha(865)$. After the resulting calibration adjustment, the atmospheric correction algorithm was invoked and the error in the retrieved value of $\rho_w(443)$ derived. The simulated aerosol was assumed to be M80, with $\tau_a(865) = 0.05, 0.1$, and 0.2 . The results are presented in Figure 13.10, showing that the error in the retrieved $\rho_w(443)$ is within the desired ± 0.002 as long as $\tau_a(865)$ is not too large. For $\tau_a(865) \leq 0.1$ the retrievals could be made with excellent accuracy for $|\alpha(865)| \leq 15\%$. However, the error grows with increasing $|\alpha(865)|$ and becomes excessive when $|\alpha(865)| \approx 15\%$ for $\tau_a(865) = 0.20$. We can see what this means in terms of ocean coverage by examining long-term averages of aerosol optical depths as shown in Figure 13.11. Clearly, over much of the open oceans, $\tau_a(550)$ is 0.222 or less; with large areas 0.111 or less.²² However, there are large areas, in particular in the vicinity of coasts, where $0.222 \leq \tau_a(550) \leq 0.333$, and for these, $|\alpha(865)| \gtrsim 15\%$ would yield a very poor atmospheric correction. So, it seems that for any sensor which displays SeaWiFS-like sensitivity decay, vicarious calibration without sensitivity monitoring, would be successful only for a limited time after launch.²³ The prudent approach is to provide such stability monitoring.

13.4 Concluding Remarks

The experience with the CZCS showed that a very rudimentary ocean color sensor could provide important data concerning biological processes in the ocean. Indeed, one need only examine Figure 13.5 see the close coupling between physical and biological processes, i.e., the motion of the water (physics) is a principal driver of biology through the resupplying of nutrients. In the early 1990s a well-known-marine biologist, presenting a plenary paper at an international meeting on the most important developments in marine biology in the past 50 years, placed the CZCS and ocean color among the top six. This was his assessment even though the uncertainty in the derived biological produce C_P was about a factor of two at that time. If the uncertainty could be significantly reduced, the potential of ocean color sensors seemed unlimited because of its global nature and the perspective offered by the view from space. Of particular interest was the role of phytoplankton in global carbon cycle.

Now it is clear from the experiences with SeaWiFS and MODIS that ocean color sensors that are well characterized and vicariously calibrated can provide water-leaving reflectances and Chlorophyll concentrations with accuracies closely approaching those made with surface measurements. In the case of Chlorophyll *a*, it is clear that much of the dispersion

²²The relationship between $\tau_a(550)$ and $\tau_a(865)$ depends on the aerosol type. For M80, $\tau_a(550)$ is about 10% larger than $\tau_a(865)$ (see Chapter 4, Figure 4.9).

²³The VIIRS instrument shows a larger sensitivity decline with time than SeaWiFS.

between the remotely-retrieved and the surface-measured concentration is due to similar dispersion between surface-measured radiances and concentration, i.e., uncertainty within the bio-optical method itself. Much of the latter uncertainty is the result of natural variations, presumably due to variations of IOPs within and among species for a given Chlorophyll value (Chapter 5). This simply points to the fact that the Chlorophyll concentration is not an adequate characterization of the retinue of plankton present in Case 1 waters. It is indeed remarkable that its characterization does as well as it does.

The vicarious calibration and validation are ongoing processes that are basically unending. Any improvement (or change) in the algorithms, in the understanding of the operation of the sensor, or in the processing of the MOBY vicarious calibration data, not only requires a reanalysis of retrievals at the validation targets, but prior to that, a reanalysis of the vicarious calibration data to determine any effect the change may have on the calibration adjustments. In this sense, while ocean color data may be of some value in real time, e.g., to locate ideal positions for fishing, its greater value must of necessity be retrospective in nature.

13.5 Bibliographic Notes

13.2.1 Surface Measurement of $[\rho_w]_N$

A set of protocols have been developed for surface measurements in support of validation, e.g., measurement of $[\rho_w]_N$, IOPs, etc. They were developed as part of the SIMBIOS project and are reported in the Bibliography under *Mueller, Fargion, and McClain, Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Volumes I–VI* [Mueller et al., 2003a,b,c,d,e,f].

13.3.1 Verification of CZCS Concept

The first demonstration of the capability of CZCS for deriving C_P was described in Gordon et al. [1980]. Later, with expanded computational resources and a better understanding of the sensor, a more complete verification was presented in Gordon et al. [1983]. Study of the stability of the CZCS in orbit, principally variations in its radiometric sensitivity, were first described in Gordon et al. [1983] and then, over the course of the entire mission, by Evans and Gordon [1994]. The “clear water radiance” concept used in the verification of CZCS was initiated in Gordon and Clark [1981].

13.3.2 Validation of MODIS Performance

Discussion of the post-CZCS needs and requirements for the validation of atmospheric correction are given in [Clark et al. \[1997\]](#). Various components of NASA's SeaWiFS validation program and initial results are outlined in [Hooker and McClain \[2000\]](#). The MOBY system is described in [Clark et al. \[2003\]](#), and a discussion of propagating the MOBY measurements to the surface is found in [Voss et al. \[2017\]](#). An initial vicarious calibration was described by [Eplee Jr. et al. \[2001\]](#), the OPBG's definitive calibration reported by [Franz et al. \[2007\]](#), and a report comparing the calibration of several sensors is [Eplee Jr. and Bailey \[2014\]](#). Details for insitu measurements of the relevant vicarious calibration and validation quantities are found in [Zibordi and Voss \[2014\]](#). The bio-optical algorithm for MODIS (OC3M) was originally developed by [O'Reilly et al. \[2000\]](#).

Validation using above surface measurements are mostly based on a modified AERONET sensor [[Zibordi et al., 2009](#)]. Match ups with above-surface measurements and several space-borne sensors are presented in [Zibordi et al. \[2006\]](#), with some additional comparisons in [Zibordi and Voss \[2014\]](#).

13.3.3 Effect and Robustness of Vicarious Calibration

Tables [13.2-13.5](#) were compiled from similar tables in [Franz et al. \[2007\]](#). [Wang and Gordon \[2002\]](#) suggested that the pre-launch calibration at 865 was really not important, and their calculations were used in the preparation of Figure [13.10](#).

13.6 Figures

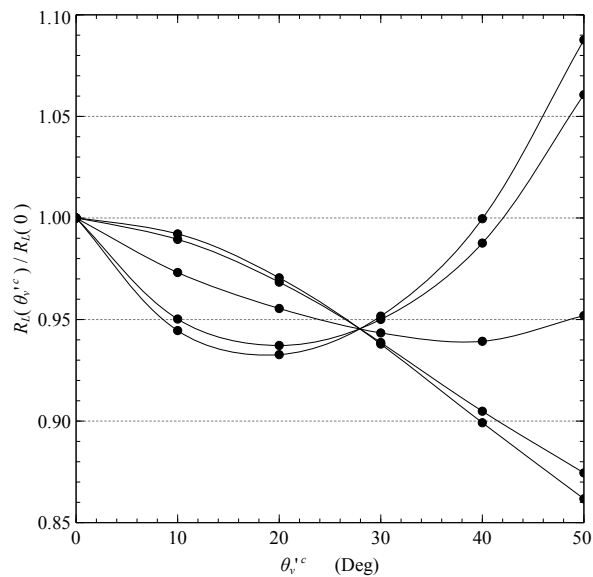


Figure 13.1: QSSA estimate of the ratio of $R_L(\theta_v'^c)$ to R_L at nadir, where $\theta_v'^c (= \pi - \theta_v')$ is the angle that the viewing direction makes with the *upward normal*, i.e., \hat{n}_S , just beneath the surface. The curves correspond to computations using a combination of water and particles, which are characterized by the Petzmas phase function. The relative particle contribution is determined by $x = (b_b)_p / (b_b)_w$. The curves are for $x = 0.01, 0.1, 1.0, 10.0$ and 100.0 (from top to bottom near $\theta_v'^c = 10^\circ$).

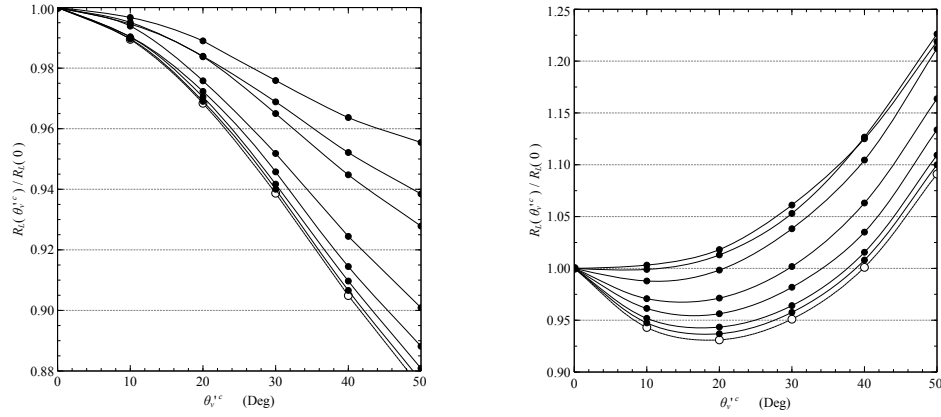


Figure 13.2: Exact computations of the ratio of $R_L(\theta_v^c)$ to R_L at nadir as in Figure 13.1 compared with the QSSA estimate. The solid symbols refer to $\omega_0 = 0.1, 0.2, 0.4, 0.6, 0.8, 0.85$, and 0.9 from the bottom to the top. The open symbols at the very bottom are the result of the QSSA approximation. Left panel: $(b_b)_p/(b_b)_w = 0.1$. Right panel: $(b_b)_p/(b_b)_w = 100$.

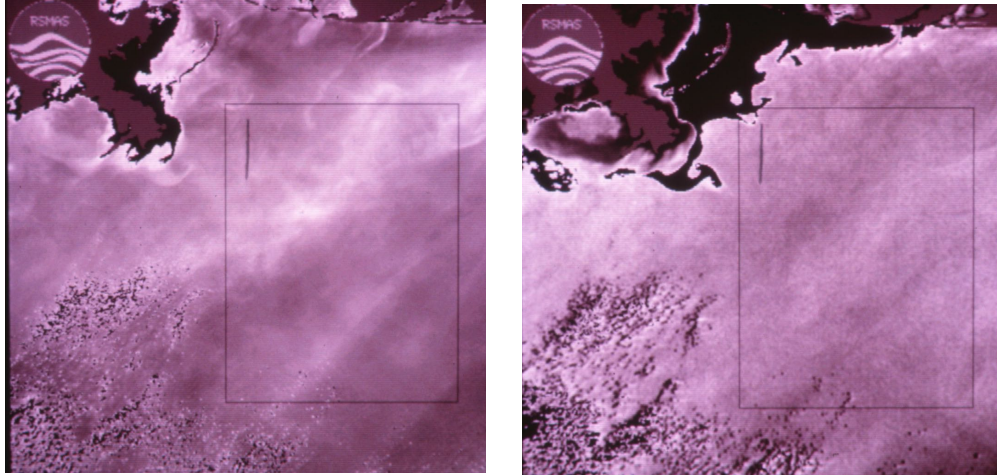


Figure 13.3: Left Panel: Image of $L_a(670) = L_t(670) - L_r(670)$ from CZCS Orbit 130 over the Mississippi Delta region. Right Panel: Image of $\varepsilon(520, 670)$ from CZCS Orbit 130 over the Mississippi Delta region.

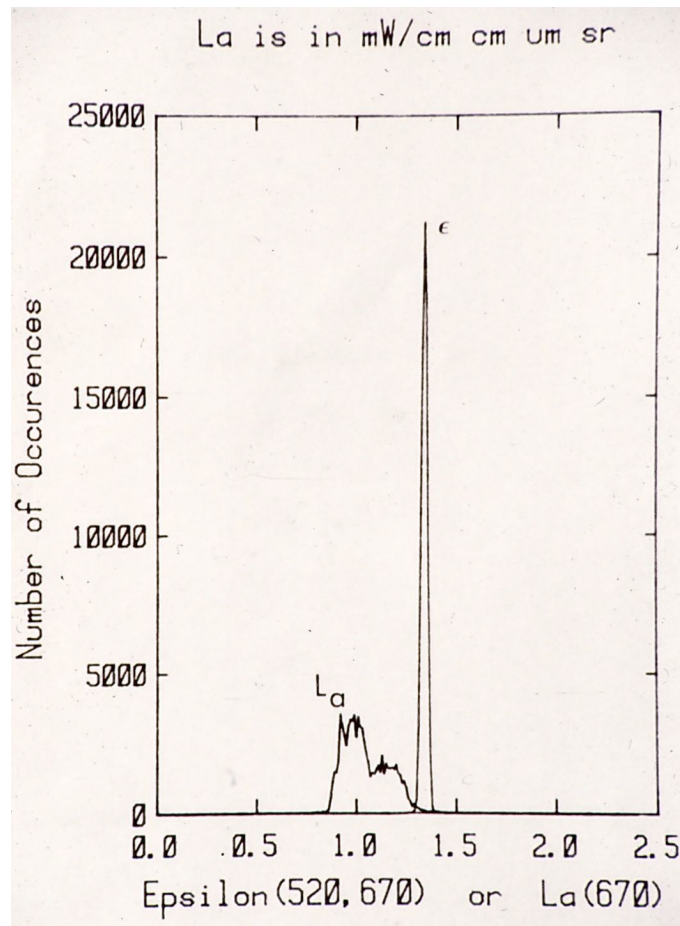


Figure 13.4: Histogram of the variations $L_a(670)$ and $\epsilon(520, 670)$ from pixels in the box drawn on Figure 13.3.

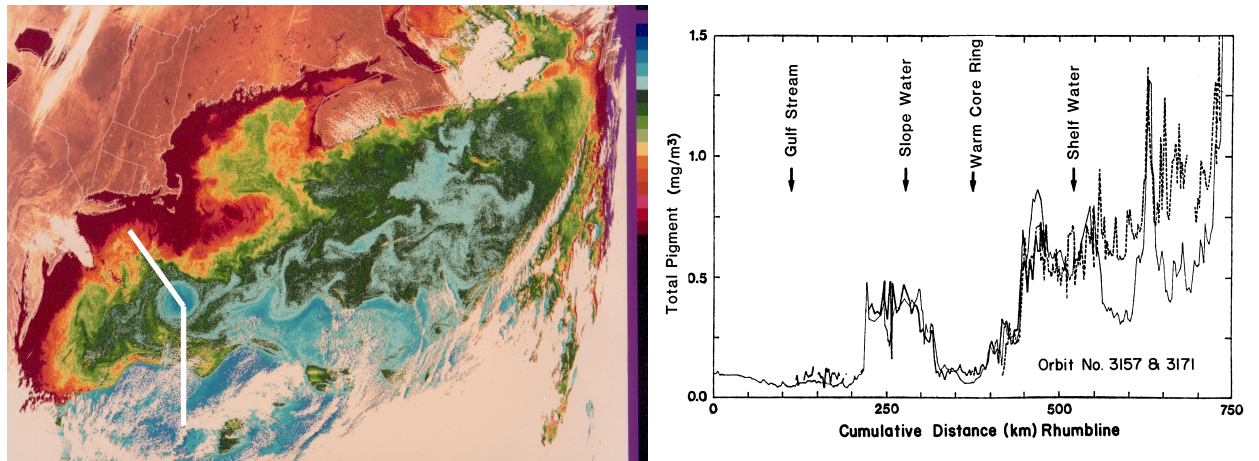


Figure 13.5: Left Panel: A four-minute CZCS scene from Orbit 3226 over the Northern Sargasso Sea and the Gulf Stream. Long Island can be seen at the extension of the top of the white line toward the northwest. The image displays the retrieved C_P in false color (blue \rightarrow green \rightarrow red in the direction of increasing C_P). Right Panel: The retrieved values of C_P (thick line) along the ship track (white line shown in the right panel), compared to ship measurements of C_P (thin line) along the same track. The ship and CZCS data were obtained contemporaneously with Orbits 3151 and 3171 several days before this image. Thus, the ship track spans two days. (Thick solid line is from Orbit 3157 and thick dashed line is from Orbit 3171.)

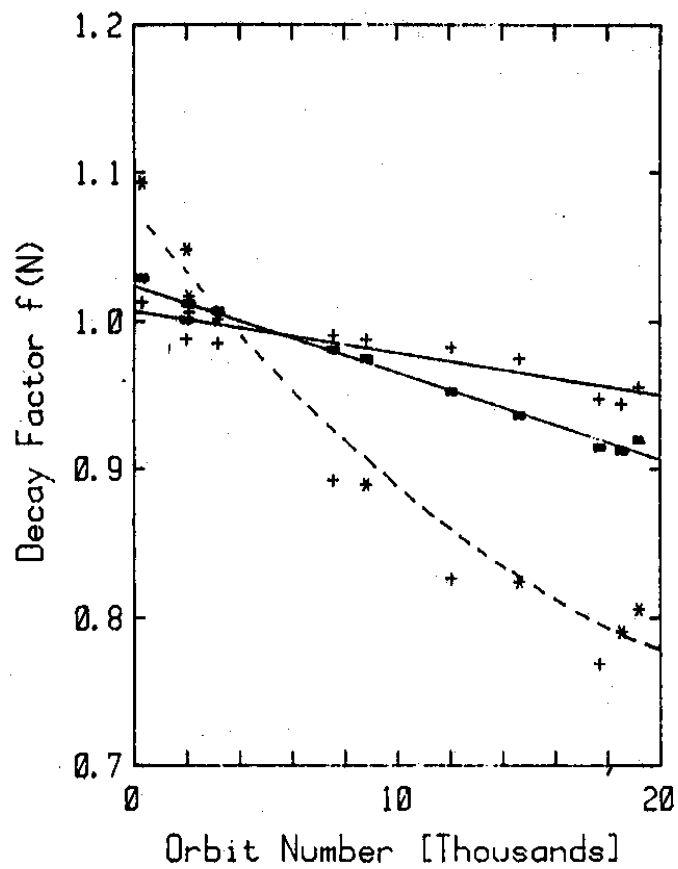


Figure 13.6: Decay factor $f(N)$ as a function of CZCS Orbit Number N . The order of the spectral bands from bottom to top near Orbit Number 20,000 is 443, 520, and 550 nm. For 443 nm the (+) symbols refer to early spring, March and April, while the (*) symbols are for June through September. From [Gordon et al. \[1983\]](#).

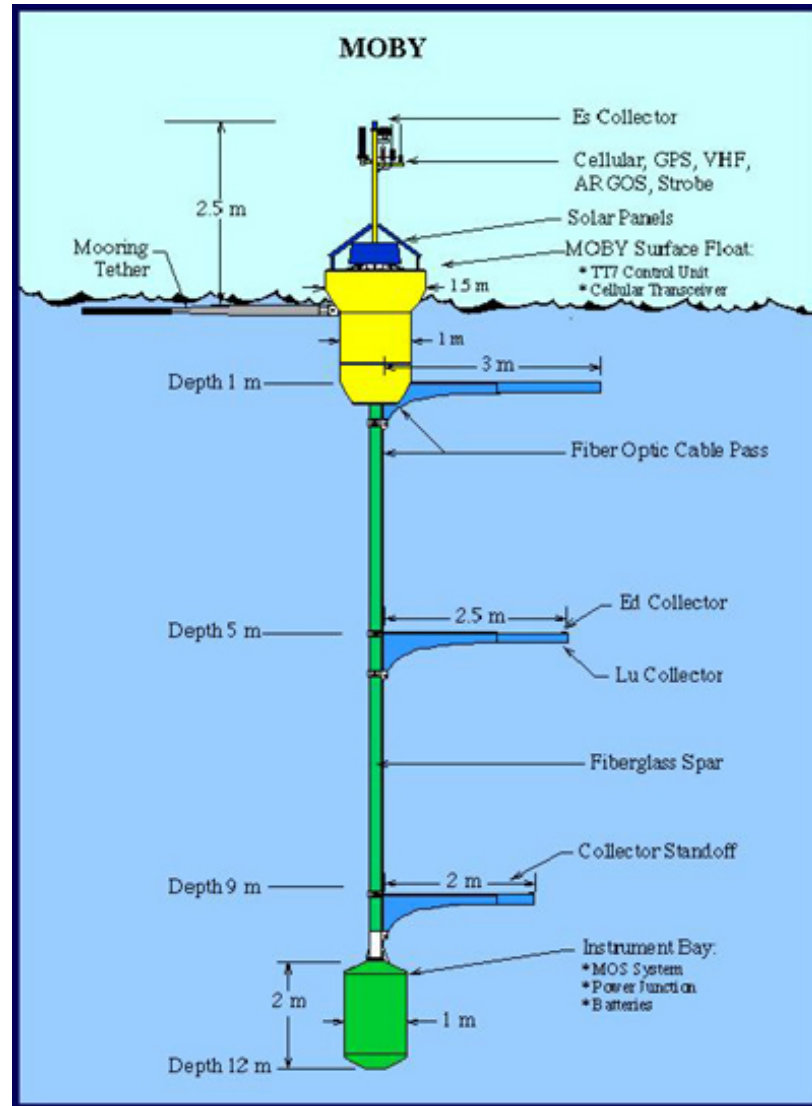


Figure 13.7: Schematic of the Marine Optical Buoy (MOBY). The key parts for vicarious calibration of MODIS are the three arms on the right, which measure $L_u(\hat{n}_S, \hat{\xi}_0', \lambda)$ and $E_d(\lambda, \hat{\xi}_0')$ at three different depths — 1, 5, and 9 m. Figure downloaded from <https://www.mlml.calstate.edu/moby/moby-photo-gallery/>.

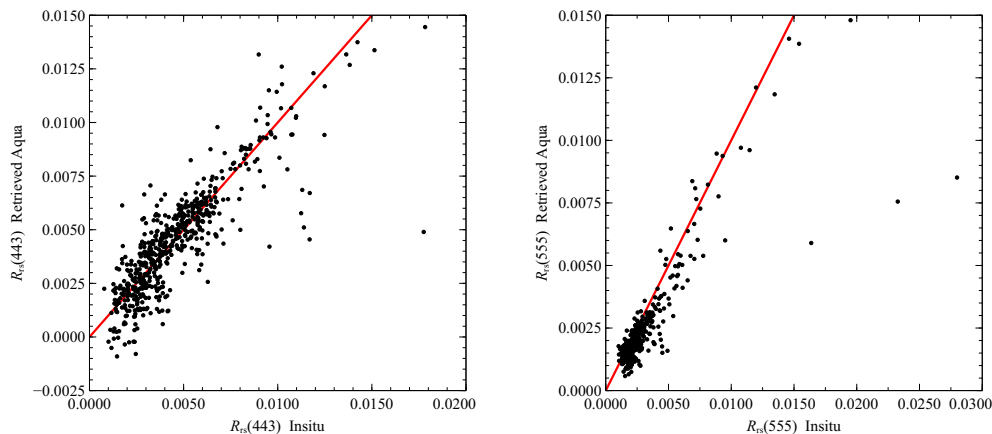


Figure 13.8: Comparison of $R_{rs} = [\rho_w]_N/\pi$ retrieved from MODIS Aqua with in-situ surface measurements for 443 nm (left) and 555 nm (right). Solid line is the one:one line. Data are from the 2018 reprocessing. Data citation: NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Ocean Color Data; 2018 Reprocessing. NASA OB.DAAC, Greenbelt, MD, USA. doi: 10.5067/AQUA/MODIS/L2/OC/2018. Accessed on 04/27/2018.

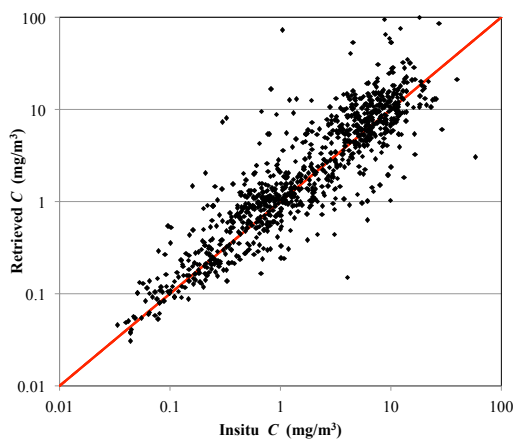


Figure 13.9: Comparison of Chlorophyll *a* concentration retrieved from MODIS Aqua with in-situ surface measurements. Solid line is the one:one line. Data are from the 2018 reprocessing. Data citation: NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Ocean Color Data; 2018 Reprocessing. NASA OB.DAAC, Greenbelt, MD, USA. doi: 10.5067/AQUA/MODIS/L2/OC/2018. Accessed on 04/27/2018.

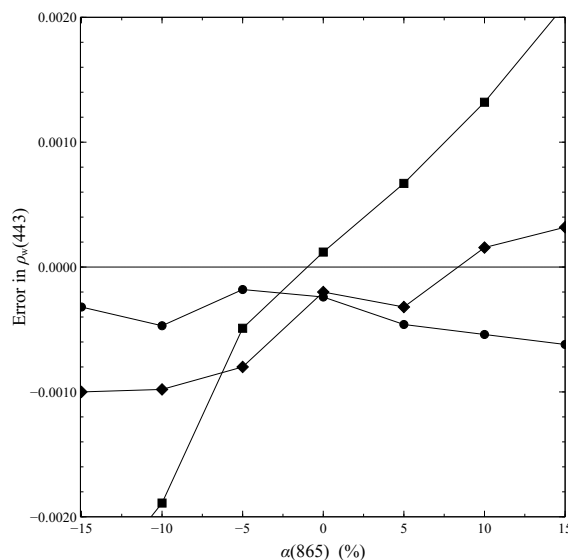


Figure 13.10: Simulated error in the retrieved value of $\rho_w(443)$, after vicarious calibration, as a function of the prelaunch calibration error in $\rho_t(865)$. The different sets correspond to different values of $\tau_a(865)$ at the target site: Circles, $\tau_a(865) = 0.05$; Diamonds, $\tau_a(865) = 0.10$; Squares, $\tau_a(865) = 0.20$.

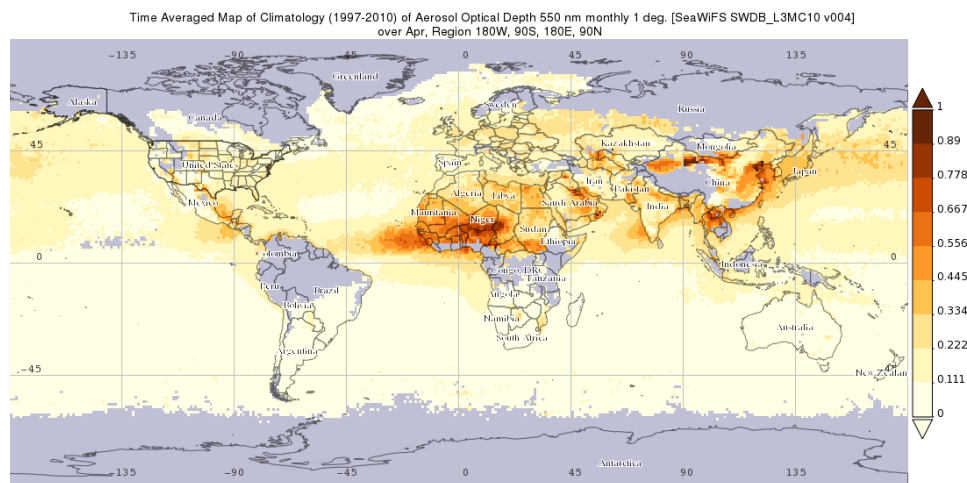


Figure 13.11: Image of $\tau_a(550)$ for the month of April averaged for 1997-2010 from SeaWiFS retrievals. Citation: N. Christina Hsu, Andrew M. Sayer, M.-J. Jeong, and Corey Bettenhausen (2013), SeaWiFS Deep Blue Aerosol Optical Depth and Angstrom Exponent Monthly Level 3 Data Gridded at 0.5 Degrees V004, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed [May 6, 2018] 10.5067/MEASURES/SWDB/DATA303.

Chapter 14

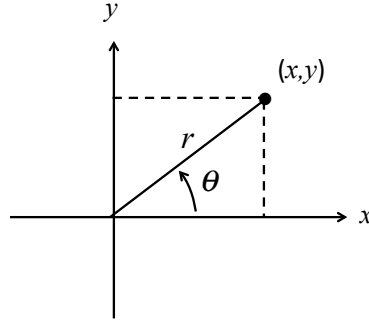
Mathematical Appendix

In this appendix we provide most of the mathematics (other than basic algebra) that is required to understand the mathematical operations used in the text. We start with the algebra of complex numbers. Next, we describe the operations of single variable calculus followed by multivariable calculus and vectors, which are best treated together. This is followed by a description of matrix algebra and coordinate transformations, which is useful for understanding and manipulating polarized light. We end the appendix with some selected topics that it is hoped will enhance understanding of some of the derivations and concepts in the text.

14.1 Complex Numbers and Their Manipulation

The quadratic equation $x^2 + 1 = 0$ has no real number x as a solution; however, if the number system is expanded to include *complex* numbers, there is a solution. Complex numbers are ordered pairs of real numbers (a, b) usually written in the form $z = a + ib$, with $i \triangleq \sqrt{-1}$. In this expanded system of numbers, the solutions to $z^2 + 1 = 0$ are the complex numbers $z = \pm i$. If $z = a + ib$, a is called the *real part*, often written $\Re(z)$, and b is called the *imaginary part*, written $\Im(z)$. The manipulation of complex numbers is easy when written in the form $z = a + ib$ because their manipulation follows all of the usual rules of algebra, e.g., the sum of two complex numbers is

$$\begin{aligned} z_1 + z_2 &= (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2) = z_3 \\ &\implies a_3 = a_1 + a_2 \text{ and } b_3 = b_1 + b_2, \end{aligned}$$



and the product of two complex numbers is

$$\begin{aligned} z_1 z_2 &= (a_1 + ib_1)(a_2 + ib_2) = a_1 a_2 + i(a_1 b_2 + a_2 b_1) + i^2(b_1 b_2) \\ &= a_1 a_2 - b_1 b_2 + i(a_1 b_2 + a_2 b_1) = z_3, \end{aligned}$$

where $a_3 = a_1 a_2 - b_1 b_2$ and $b_3 = a_1 b_2 + a_2 b_1$.

A useful way to think about complex numbers is that they are a point (a, b) in the *complex plane*. The complex plane is prescribed as follows. Consider a plane where each point has cartesian coordinates (x, y) . Then the complex number $z = x + iy$ is just the point (x, y) in that plane, as shown in the figure above, in which the complex number z is represented by the large dot.¹ One can also use polar coordinates in the plane to represent z through

$$x = r \cos \theta, \quad y = r \sin \theta, \quad \text{where} \quad r = \sqrt{x^2 + y^2} \quad \text{and} \quad \theta = \arctan(y/x).$$

Thus, $z = x + iy = r(\cos \theta + i \sin \theta)$. The beauty of this is that, as we show later using power series, $(\cos \theta + i \sin \theta) = \exp(i\theta)$, so

$$z = x + iy = r \exp(i\theta),$$

which greatly facilitates the manipulation of complex numbers. For example, if we are given the complex number $1/(a+ib)$, how do we write it in the form $x+iy$? Let $a+ib = r \exp(i\theta)$, then $1/(a+ib) = r^{-1} \exp(-i\theta) = r^{-1}(\cos \theta - i \sin \theta)$, so

$$x = \frac{1}{\sqrt{a^2 + b^2}} \cos \theta, \quad y = -\frac{1}{\sqrt{a^2 + b^2}} \sin \theta \quad \text{and} \quad \theta = \arctan(b/a).$$

Before continuing we need to define some additional quantities. The *magnitude* of a complex number $z = x + iy = r \exp(i\theta)$, written $|z|$, is $\sqrt{x^2 + y^2} = r$. The *complex*

¹If the line from the origin to the dot on the figure is thought of as a vector in this plane (pointing toward the dot), the sum of two complex numbers $z_1 + z_2$ is determined in a manner identical to the sum of the two vectors.

conjugate of $z = x + iy = r \exp(i\theta)$, written z^* , is $x - iy = r \exp(-i\theta)$. Notice that the product $z^*z = r^2 = |z|^2$. The notion of the complex conjugate provides another way to write $z = 1/(a + ib)$ in a simpler form: multiply z by the complex conjugate of $a + ib$ divided by itself, i.e.,

$$z = z \frac{z^*}{z^*} = \frac{1}{a + ib} \left(\frac{a - ib}{a - ib} \right) = \frac{a - ib}{a^2 + b^2}, \quad \text{and so} \quad x = \frac{a}{a^2 + b^2}, \quad y = \frac{-b}{a^2 + b^2}.$$

One should note that the meaning of the statement $z_1 = z_2$ is just $a_1 = a_2$ and $b_1 = b_2$, while statements like $z_1 < z_2$ are meaningless; however, the statement $|z_1| < |z_2|$ is well defined.

Many functions carry over from the real to the complex domain quite simply. For example, what is $\sin(i\vartheta)$ in terms of elementary functions? Noting that

$$\sin(\vartheta) = \frac{1}{2i} [\exp(i\vartheta) - \exp(-i\vartheta)],$$

we have

$$\begin{aligned} \sin(i\vartheta) &= \frac{1}{2i} [\exp(ii\vartheta) - \exp(-ii\vartheta)], \\ &= \frac{1}{2i} [\exp(-\vartheta) - \exp(\vartheta)], \\ &= -\frac{1}{2i} [\exp(+\vartheta) - \exp(-\vartheta)], \\ &= i \sinh(\vartheta), \end{aligned}$$

so $\sin(i\vartheta)$ is pure imaginary (no real part). Many trigonometric identities are easily derived using complex numbers. For example, consider the following

$$\begin{aligned} \exp(i\theta_1) \exp(i\theta_2) &= \exp[i(\theta_1 + \theta_2)] = \cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2), \\ &= (\cos \theta_1 + i \sin \theta_1)(\cos \theta_2 + i \sin \theta_2), \\ &= (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i(\cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2), \end{aligned}$$

which implies

$$\begin{aligned} \cos(\theta_1 + \theta_2) &= (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2), \\ \sin(\theta_1 + \theta_2) &= (\cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2). \end{aligned}$$

The study of functions of a complex variable is vast and rewarding, but we have covered all that is necessary for our purpose.

14.2 Single Variable Calculus

In this section we review single variable calculus focussing on the concepts of differentiation and integration, including a discussion of the Dirac delta function. We end the section with

a detailed description of the solution of second-order differential equations with constant coefficients.

14.2.1 Derivatives of a Function

Consider a function f of a single variable x . This is usually written $f(x)$, where the notation means that given a value of x , say x_0 , then the *value* of the function f is $f(x_0)$.² It is important to note that $f(x_0)$ is a number (which may or may not have dimensions, e.g., meters. etc.). To emphasize this, one often writes $y = f(x)$, where x and y are numbers. The subject of calculus is centered on the question; if I change x a little, e.g., by Δx , how much does f change? The change in f , written Δf is given by

$$\Delta f(x) \approx f'(x)\Delta x$$

where $f'(x)$ is called the derivative of f with respect to x and is usually written df/dx . The approximation gets better and better as Δx gets smaller and smaller. Formally we define $f'(x)$ using the above as

$$f'(x) \triangleq \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad \text{or} \quad f'(x) \triangleq \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x/2) - f(x - \Delta x/2)}{\Delta x},$$

so as Δx gets smaller and smaller, the quotients in the above formulas get closer and closer to $f'(x)$.³ Note that if we make a graph of $y = f(x)$, then $f'(x)$ is simply the *slope* of the curve of $f(x)$ at the point x . Often in the text, Δx and dx are used almost interchangeably as shortcuts, but one can easily figure out the actual meaning by the context. The fact that $f'(x)$ is the slope of $f(x)$ at x shows that values of x for which $f'(x) = 0$ are either maxima or minima of $f(x)$.⁴

The definition of the derivative can be used to find its value (but it's not always simple), e.g., let $f(x) = x^2$. Then

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{2x\Delta x + (\Delta x)^2}{\Delta x} = \lim_{\Delta x \rightarrow 0} (2x + \Delta x) = 2x.$$

²Colloquially, feed x_0 to the function f and you get its value $f(x_0)$.

³At the risk of being too formal, we define what is meant by a limit. The limit, $\lim_{x \rightarrow a} f(x) = A$, means that if we chose an $\epsilon > 0$, one can always find a $\delta > 0$, such that if $|x - a| < \delta$, then $|f(x) - A| < \epsilon$. Similarly, $\lim_{x \rightarrow \infty} f(x) = A$ means, choosing an $\epsilon > 0$, one can always find an a , such that if $x > a$, then $|f(x) - A| < \epsilon$.

⁴To determine which, i.e., maximum or minimum, one computes $f''(x)$ at the point x where $f'(x) = 0$. If $f''(x) < 0$, $f(x)$ is a maximum, if $f''(x) > 0$, $f(x)$ is a minimum, and if $f''(x) = 0$, $f(x)$ is said to have a point of inflection at x .

A sample of the derivatives of some well-known functions are:

$$\begin{aligned}\frac{dx^n}{dx} &= nx^{n-1}, \\ \frac{d \sin(x)}{dx} &= \cos(x), \\ \frac{d \cos(x)}{dx} &= -\sin(x), \\ \frac{d \exp(x)}{dx} &= \exp(x), \\ \frac{d \ln(x)}{dx} &= \frac{1}{x}.\end{aligned}$$

From these, others can be derived,⁵ e.g., given the second, we know that

$$\frac{d \sin(ax)}{d(ax)} = \cos(ax) \quad \text{so} \quad \frac{d \sin(ax)}{d(x)} = a \cos(ax).$$

Higher order derivatives are indicated by more primes and are defined in a similar manner, i.e., the second derivative of f is

$$\frac{d^2 f}{dx^2} = f''(x) \triangleq \lim_{\Delta x \rightarrow 0} \frac{f'(x + \Delta x) - f'(x)}{\Delta x}.$$

The number of primes on the function refer to the number of derivatives, or the *order* of the differentiation.

14.2.2 Taylor Series Representation of a Function

Consider the function $f(x)$. Can it be represented as an infinite series

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots = \sum_{n=1}^{\infty} a_n x^n ? \quad (14.1)$$

If it can, the a_n 's are easy to find. Let $x = 0$, then $a_0 = f(0)$. Again, if it can be represented as above, compute

$$f'(x) = a_1 + 2a_2x + 3a_3x^2 \cdots ,$$

and again letting $x = 0$ shows that $a_1 = f'(0)$. Continuing in this manner,

$$a_n = \frac{f^{(n)}(0)}{n!},$$

⁵Another example is the derivative of a product. From the definition of the derivative, it is easy to show that, given $f(x)$ and $g(x)$, then $(fg)' = f'g + fg'$.

where the superscript “ (n) ” on f indicates the n^{th} derivative of f . Thus for $f(x)$ to be represented by the series, it is necessary that it can be differentiated an infinite number of times.⁶ Some useful (and hopefully familiar) series that are easily derived are

$$\begin{aligned}\frac{1}{1-x} &= 1 + x + x^2 + x^3 \cdots, \\ \sin(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \cdots, \\ \cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \cdots, \\ \exp(x) &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} \cdots.\end{aligned}$$

If we write the last series for ix , where $i^2 = -1$, rather than just x , and compare the result with the second and third series, we find

$$\begin{aligned}\exp(ix) &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} \cdots, \\ &= \cos(x) + i\sin(x),\end{aligned}$$

an important relationship when dealing with functions involving complex numbers, as we have seen earlier. A generalization of Eq. (14.1) is

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots = \sum_{n=1}^{\infty} a_n(x - x_0)^n. \quad (14.2)$$

In this case,

$$a_n = \frac{f^{(n)}(x_0)}{n!},$$

where the notation $f^{(n)}(x_0)$ means; differentiate f a total of n times and evaluate the result at $x = x_0$. Series of the type indicated in Eq. (14.2) are called *Taylor* series, and functions for which Taylor series can be written (and converge) are called *analytic* functions.

An important application of Taylor series is in the evaluation of limits. Suppose one wants to evaluate

$$\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} \text{ but both } \lim_{x \rightarrow 0} f(x) = 0 \text{ and } \lim_{x \rightarrow 0} g(x) = 0,$$

⁶We will not discuss the important question of *convergence* of such series, i.e., for what range of values of x the sum in Eq. (14.1) actually adds up to the value of the function at the given x . All of the series discussed here do converge to the function they represent for some range of x .

yielding $0/0$ for the limit; and indeterminate form. If both $f(x)$ and $g(x)$ can be expanded in a Taylor series about $x = 0$, i.e.,

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots = \sum_{n=1}^{\infty} a_n x^n$$

$$g(x) = b_0 + b_1x + b_2x^2 + \cdots = \sum_{n=1}^{\infty} b_n x^n,$$

where in the proposed problem $a_0 = 0$ and $b_0 = 0$. Then

$$\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 0} \left(\frac{\sum_{n=1}^{\infty} a_n x^n}{\sum_{n=1}^{\infty} b_n x^n} \right) \rightarrow \frac{a_1}{b_1} = \frac{f'(0)}{g'(0)},$$

assuming that $a_1/b_1 \neq 0/0$. If $a_1/b_1 = 0/0$, one can continue this procedure n times until $a_n/b_n \neq 0/0$, in which case the required limit is $f^n(0)/g^n(0)$. This is known as *L'hospital's rule*. A very simple example of L'hospital's rule is $\lim_{x \rightarrow 0} (\sin(x)/x) = 1$.

14.2.3 Integration

An important concept in calculus is integration. We review this now. What do we mean by the *definite* integral $I = \int_a^b f(x) dx$? I is operationally defined as follows. Divide the x axis between a and b into N intervals and let $\Delta x = (b - a)/N$, then

$$I \triangleq \lim_{N \rightarrow \infty} \sum_{i=1}^N f(\xi_i) \Delta x, \quad (14.3)$$

where ξ_i is a point on the x axis *anywhere* in the i^{th} interval, i.e., $x_{i-1} \leq \xi_i \leq x_i$.⁷ If we graph the function $y = f(x)$, from the definition it should be clear the I is the area between the curve of f and the x -axis, with the area positive above the axis and negative below it.

An important theorem regarding integration can be justified by examining

$$\int_a^b f'(x) dx = \lim_{N \rightarrow \infty} \sum_{i=1}^N f'(\xi_i) \Delta x.$$

⁷This is not the most general definition of the integral. More generally, the intervals (Δx 's) do not have to all be the same size, but then in the limiting procedure one must require that the largest interval goes to zero as $N \rightarrow \infty$.

Writing $f'(x_i)\Delta x = (f(x_{i+1}) - f(x_{i-1}))/2$, and writing out the sum term by term, we discover that all of the f 's cancel except the two at the endpoints, i.e.,

$$\int_a^b f'(x) dx = f(b) - f(a). \quad (14.4)$$

This suggests that the *indefinite* integral of $f'(x)$ is

$$\int f'(x) dx = f(x) + C,$$

where C is an arbitrary constant. Then the definite integral (i.e., that with the limits) is

$$\int_a^b f'(x) dx = f(x) \Big|_a^b.$$

This establishes $f(x)$ as the *antiderivative* of $f'(x)$. Equation Eq. (14.4) is sometimes called the *Fundamental Theorem of Integral Calculus*.

14.2.4 The Dirac Delta Function

Consider the function $\delta_n(x)$, where

$$\begin{aligned} \delta_n(x) &= n, & -\frac{1}{2n} \leq x \leq +\frac{1}{2n} \\ \delta_n(x) &= 0 & \text{otherwise.} \end{aligned}$$

Then

$$A_n = \int_{-\infty}^{+\infty} \delta_n(x) dx = 1,$$

and $\lim_{n \rightarrow \infty} (A_n) = 1$. The function $\delta_n(x)$ is well defined and integrable, but δ_∞ is poorly defined, it is a “spike” of infinite height and zero thickness, and is hardly integrable as defined in Eq. (14.3).⁸ However, the sequence of integrals A_n is well defined and, in the limit, they converge to 1 (because each term in the sequence is unity). Now consider

$$B_n = \int_{-\infty}^{+\infty} \delta_n(x) f(x) dx,$$

⁸There are other representations that can be used for $\delta_n(x)$ rather than the “box” function used here. Two *continuous* examples are

$$\delta_n(x) = \frac{n}{\sqrt{\pi}} \exp(-n^2 x^2) \quad \text{and} \quad \delta_n(x) = \frac{1}{n\pi} \frac{\sin^2(nx)}{x^2},$$

which become larger and narrower near the origin as n becomes larger. Both of these, along with the box function, have the property that $\int_{-\infty}^{+\infty} \delta_n(x) dx = 1$; however, these two have continuous and finite derivatives.

where $f(x)$ is a continuous function. Clearly,

$$B_n = f(\xi_n), \quad -\frac{1}{2n} \leq \xi_n \leq +\frac{1}{2n},$$

where ξ_n is some value of x in the above range. Clearly, $\lim_{n \rightarrow \infty} (B_n) = f(0)$. We define the *Dirac delta function*, $\delta(x)$ through

$$\int_{-\infty}^{+\infty} \delta(x) f(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} \delta_n(x) f(x) dx = f(0).$$

This is easily extended to

$$\int_{-\infty}^{+\infty} \delta(x - x_0) f(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} \delta_n(x - x_0) f(x) dx = f(x_0), \quad (14.5)$$

where

$$\begin{aligned} \delta_n(x - x_0) &= n, & -\frac{1}{2n} \leq (x - x_0) \leq +\frac{1}{2n} \\ \delta_n(x - x_0) &= 0 & \text{otherwise.} \end{aligned}$$

Thus, the Dirac delta function has the property that when integrated with another function it provides the value of the other function at the point where the argument of the delta function is zero. This is sometimes called the *sifting* property. The neat thing about the delta function as defined in Eq. (14.5) is that (often) it can be treated as an ordinary function. For example, treating $\delta(x)$ as an ordinary continuous function and using integration by parts,⁹

$$f'(x_0) = \int_{-\infty}^{+\infty} \delta(x) f'(x) dx = \delta(x) f(x) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \delta'(x) f(x) dx = - \int_{-\infty}^{+\infty} \delta'(x) f(x) dx,$$

so the meaning of $\delta'(x)$ is defined through¹⁰

$$\int_{-\infty}^{+\infty} \delta'(x - x_0) f(x) dx = -f'(x_0).$$

⁹Integration by parts follows from the derivative of a product. Thus integrating

$$\int_a^b d(fg) = \int_a^b f dg + \int_a^b g df \implies \int_a^b f dg = fg \Big|_a^b - \int_a^b g df.$$

¹⁰This can be understood qualitatively by using one of the continuous representations of $\delta_n(x)$, for which the derivative results in two spikes (one positive and one negative) near the origin.

Using the sequences, or treating $\delta(x)$ as an ordinary function (it isn't), we can develop the following relationships;

$$\begin{aligned}\delta(ax) &= \frac{1}{|a|}\delta(x), \quad a \neq 0, \\ x\delta(x) &= 0, \\ \delta(x^2 - a^2) &= \frac{1}{2a}[\delta(x + a) + \delta(x - a)], \quad a \neq 0.\end{aligned}$$

There should be joy in seeing a Dirac delta function in an integral as it means evaluation of the integral is trivial (but remember, the delta function only has meaning under an integral sign).

14.2.5 Differential Equations

Differential equations are equations that relate derivatives of various orders. For example, consider

$$f'(x) = f(x).$$

From the list of derivatives above we see that $f(x) = \exp(x) + C$, where C is an arbitrary number (constant). Why must we add C ? Because by specifying the derivative of f , we are providing only information about the *slope* of the function at x . All of the functions $f(x) = \exp(x) + C$ have the same slope at x . Thus, there are an infinite number of solutions to the differential equation above. If the differential equation represents a physical process, we need to figure out which value of C applies in the given problem, and this requires some additional information. This information is usually referred to as *initial* or *boundary* conditions. A simple example illustrating this is radioactive decay, for which the *rate of change* in the number of decay candidates present at any given time is proportional to the number of decay candidates (N) present at any given time:

$$\frac{dN}{dt} = -\lambda N,$$

where λ is a constant and the minus sign is due to the fact that N *decreases* with time, so its change must be negative. Rewriting, using our table of derivatives, and working backward, we have

$$d \ln(N) = \frac{dN}{N} = -\lambda dt \quad \text{so} \quad \ln N = \lambda t + C$$

What must C be? If we have $N = N_0$ present at $t = 0$, then $C = \ln N_0$, so

$$N = N_0 \exp(-\lambda t).$$

This provides the number of undecayed atoms present at time t , given N_0 present at $t = 0$.

Many of the differential equations of mathematical physics are second order. A simple example is Newton's second law, which states that in one dimension, e.g., x ,

$$F = ma = m \frac{dv}{dt} = m \frac{d^2x}{dt^2},$$

where F is the force acting on a particle, m is the (constant) mass of the particle, v is the velocity of the particle, and x is the position of the particle along the x -axis. A simple, and poignant, example is that of a mass attached to a spring (force constant k) moving in a viscous fluid (damping constant b) and subjected to an external force that is harmonic in time. For this, Newton's law reads

$$m \frac{d^2x}{dt^2} = -k(x - x_0) - b \frac{dx}{dt} + F_0 \cos(\omega t),$$

where x_0 is the equilibrium position of the mass. If we move the origin to $x = x_0$ rather than $x = 0$, then we can remove x_0 from the equation and rewrite this as

$$\frac{d^2x}{dt^2} + 2\gamma \frac{dx}{dt} + \omega_0^2 x = \frac{F_0}{m} \cos(\omega t), \quad (14.6)$$

where $\gamma = b/2m$ and $\omega_0^2 = k/m$. Equation (14.6) is in the form

$$a_2 \frac{d^2x}{dt^2} + a_1 \frac{dx}{dt} + a_0 x = f(t). \quad (14.7)$$

It is called a second order *inhomogeneous* differential equation with constant coefficients (the a_i 's). If $f(t) = 0$ the differential equation is said to be *homogeneous*. We now describe how to solve differential equations such as Eq. (14.7).

Consider first the homogeneous counterpart to the above differential equation,

$$a_2 \frac{d^2x}{dt^2} + a_1 \frac{dx}{dt} + a_0 x = 0,$$

There is a theorem that states that the general solution of this equation consists of an arbitrary linear sum of two *linearly independent* functions $x_1(t)$ and $x_2(t)$, each of which satisfy the differential equation, i.e., if

$$a_2 \frac{d^2x_1}{dt^2} + a_1 \frac{dx_1}{dt} + a_0 x_1 = 0,$$

and

$$a_2 \frac{d^2x_2}{dt^2} + a_1 \frac{dx_2}{dt} + a_0 x_2 = 0,$$

then the general solution is

$$x(t) = C_1 x_1(t) + C_2 x_2(t).$$

What does linear independence mean? By definition $x_1(t)$ and $x_2(t)$ are linearly independent if the only solution to $c_1 x_1(t) + c_2 x_2(t) = 0$ (where the c 's are constants) that is valid for all t , is $c_1 = c_2 = 0$. Thus, for example, $x_1(t) = t$ and $x_2(t) = 2t$ are not linearly independent because, if $c_1 + 2c_2 = 0$, the condition is satisfied for all t without $c_1 = c_2 = 0$; however, if $x_1(t) = \exp(t)$ and $x_2(t) = \exp(2t)$, these functions are linearly independent since, although $c_1 \exp(t) + c_2 \exp(2t) = 0$ can be satisfied for some particular t (and c 's), it cannot be satisfied for all t unless $c_1 = c_2 = 0$.

There is a simple prescription for finding the general solution of Eq. (14.7): first find the general solution $x_h(t)$ to the homogeneous counterpart,

$$a_2 \frac{d^2 x_h}{dt^2} + a_1 \frac{dx_h}{dt} + a_0 x_h = 0,$$

which must contain two arbitrary constants as above; second find *any* function $x_i(t)$ such that

$$a_2 \frac{d^2 x_i}{dt^2} + a_1 \frac{dx_i}{dt} + a_0 x_i = f(t);$$

then the general solution to Eq. (14.7) is

$$x(t) = x_h(t) + x_i(t).$$

For the special case we are considering, where the coefficients (the a_i 's) are constants, there is a particularly simple way of solving the homogeneous equation for $x_h(t)$: we assume a solution of the form $x_h(t) = C \exp(pt)$, where C and p are constants and determine the value of p that makes x_h a solution. Let's try it on the homogeneous equation above:

$$\begin{aligned} a_2 \frac{d^2 x_h}{dt^2} + a_1 \frac{dx_h}{dt} + a_0 x_h &= 0 \quad \text{with} \quad x_h(t) = C \exp(pt) \\ \implies a_2 p^2 + a_1 p + a_0 &= 0 \quad \text{or} \quad p = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2 a_0}}{2a_2}, \end{aligned}$$

using the quadratic formula. Each of the two values of p provides a solution, so the general solution for $x_h(t)$ is¹¹

$$x_h(t) = C_+ \exp(p_+ t) + C_- \exp(p_- t).$$

There are several methods for finding a single solution for $x_i(t)$; however, as we will see below, a simple guess may sometimes suffice.

¹¹If $a_1^2 - 4a_2 a_0 = 0$, there is only one value of p , so only one solution, $x_1(t)$, is provided. In this case, a second, linearly independent solution can be shown to be $x_2(t) = t x_1(t)$.

As an illustrative example of the whole procedure, we now apply these steps to Eq. (14.6). First, let $x_h = C \exp(pt)$. Then

$$p^2 + 2\gamma p + \omega_0^2 = 0 \quad \text{or} \quad p = -\gamma \pm \sqrt{\gamma^2 - \omega_0^2}.$$

There are three possibilities depending on the relative magnitudes of γ and ω_0 :

$$\begin{aligned} (1) \quad & \gamma = \omega_0 \implies x_h(t) = (C_1 + C_2 t) \exp(-\gamma t); \\ (2) \quad & \gamma > \omega_0 \implies x_h(t) = C_+ \exp(-\gamma_+ t) + C_- \exp(-\gamma_- t); \text{ and} \\ (3) \quad & \gamma < \omega_0 \implies x_h(t) = \exp(-\gamma t) \left(C_+ \exp(+i\omega_1 t) + C_- \exp(-i\omega_1 t) \right). \end{aligned} \tag{14.8}$$

In these equations,

$$\begin{aligned} \gamma_{\pm} &= -\gamma \pm \sqrt{\gamma^2 - \omega_0^2}, \\ \omega_1^2 &= \omega_0^2 - \gamma^2, \end{aligned}$$

and $i = \sqrt{-1}$. Now that we have found the general solution the homogeneous counterpart of Eq. (14.6), we need to find *any* solution to the inhomogeneous equation. To effect this we consider its complex counterpart

$$\frac{d^2 z}{dt^2} + 2\gamma \frac{dz}{dt} + \omega_0^2 z = \frac{F_0}{m} \exp(i\omega t), \tag{14.9}$$

where $z = x + iy$. Note that the real part of Eq. (14.9) is just Eq. (14.6), so the solution $x(t)$ is just $\Re[z(t)]$. Since the derivative of an exponential function is proportional to the same exponential function, we can guess that a solution must be of the form $z(t) = A \exp(i\omega t)$, where the constant A is yet to be determined. Inserting our guess into Eq. (14.9) results in

$$-\omega^2 A + 2i\gamma\omega A + \omega_0^2 A = \frac{F_0}{m} \quad \text{or} \quad A = \frac{F_0/m}{\omega_0^2 - \omega^2 + 2i\gamma\omega}.$$

As might have been expected, A is a complex number, which we need to simplify. In the manner discussed earlier, we write

$$\omega_0^2 - \omega^2 + 2i\gamma\omega = \sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2} \exp(i\phi),$$

where

$$\tan \phi = \frac{2\gamma\omega}{\omega_0^2 - \omega^2}$$

so

$$A = \frac{(F_0/m) \exp(-i\phi)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}}$$

and

$$z(t) = \frac{(F_0/m) \exp(i\omega t - i\phi)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}} \quad \text{or} \quad x_i(t) = \frac{(F_0/m)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}} \cos(\omega t - \phi).$$

The general solution to Eq. (14.6) is then

$$x(t) = x_h(t) + \frac{(F_0/m)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}} \cos(\omega t - \phi), \quad (14.10)$$

where $x_h(t)$ is one of the three possibilities given in Eq. (14.8). We can evaluate the arbitrary constants C_+ and C_- (or C_1 and C_2) in the solution by using the *initial conditions*, the position and velocity of the mass at some arbitrary time, e.g., say $t = 0$, after which Eq. (14.10) provides the position of the mass for all time. In all three cases in Eq. (14.8), $x_h \rightarrow 0$ as $t \rightarrow \infty$, i.e., the solution to the homogeneous equation dies away in time and we are left only with

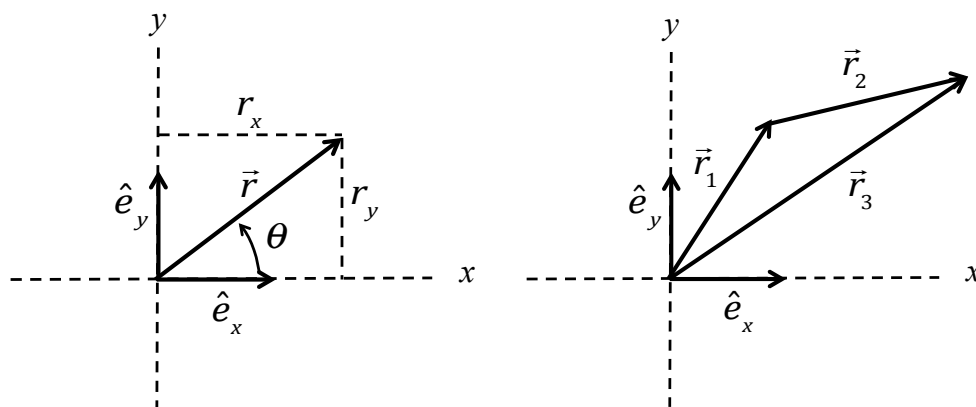
$$x(t) \rightarrow \frac{(F_0/m)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}} \cos(\omega t - \phi) \quad \text{as} \quad t \rightarrow \infty.$$

This is called the *steady state* part of the solution and x_h is called the *transient* part of the solution. Thus, after the transient dies away, the motion of the mass is purely sinusoidal with a definite phase difference (ϕ) in time between it and the driving force. The fact that the denominator in the solution becomes small for ω close to ω_0 , leading to large values of x , is the mathematical demonstration of the phenomenon called *resonance*. In our study of electromagnetic fields in matter, we used the techniques described here (with some notational variations) to discuss the dynamic properties of dielectrics (Section 1.4.5).

In Chapter 1, we generally represented the complex electric field of a traveling wave (in the $+x$ direction) as $\vec{E}(x, t) = \vec{E}_0 \exp[i(\kappa x - \omega t)]$, and when we needed to know the actual electric field, we took the real part to get $\vec{E}(x, t) = \vec{E}_0 \cos(\kappa x - \omega t)$. If we wanted to compute the average energy density associated with one oscillation of the electric field, it would be

$$\begin{aligned} \langle u_e \rangle &= \frac{1}{\tau} \int_0^\tau \frac{\epsilon_0}{2} \vec{E}(x, t) \bullet \vec{E}(x, t) dt = \left[\frac{\omega}{2\pi} \right] \frac{\epsilon_0}{2} \int_0^{2\pi/\omega} \vec{E}(x, t) \bullet \vec{E}(x, t) dt \\ &= \left[\frac{\omega}{2\pi} \right] \frac{\epsilon_0}{2} \int_0^{2\pi/\omega} E_0^2 \cos^2(\kappa x - \omega t) dt = \frac{1}{2} \frac{\epsilon_0 E_0^2}{2} = \frac{1}{2} \frac{\epsilon_0}{2} \vec{E}^*(x, t) \bullet \vec{E}(x, t), \end{aligned}$$

where τ is the period of one oscillation. Thus, the product $\vec{E}^*(x, t) \bullet \vec{E}(x, t)$ is related to the time-average energy density, *not* the *instantaneous* energy density. More generally, the time average of the square of an oscillatory function $f(t)$ over one cycle can be replaced by $f^*(t)f(t)/2$; however, it is safer to convert the complex functions to their real counterparts before computing quantities like energy that are dependent on their *products*.



14.3 Vectors and Multivariable Calculus

The calculus of functions of more than one variable is closely related to the calculus of vectors, and are reviewed together in this section.

14.3.1 Vectors — the Elements

In elementary physics we usually call quantities like temperature, that are described by a single number, *scalars*. In contrast, quantities like velocity, which have a certain magnitude (e.g., 10 mph) and direction (e.g., northeast) are called *vectors*. Another example of a vector is the displacement vector \vec{d} defined as follows.¹² Assume you are walking on a plane (the x - y plane shown in the figure above; left panel). You start at the origin and walk along the $+x$ axis a distance r_x and then parallel to the y axis a distance r_y . Draw an arrow from the point $(0,0)$ to the point (r_x, r_y) (pointing *away* from the origin) and label it \vec{r} . The vector \vec{r} is the displacement vector associated with your walk. It has units of length, e.g., meters, and makes an angle θ with the x axis. The *magnitude* of the displacement vector is simply its length, i.e., $\sqrt{r_x^2 + r_y^2}$, and is indicated by r (without the arrow) or by $|\vec{r}|$. The magnitude of a vector is a scalar. A *unit vector* pointing in the same direction as \vec{r} , and indicated by \hat{r} , is defined through $\hat{r} \triangleq \vec{r}/r$. Note that the magnitude of the unit vector \hat{r} is $|\hat{r}| = |\vec{r}/r| = |\vec{r}|/r = r/r = 1$ (with *no* units!). Here we see that multiplication of a vector (\vec{r}) by a scalar ($1/r$) simply changes the magnitude of the vector, *not* its direction. In the left figure, the quantities \hat{e}_x and \hat{e}_y are (cartesian) unit vectors pointing toward the $+x$ and $+y$ directions, respectively.

¹²Here vectors are indicated by an arrow over the symbol. Unit vectors are indicated by a carrot over the symbol, e.g., \hat{d} .

Now look at the right panel in the figure. Assume that you first take a path that results in at the displacement vector \vec{r}_1 , and then take another path that results in \vec{r}_2 . The displacement vector \vec{r}_3 from the original starting point to the finish is given by $\vec{r}_3 = \vec{r}_1 + \vec{r}_2$, which defines vector addition. Using this rule for vector addition one sees that in the left panel of the figure

$$\vec{r} = \hat{e}_x r_x + \hat{e}_y r_y = \hat{e}_x r \cos \theta + \hat{e}_y r \sin \theta,$$

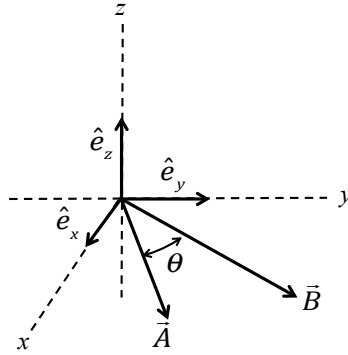
and in the right panel

$$\vec{r}_3 = \hat{e}_x r_{1x} + \hat{e}_y r_{1y} + \hat{e}_x r_{2x} + \hat{e}_y r_{2y} = \hat{e}_x (r_{1x} + r_{2x}) + \hat{e}_y (r_{1y} + r_{2y}) = \hat{e}_x r_{3x} + \hat{e}_y r_{3y}$$

The quantities multiplying the cartesian unit vectors in these equations are called the *components* of the given vector, e.g., r_{1x} is the x component of \vec{r}_1 , etc., so we see that to add two vectors one need only add their individual x and y components.

The negative of a vector is just a vector pointing in the opposite direction (or a vector whose components are the negative of the original vector. Thus, $\vec{r} - \vec{r} = \vec{r} + (-\vec{r}) = \vec{0}$ according to the rule of vector addition (usually the vector sign is left off the null or zero vector). A vector times a scalar simply changes the magnitude of the vector, not its direction, i.e., $|a\vec{r}| = a|\vec{r}|$.

We have seen that vectors can be added and subtracted. They can also be multiplied. There are two ways for multiplication. The first yields a scalar and is called the “dot” or “scalar” product. Referring to the figure below, where the vectors \vec{A} and \vec{B} are both in



the x - y plane, the scalar product is given by

$$\vec{A} \bullet \vec{B} \triangleq AB \cos \theta.$$

The other product of two vectors yields a vector as follows. If the two vectors are in the x - y plane (as in the figure) then the “cross” or “vector” product is defined as

$$\vec{A} \times \vec{B} \triangleq (AB \sin \theta) \hat{e}_z,$$

so the magnitude of the cross product is $|AB \sin \theta|$ and the direction is \hat{e}_z . Finding the direction of the cross product seems give some people trouble, but it is easy. In words, $\vec{A} \times \vec{B}$ is perpendicular to the plane formed by \vec{A} and \vec{B} . To find which direction it points normal to the plane; (1) place the fingers of your right hand along the first vector (\vec{A}) with the palm of your right hand pointing toward the second vector (\vec{B}) through the *smaller* of the two angles between \vec{A} and \vec{B} (in the figure the smaller angle is θ and the larger is $\pi - \theta$), and (2) the direction of $\vec{A} \times \vec{B}$ is given by the general direction in which your thumb is pointing. Try it and verify that if $\theta > \pi$ the direction for the case in the figure will be $-\hat{e}_z$, just as the formula says. Note $\vec{B} \times \vec{A} = -\vec{A} \times \vec{B}$, but $\vec{B} \bullet \vec{A} = \vec{A} \bullet \vec{B}$.

Can we find these products of two vectors in terms of component vectors? Let's try the dot product.

$$\begin{aligned}\vec{A} \bullet \vec{B} &= (A_x \hat{e}_x + A_y \hat{e}_y) \bullet (B_x \hat{e}_x + B_y \hat{e}_y) \\ &= A_x \hat{e}_x \bullet B_x \hat{e}_x + A_x \hat{e}_x \bullet B_y \hat{e}_y + A_y \hat{e}_y \bullet B_x \hat{e}_x + A_y \hat{e}_y \bullet B_y \hat{e}_y,\end{aligned}$$

but $\hat{e}_x \bullet \hat{e}_x = \hat{e}_y \bullet \hat{e}_y = 1$ and $\hat{e}_y \bullet \hat{e}_x = 0$, so

$$\vec{A} \bullet \vec{B} = A_x B_x + A_y B_y.$$

For the cross product,

$$\begin{aligned}\vec{A} \times \vec{B} &= (A_x \hat{e}_x + A_y \hat{e}_y) \times (B_x \hat{e}_x + B_y \hat{e}_y) \\ &= A_x \hat{e}_x \times B_x \hat{e}_x + A_x \hat{e}_x \times B_y \hat{e}_y + A_y \hat{e}_y \times B_x \hat{e}_x + A_y \hat{e}_y \times B_y \hat{e}_y,\end{aligned}$$

but $\hat{e}_x \times \hat{e}_x = \hat{e}_y \times \hat{e}_y = 0$ and $\hat{e}_x \times \hat{e}_y = \hat{e}_z = -\hat{e}_y \times \hat{e}_x$, so

$$\vec{A} \times \vec{B} = (A_x B_y - A_y B_x) \hat{e}_z.$$

In three dimensions (and cartesian coordinates), where $\vec{A} = A_x \hat{e}_x + A_y \hat{e}_y + A_z \hat{e}_z$ and $\vec{B} = B_x \hat{e}_x + B_y \hat{e}_y + B_z \hat{e}_z$, it is easy to show that

$$\vec{A} \bullet \vec{B} = A_x B_x + A_y B_y + A_z B_z,$$

and the cross product is given by the determinant

$$\vec{A} \times \vec{B} = \begin{vmatrix} \hat{e}_x & \hat{e}_y & \hat{e}_z \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}$$

14.3.2 Functions of Several Variables

In complete analogy to a function $f(x)$ of a single variable x , we can have functions of several variables, e.g., $f(x, y, \dots)$. Examples are $f(x, y) = \sin(x) \cos(y)$, or $f(x, y) = xy$.

To emphasize the fact that $f(x, y)$ is a single number (with or without dimensions), it is sometimes written $z = f(x, y)$. Another way to write $f(x, y)$ is $f(\vec{r})$, where \vec{r} is the vector shown on the left graph page 909, but now has components x and y , i.e., $\vec{r} = x\hat{e}_x + y\hat{e}_y$. When written this way, \vec{r} is called the *position* vector: it indicates the “position” of the point (x, y) .¹³ The components of \vec{r} are related to *direction cosines* of \vec{r} : $x = \vec{r} \bullet \hat{e}_x$ and $y = \vec{r} \bullet \hat{e}_y$, so

$$\vec{r} = (\vec{r} \bullet \hat{e}_x) \hat{e}_x + (\vec{r} \bullet \hat{e}_y) \hat{e}_y,$$

where the dot products $(\vec{r} \bullet \hat{e}_x)$ and $(\vec{r} \bullet \hat{e}_y)$ are r times the cosines of the angles that \vec{r} makes with the x and y axes, respectively, the “direction cosines.”

14.3.3 Differentiation of Multivariable Functions and Vectors

As in the case of single variable calculus, differentiation is motivated by asking questions of the sort — “if I change x a little but do not change y , how much does f change?” We define the change in f for an infinitesimal change in x to be the partial derivative of f with respect to x , i.e.,

$$\frac{\partial f}{\partial x} \triangleq \lim_{\Delta x \rightarrow 0} \left[\frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} \right].$$

Similarly,

$$\frac{\partial f}{\partial y} \triangleq \lim_{\Delta y \rightarrow 0} \left[\frac{f(x, y + \Delta y) - f(x, y)}{\Delta y} \right].$$

When there might be confusion as to what is being held constant in the derivative (in the last equation it is x), one often indicates this by putting parenthesis around the derivative with a subscript indicating the variable(s) held constant, i.e., $(\partial f / \partial y)_x$.¹⁴ What if both x and y change a little? The change in f , Δf is

$$\begin{aligned} \Delta f &= f(x + \Delta x, y + \Delta y) - f(x, y) \\ &= f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y) + f(x + \Delta x, y) - f(x, y) \\ &= \frac{\partial}{\partial y}(f(x + \Delta x, y))\Delta y + \frac{\partial}{\partial x}f(x, y)\Delta x, \end{aligned}$$

or letting Δx and Δy become infinitesimals,

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy.$$

¹³This can clearly be extended to any number of variables, i.e., dimensions.

¹⁴It's of course possible that something peculiar is being held constant, e.g., $2x + y$, then the notation would be $(\partial f / \partial y)_{2x+y}$, and when y changes x changes as well, but in the prescribed manner.

If we write $\vec{r} = \hat{e}_x x + \hat{e}_y y$, then the change in \vec{r} is $\Delta\vec{r} = \hat{e}_x \Delta x + \hat{e}_y \Delta y$. It is useful to define a vector called the gradient of f , written ∇f as

$$\nabla f \triangleq \hat{e}_x \frac{\partial f}{\partial x} + \hat{e}_y \frac{\partial f}{\partial y},$$

then

$$\Delta f = \nabla f \bullet \Delta\vec{r} \quad \text{or} \quad df = \nabla f \bullet d\vec{r}. \quad (14.11)$$

The symbol ∇ is called the “gradient operator” or just “*del*.” It is a vector operator¹⁵ defined through

$$\nabla = \hat{e}_x \frac{\partial}{\partial x} + \hat{e}_y \frac{\partial}{\partial y}.$$

Consider a point \vec{r} . If one moves from \vec{r} to $\vec{r} + \Delta\vec{r}$, the change in f is $\nabla f(\vec{r}) \bullet \Delta\vec{r}$. If $f(\vec{r} + \Delta\vec{r})$ is the same as $f(\vec{r})$, i.e., f doesn’t change, then either $\nabla f(\vec{r}) = 0$ or $\nabla f(\vec{r})$ is perpendicular to $\Delta\vec{r}$. If f doesn’t change that means one has moved along a line of constant f , i.e., along a contour line of f .¹⁶ Thus, $\nabla f(\vec{r})$ is everywhere perpendicular to lines of constant values of f , and in fact points in the direction of *increasing* f (i.e., uphill on a topographic map). Thus, if we have a curve $y = f(x)$, which we could write as $g(x, y) = y - f(x) = c$, where c is a constant and equal to zero for the curve we are interested in, we can find a vector \vec{n} that is normal to the curve through $\vec{n} = \nabla g(x, y)$, or $\hat{n} = \nabla g(x, y) / |\nabla g(x, y)|$. If we write $d\vec{r}$ in Eq. (14.11) as $d\vec{r} = \hat{\ell} d\ell$, then the quantity

$$\frac{df}{d\ell} = \nabla f(\vec{r}) \bullet \hat{\ell}$$

is called the *directional derivative* of f in the direction $\hat{\ell}$.

The extension of this to three (or more) dimensions is immediate: just add another coordinate, e.g., in three dimensions

$$\nabla = \hat{e}_x \frac{\partial}{\partial x} + \hat{e}_y \frac{\partial}{\partial y} + \hat{e}_z \frac{\partial}{\partial z} \quad \text{and} \quad \vec{r} = \hat{e}_x x + \hat{e}_y y + \hat{e}_z z, \text{ etc.,}$$

and $\nabla f(\vec{r})$ is normal to *surfaces* of constant $f(\vec{r})$, and can be used to find the normal to such surfaces.¹⁷

There are other operations that can be defined with the gradient operator. Two that we need are the “divergence” and the “curl.” Given a vector function of the position vector,

¹⁵It is customary to leave *off* the vector sign for this operator.

¹⁶The term “contour line” comes from an analogy to topographic maps, where lines of constant height above sea level are called contour lines.

¹⁷As an example, consider a sphere centered on the origin, i.e., $x^2 + y^2 + z^2 = R^2$. Write this as $g(x, y, z) = x^2 + y^2 + z^2 - R^2 = c$, where $c = 0$ for the particular sphere of interest. Then $\nabla g = 2x\hat{e}_x + 2y\hat{e}_y + 2z\hat{e}_z = 2\vec{r}$, so the unit normal to the surface is \hat{r} .

$\vec{A}(\vec{r}) = A_x(\vec{r})\hat{e}_x + A_y(\vec{r})\hat{e}_y + A_z(\vec{r})\hat{e}_z$ in three dimensions,¹⁸ the divergence of \vec{A} is defined as

$$\nabla \bullet \vec{A} \triangleq \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z},$$

and the curl of \vec{A} is given by

$$\nabla \times \vec{A} = \begin{vmatrix} \hat{e}_x & \hat{e}_y & \hat{e}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}$$

where the partial derivatives operate on the components of \vec{A} , i.e.,

$$\nabla \times \vec{A} = \hat{e}_x \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + \hat{e}_y \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + \hat{e}_z \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right).$$

A second derivative which can be formed and is very important is $\nabla(\nabla \bullet \vec{A}) \triangleq \nabla^2 \vec{A}$, and is easily seen to be

$$\nabla^2 \vec{A} = \frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_y}{\partial y^2} + \frac{\partial^2 A_z}{\partial z^2},$$

where

$$\frac{\partial^2 A_x}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial A_x}{\partial x} \right), \text{ etc.}$$

the operator ∇^2 is called the *Laplacian* operator. Note that one can apply the Laplacian operator to both scalar and vector functions.

There are many relationships among vectors and their derivatives that are normally called “vector identities.” We list several below.

$$\begin{aligned} \vec{A} \bullet (\vec{B} \times \vec{C}) &= \vec{B} \bullet (\vec{C} \times \vec{A}) = \vec{C} \bullet (\vec{A} \times \vec{B}) \\ \vec{A} \times (\vec{B} \times \vec{C}) &= \vec{B}(\vec{A} \bullet \vec{C}) - \vec{C}(\vec{A} \bullet \vec{B}) \\ \nabla \times (\vec{A} \times \vec{B}) &= \vec{B} \bullet (\nabla \times \vec{A}) - \vec{A} \bullet (\nabla \times \vec{B}) \\ \nabla(fg) &= f\nabla(f) + g\nabla(f) \\ \nabla \bullet (f\vec{A}) &= f(\nabla \bullet \vec{A}) + \vec{A} \bullet \nabla(f) \\ \nabla \times (f\vec{A}) &= f(\nabla \times \vec{A}) - \vec{A} \times \nabla(f) \\ \nabla \bullet (\nabla \times \vec{A}) &= 0 \\ \nabla \times (\nabla f) &= 0 \\ \nabla \times (\nabla \times \vec{A}) &= \nabla(\nabla \bullet \vec{A}) - \nabla^2 \vec{A}. \end{aligned} \tag{14.12}$$

¹⁸This is usually called a *vector field*. At each point in space there is an associated vector \vec{A} , e.g., the velocity of a fluid at each point in space.

14.3.4 Integration of Multivariable Functions and Vectors

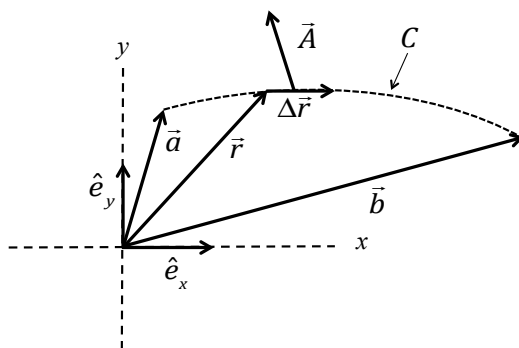
The integration of functions of more than one variable proceeds in a manner similar to single variables. Thus,

$$\int_a^b \int_c^d f(x, y) dx dy \triangleq \lim_{N_x \rightarrow \infty} \lim_{N_y \rightarrow \infty} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} f(\xi_i, \zeta_j) \Delta x \Delta y,$$

where $N_x \Delta x = b - a$, $N_y \Delta y = d - c$, $x_i - \Delta x/2 \leq \xi_i \leq x_i + \Delta x/2$, and $y_i - \Delta y/2 \leq \zeta_i \leq y_i + \Delta y/2$. One also encounters “line integrals,” for example

$$\int_{\vec{a}}^{\vec{b}} \vec{A}(\vec{r}) \bullet d\vec{r},$$

where $d\vec{r}$ is along some curve C , as shown for the case of two dimensions in the figure below (the dotted line). The definition of the integral is



$$\int_{\vec{a}}^{\vec{b}} \vec{A}(\vec{r}) \bullet d\vec{r} \triangleq \lim_{N \rightarrow \infty} \sum_{i=1}^N \vec{A}(\vec{\xi}_i) \bullet \Delta \vec{r}_i,$$

where the $\Delta \vec{r}_i$'s, joined tip-to-tail, constitute the curve C , and $\vec{\xi}_i$ is a point on the curve between \vec{r} and $\vec{r} + \Delta \vec{r}$. Choose $|\Delta \vec{r}_i| = L/N$, so that $L = \lim_{N \rightarrow \infty} \sum_{i=1}^N |\Delta \vec{r}_i|$ is the arc length of the curve C . Clearly, it follows that

$$\int_{\vec{a}}^{\vec{b}} d\vec{r} = \lim_{N \rightarrow \infty} \sum_{i=1}^N \Delta \vec{r}_i = \vec{b} - \vec{a}.$$

This is an example of an integral that is independent of the path C , i.e., it only depends on the end points \vec{b} and \vec{a} . Another path integral that doesn't depend on the path of

integration is

$$\int_{\vec{a}}^{\vec{b}} df(\vec{r}) = \int_{\vec{a}}^{\vec{b}} \nabla(f(\vec{r})) \bullet d\vec{r} = f(\vec{b}) - f(\vec{a}). \quad (14.13)$$

Note that when the path is a closed curve, i.e., $\vec{b} = \vec{a}$,

$$\oint \nabla(f(\vec{r})) \bullet d\vec{r} = 0.$$

Area and volume integrals are defined in a similar manner.

Two very important theorems relating multiple integrals are the divergence theorem and Stokes theorem. The divergence theorem states

$$\int_V \nabla \bullet \vec{B}(\vec{r}) dV = \int_S \vec{B}(\vec{r}) \bullet \hat{n} dA, \quad (14.14)$$

where V is a volume bounded by the surface S . The quantity dV is the volume element, (e.g., $dx dy dz$ in cartesian coordinates), dA is an element of surface area (e.g., $dx dy$ if a part of the surface is parallel to the x - y plane), \hat{n} is the (outward) normal to the surface at dA . The name “divergence” comes from the fact that if, for example, \vec{B} were a fluid velocity, when the surface integral is positive the fluid would have a net flow *away* from the volume, i.e., *diverge* from the interior of the volume.

Stokes theorem states

$$\int_S \hat{n} \bullet (\nabla \times \vec{B}(\vec{r})) dA = \oint_C \vec{B} \bullet d\vec{r}, \quad (14.15)$$

where S is a surface bounded by the curve C , i.e., C is the *edge* of S . The unit vector \hat{n} is again the surface normal at dA , but as the surface in general is not closed, how do we specify the direction? The direction of \hat{n} is related to the direction of integration around the closed curve. To see how, imagine that S is a disk and C is its edge. If one wraps one’s fingers of the right hand around C in the direction of integration (the direction $d\vec{r}$ ’s point), then \hat{n} is normal to the area pointing in the general direction of the thumb. If S is morphed into any surface, but still bounded by C , the normal is still on the *same* side of S . This is a quite amazing theorem. Consider a coffee cup and let C be the rim. The theorem says the the integral over the surface formed by the plane of the top is the same as the integral over the sides, since both end on C . In addition, what if $\vec{B} = \nabla f$? Then the integral on the right hand side is zero, so $\int_S \hat{n} \bullet (\nabla \times \nabla f(\vec{r})) dA = 0$. But this is true regardless of the shape of the surface S . The only way this can be true for all surfaces is if the integrand is zero, i.e., $\nabla \times \nabla f(\vec{r}) = 0$. This is one of the vector identities listed earlier. In a manner similar to Eq. (14.4), Eqs. (14.13), (14.14), and (14.15) are fundamental theorems of vector calculus.

In analogy to single variable calculus, we can define the Dirac delta function in multiple dimensions. Multidimensional Dirac delta functions work in a manner similar to the single dimensional version we discussed earlier. For example, in two dimensions

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x - x_0) \delta(y - y_0) dx dy = f(x_0, y_0).$$

In vector notation this becomes

$$\int_{\text{All A}} f(\vec{r}) \delta^{(2)}(\vec{r} - \vec{r}_0) dA = f(\vec{r}_0),$$

where dA is the element of area in the plane.¹⁹ Now, what if we use plane polar coordinates for dA , i.e., $dA = r dr d\theta$, in the above equation. We may be tempted to write $\delta^{(2)}(\vec{r} - \vec{r}_0) = \delta(r - r_0) \delta(\theta - \theta_0)$. However, if we did that, then the result would be

$$\int_{\text{All A}} f(\vec{r}) \delta^{(2)}(\vec{r} - \vec{r}_0) dA = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} f(r, \theta) \delta(r - r_0) \delta(\theta - \theta_0) r dr d\theta = r_0 f(r_0, \theta_0),$$

which is not the same answer found using cartesian coordinates (the correct one). To get the correct answer we need to take

$$\delta^{(2)}(\vec{r} - \vec{r}_0) = \frac{\delta(r - r_0) \delta(\theta - \theta_0)}{r} = \frac{\delta(r - r_0) \delta(\theta - \theta_0)}{r_0}.$$

In the general case, if the multidimensional volume element $dx_1 dx_2 dx_3, \dots$, becomes $f(y_1, y_2, y_3, \dots) dy_1 dy_2 dy_3 \dots$ when a transformation is made from the cartesian $\vec{x} = (x_1, x_2, x_3, \dots)$ to the non-cartesian $\vec{y} = (y_1, y_2, y_3, \dots)$ coordinate system, then the multidimensional Dirac delta function becomes

$$\delta^{(n)}(\vec{r} - \vec{r}_0) = \delta(x_1 - x_{10}) \delta(x_2 - x_{20}) \delta(x_3 - x_{30}) \dots \longrightarrow \frac{\delta(y_1 - y_{10}) \delta(y_2 - y_{20}) \delta(y_3 - y_{30}) \dots}{f(y_1, y_2, y_3, \dots)}.$$

Clearly, the appearance of a Dirac delta function in an integral greatly simplifies its evaluation. In the text we used a two dimensional delta function to represent the radiance of a parallel beam of irradiance F_0 , i.e., $L(\hat{\xi}) = F_0 \delta^{(2)}(\hat{\xi} - \hat{\xi}_0)$, so

$$\int_{\Omega} L(\hat{\xi}) d\Omega(\hat{\xi}) = F_0,$$

if the direction $\hat{\xi}_0$ is in the range of directions within the solid angle Ω , and zero otherwise.

¹⁹Note, the integral need not extend over the whole plane. If it covers an area S in the plane, then the value is $f(\vec{r}_0)$ if the point \vec{r}_0 is within S , and zero otherwise.

Finally, we state the Leibnitz integral rule for differentiating an integral. It is

$$\frac{d}{dy} \left[\int_{a(y)}^{b(y)} f(x, y) dx \right] = \int_{a(y)}^{b(y)} \frac{\partial f(x, y)}{\partial y} dx + f(b, y) \frac{\partial b(y)}{\partial y} - f(a, y) \frac{\partial a(y)}{\partial y}.$$

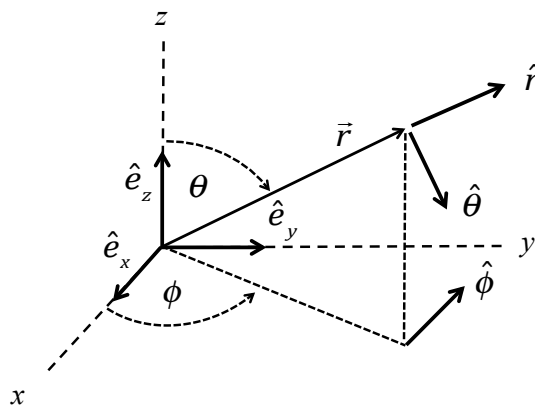
Therefore if the limits are fixed, an ordinary derivative operating on an integral becomes a *partial* derivative operating on the *integrand*.

14.3.5 Functions and Vectors in Spherical Coordinates

In many cases it is advantageous to represent a vector in another coordinate system rather than the cartesian system we have used so far in our discussion of multivariable functions and vectors. We now examine their representation in *spherical coordinates*. The transformation from cartesian coordinates (x, y, z) to spherical coordinates (r, θ, ϕ) is given by (see figure below)

$$\begin{aligned} x &= r \sin \theta \cos \phi, \\ y &= r \sin \theta \sin \phi, \\ z &= r \cos \theta, \end{aligned}$$

so



$$\vec{r} = \hat{e}_x x + \hat{e}_y y + \hat{e}_z z = \hat{e}_x r \sin \theta \cos \phi + \hat{e}_y r \sin \theta \sin \phi + \hat{e}_z r \cos \theta,$$

and

$$\hat{r} = \hat{e}_x \sin \theta \cos \phi + \hat{e}_y \sin \theta \sin \phi + \hat{e}_z \cos \theta.$$

The other unit vectors $\hat{\theta}$ and $\hat{\phi}$ are seen from the figure above to be²⁰

$$\begin{aligned}\hat{\theta} &= \hat{e}_x \cos \theta \cos \phi + \hat{e}_y \cos \theta \sin \phi - \hat{e}_z \sin \theta, \\ \hat{\phi} &= -\hat{e}_x \sin \phi + \hat{e}_y \cos \phi.\end{aligned}$$

The unit vector $\hat{\theta}$ is in the plane formed by \hat{e}_z and \vec{r} , while the unit vector $\hat{\phi}$ is in the x - y plane, normal to the projection of \vec{r} on that plane. Note that the set of unit vectors are mutually orthogonal so $\vec{r} \bullet \hat{\theta} = 0$, $\vec{r} \times \hat{\theta} = \hat{\phi}$, etc. We can write a small change in the position of a particle, i.e., $d\vec{r}$, as

$$d\vec{r} = \hat{e}_x dx + \hat{e}_y dy + \hat{e}_z dz = \hat{r} dr + \hat{\theta} r d\theta + \hat{\phi} r \sin \theta d\phi.$$

The volume and area elements in cartesian coordinates, $dV = dx dy dz$ and $dA = dx dy$, respectively, become $dV = r^2 \sin \theta dr d\theta d\phi$ and $dA = r^2 \sin \theta d\theta d\phi$, where dA is on the surface of a sphere centered at $r = 0$. Note that the volume and area elements in spherical coordinate are not just the products of the differentials (they couldn't be because the differentials do not all have the units of length!), but include other factors. In general, if one makes a transformation from cartesian coordinates (x, y) to another system (u, v) , where $u = u(x, y)$ and $v = v(x, y)$, the product of differentials transform according to

$$du dv = \left| \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} \right| dx dy = |J(u, v; x, y)| dx dy.$$

The determinant J is called the *Jacobian* of the transformation. In higher dimensions, the Jacobian determinant simply becomes larger (in three dimensions, it is 3×3). For a particularly simple example, consider two dimensions and the transformation from cartesian to plane polar coordinates for which the transformation equations are given by $x = r \cos \phi$, $y = r \sin \phi$. Then

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & +r \cos \phi \end{vmatrix} = r$$

and so, $dx dy = r dr d\theta$. One can of course write formulas for the gradient, divergence, and curl in terms of derivatives with respect to r , θ , and ϕ ; however, we shall not write them down as they are not actually used in the text.²¹

²⁰In the text, \hat{r} , $\hat{\theta}$, and $\hat{\phi}$, are also referred to as \hat{e}_r , \hat{e}_θ , and \hat{e}_ϕ .

²¹Virtually any book on advanced calculus has these formulas (as well as various internet sites).

14.4 Linear Transformations and Matrices

Consider a linear transformation between two coordinate systems (x, y, z) and (x', y', z') , where

$$\begin{aligned}x' &= a_{x'x}x + a_{x'y}y + a_{x'z}z \\y' &= a_{y'x}x + a_{y'y}y + a_{y'z}z \\z' &= a_{z'x}x + a_{z'y}y + a_{z'z}z,\end{aligned}$$

i.e., a point (x, y, z) in the first system is transformed to the point (x', y', z') in the second. We can write these equations as a single *matrix* equation as follows. Write the array of a 's in the following form:

$$\mathbf{A} \triangleq \begin{pmatrix} a_{x'x} & a_{x'y} & a_{x'z} \\ a_{y'x} & a_{y'y} & a_{y'z} \\ a_{z'x} & a_{z'y} & a_{z'z} \end{pmatrix}.$$

Also write the coordinates (x', y', z') and (x, y, z) as column arrays \mathbf{x}' and \mathbf{x} , where

$$\mathbf{x}' \triangleq \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \quad \text{and} \quad \mathbf{x} \triangleq \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Then

$$\mathbf{x}' = \mathbf{A}\mathbf{x},$$

where for equality we must take the sum of the products of the elements in the top row of \mathbf{A} with the first (and only) column of \mathbf{x} to get the top element of the column representing \mathbf{x}' . This procedure can be simplified by relabeling (x, y, z) as (x_1, x_2, x_3) , (x', y', z') as (x'_1, x'_2, x'_3) , and $a_{x'x}$ as a_{11} , $a_{x'y}$ as a_{12} , etc. Then

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad \mathbf{x}' = \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad (14.16)$$

and the rule for finding $\mathbf{x}' = \mathbf{A}\mathbf{x}$ can be simply stated:

$$x'_i = \sum_{j=1}^3 a_{ij} x_j.$$

The object \mathbf{A} is called a 3×3 *matrix*, and \mathbf{x}' and \mathbf{x} are 3×1 matrices or *column vectors*. This algebra can be generalized. For example let \mathbf{B} and \mathbf{C} be two 3×3 matrices, then the *product* of \mathbf{B} and \mathbf{C} ,

$$\mathbf{BC}, \quad \text{means} \quad (\mathbf{BC})_{ij} = \sum_{k=1}^3 B_{ik} C_{kj},$$

where $(\mathbf{BC})_{ij}$ is the ij^{th} element of \mathbf{BC} . Likewise

$$\mathbf{CB} \quad \text{means} \quad (\mathbf{CB})_{ij} = \sum_{k=1}^3 C_{ik} B_{kj},$$

which clearly shows that $\mathbf{BC} \neq \mathbf{CB}$, i.e., matrix multiplication is *non-commutative*. Note that it is not always possible to multiply two matrices. For multiplication, as defined according to $(\mathbf{BC})_{ij} = \sum_{k=1}^p B_{ik} C_{kj}$, the number of rows of \mathbf{B} *must* be the same as the number of columns of \mathbf{C} , i.e., \mathbf{B} must be $n \times p$ and \mathbf{C} must be $p \times m$, where n and m are arbitrary.²² The sum of two matrices is as one might expect, just the sum of the corresponding elements: $(\mathbf{A} + \mathbf{B})_{ij} = A_{ij} + B_{ij}$. Clearly, to add two matrices they must be of the same size.

Given the matrix \mathbf{A} in Eq. (14.16), many related matrices can be defined: the *transpose* of \mathbf{A} , written $\tilde{\mathbf{A}}$ is the matrix with the rows and columns interchanged, i.e.,

$$\tilde{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{32} & a_{33} \end{pmatrix};$$

if $\tilde{\mathbf{A}} = \mathbf{A}$ the matrix is said to be *symmetric*; if the *elements* of \mathbf{A} (the a_{ij} 's) are real \mathbf{A} is said to be real; the *inverse* of \mathbf{A} , written \mathbf{A}^{-1} is defined such that

$$\mathbf{A}^{-1} \mathbf{A} = \mathbf{A} \mathbf{A}^{-1} = \mathbf{1}, \quad \text{where} \quad \mathbf{1} \triangleq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is called the *unit* matrix. If $\mathbf{A}^{-1} = \tilde{\mathbf{A}}$ the matrix is said to be *orthogonal*. For an example of an orthogonal matrix, consider the transformation from (x, y) to (x', y') corresponding to a rotation of the axes in the x - y plane through an angle θ :

$$\begin{aligned} x' &= x \cos \theta + y \sin \theta \\ y' &= -x \sin \theta + y \cos \theta \end{aligned}$$

so

$$\mathbf{A} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad \mathbf{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Note that

$$\tilde{\mathbf{A}} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{A}} \mathbf{A} = \mathbf{A} \tilde{\mathbf{A}} = \mathbf{1},$$

²²Note that when we designate the size of a matrix it is always the number of *rows* \times the number of *columns*, e.g., a column vector is as 3×1 matrix.

and so $\mathbf{A}^{-1} = \widetilde{\mathbf{A}}$, i.e., the matrix \mathbf{A} is orthogonal. One can show from the definition of matrix multiplication that for two matrices \mathbf{A} and \mathbf{B} , that $\widetilde{\mathbf{AB}} = \widetilde{\mathbf{B}}\widetilde{\mathbf{A}}$. If the matrix \mathbf{A} has elements with the property that $a_{ij} = 0$ when $i \neq j$, \mathbf{A} is called a *diagonal* matrix. The sum of the diagonal elements of \mathbf{A} is called its *trace* and written $Tr(\mathbf{A}) \triangleq \sum_{i=1}^n A_{ii}$. To form the product of a matrix \mathbf{A} and a scalar s , i.e., $s\mathbf{A}$, simply multiply *all* of the elements of \mathbf{A} by s .

Often we will write vectors as column matrices, i.e., the vector $\vec{V} = V_x\hat{e}_x + V_y\hat{e}_y + V_z\hat{e}_z$ is sometimes written

$$\vec{V} = V_x\hat{e}_x + V_y\hat{e}_y + V_z\hat{e}_z, \text{ or}$$

$$\mathbf{V} = \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix},$$

so, when \vec{V} is used it normally refers to the first representation, and when \mathbf{V} is used it refers to the second; however, often these two notations are used interchangeably and confusion is unlikely. When vectors are represented by column matrices, the scalar or dot product of two vectors is

$$\vec{V}_1 \bullet \vec{V}_2 = \widetilde{\mathbf{V}}_1 \mathbf{V}_2 = (V_{1_1} \quad V_{1_2} \quad V_{1_3} \quad \cdots) \begin{pmatrix} V_{2_1} \\ V_{2_2} \\ V_{2_3} \\ \vdots \end{pmatrix} = V_{1_1}V_{2_1} + V_{1_2}V_{2_2} + V_{1_3}V_{2_3} \cdots.$$

The transformation of unit vectors can also be represented by matrices. To go from cartesian to polar representation of a vector, the unit vectors are transformed via

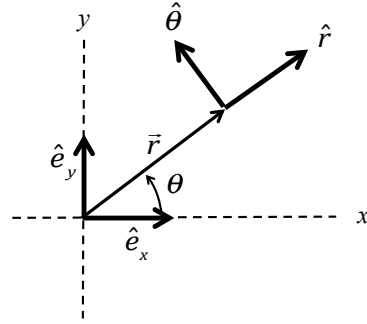
$$\hat{r} = \hat{e}_x \cos \theta + \hat{e}_y \sin \theta,$$

$$\hat{\theta} = -\hat{e}_x \sin \theta + \hat{e}_y \cos \theta,$$

which can be written

$$\begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \end{pmatrix}, \quad \text{so} \quad \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix},$$

and we see that the \hat{e}_x and \hat{e}_y are transformed into \hat{r} and $\hat{\theta}$ through²³



$$\begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix} = \mathbf{A} \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \end{pmatrix}.$$

If we have a vector $\vec{V} = V_x \hat{e}_x + V_y \hat{e}_y$ and we make this transformation, how do the components transform, i.e., in $\vec{V} = V_r \hat{r} + V_\theta \hat{\theta}$, how do V_r and V_θ relate to V_x and V_y ? This is easy to work out using the transformation equations for the unit vectors:

$$\begin{aligned} \vec{V} &= V_x \hat{e}_x + V_y \hat{e}_y \\ &= V_x(\hat{r} \cos \theta - \hat{\theta} \sin \theta) + V_y(\hat{r} \sin \theta + \hat{\theta} \cos \theta) \\ &= \hat{r}(V_x \cos \theta + V_y \sin \theta) + \hat{\theta}(-V_x \sin \theta + V_y \cos \theta) \\ &= \hat{r}V_r + \hat{\theta}V_\theta, \text{ so} \\ V_r &= V_x \cos \theta + V_y \sin \theta \\ V_\theta &= -V_x \sin \theta + V_y \cos \theta \end{aligned}$$

Thus,

$$\begin{pmatrix} V_r \\ V_\theta \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} V_x \\ V_y \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} V_r \\ V_\theta \end{pmatrix} = \mathbf{A} \begin{pmatrix} V_x \\ V_y \end{pmatrix}$$

and both the unit vectors and the components of a vector transform according to \mathbf{A} . However this is not true in general. If

$$\hat{e}' = \mathbf{M} \hat{e}, \quad \text{or} \quad \hat{e}'_i = \sum_{j=1}^2 m_{ij} \hat{e}_j,$$

²³Note that the matrix \mathbf{A} is the matrix of the *direction cosines*, i.e., writing

$$\begin{aligned} \hat{r} &= \hat{e}_x(\hat{e}_x \bullet \hat{r}) + \hat{e}_y(\hat{e}_y \bullet \hat{r}) \\ \hat{\theta} &= \hat{e}_x(\hat{e}_x \bullet \hat{\theta}) + \hat{e}_y(\hat{e}_y \bullet \hat{\theta}) \end{aligned}$$

$$\mathbf{A} = \begin{pmatrix} \hat{e}_x \bullet \hat{r} & \hat{e}_y \bullet \hat{r} \\ \hat{e}_x \bullet \hat{\theta} & \hat{e}_y \bullet \hat{\theta} \end{pmatrix},$$

where every entry is the cosine of the angle between the two unit vectors.

it is easy to show, given $\vec{V} = \sum_{i=1}^2 V_i \hat{e}_i = \sum_{i=1}^2 V'_i \hat{e}'_i$, that

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \widetilde{\mathbf{M}} \begin{pmatrix} V'_1 \\ V'_2 \end{pmatrix}, \text{ or } \mathbf{V} = \widetilde{\mathbf{M}} \mathbf{V}', \text{ and we have } \mathbf{V}' = (\widetilde{\mathbf{M}})^{-1} \mathbf{V}.$$

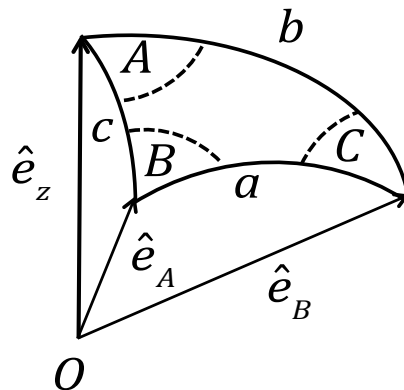
In the case above with \mathbf{A} , recall that $\mathbf{A} = \widetilde{\mathbf{A}} = \mathbf{A}^{-1}$, so $(\widetilde{\mathbf{A}})^{-1} = \mathbf{A}$.

In the text, we used matrices in the study of polarization of electromagnetic waves (Chapter 1) and in the discrete ordinate and matrix operator methods of solving the radiative transfer equation. In the application to polarization, a 4×1 column vector, the Stokes vector \mathbf{I} , described the state of polarization of radiance, and 4×4 matrices describe the action of polarizers, scattering, and changes of coordinate axes, on the Stokes vector — Mueller matrices \mathbf{M} . The action of several elements is described by the product of their individual element Mueller matrices, e.g., $\mathbf{I}' = \mathbf{M}_2 \mathbf{M}_1 \mathbf{I}$, where the light is subjected to element 1 first, then element 2. The noncommutative nature of the matrices representing polarizing elements is easily demonstrated. If natural light passes through a linear polarizer and then a quarter-wave plate, it emerges as *elliptically* polarized light, while if it first passes through a quarter-wave plate and then linear polarizer, it emerges as *linearly* polarized light.

14.5 Selected Topics

14.5.1 Spherical Triangles: The Laws of Sines and Cosines

Spherical triangles are formed by the intersection of three arcs of great circles on a sphere. In this section we derive the relationships between the sides and the angles of a spherical triangle. In the limit that the radius of the sphere becomes very large, the spherical triangle approaches a plane triangle, for which the relationships are equally valid. Let three unit vectors \hat{e}_z , \hat{e}_A and \hat{e}_B have their tails on the origin O and ends on the surface of a sphere of unit radius as shown in the figure below. The plane formed by any two of these vectors intersects the sphere along a great circle. The figure on the sphere formed by the intersection of the planes containing pairs of the three vectors is a spherical triangle. The arcs of the triangle (the sides) have angles a , b , and c given by $a = \hat{e}_A \bullet \hat{e}_B$, $b = \hat{e}_z \bullet \hat{e}_B$, $c = \hat{e}_z \bullet \hat{e}_A$. The angles formed by the intersection of the arcs on the sphere are A , B , and C , the *angles* of the spherical triangle. By convention, the angles of a spherical triangle are all $\leq 180^\circ$. Note that the arc opposing the angle A is a , etc. Assume for the moment



that \hat{e}_z is along the z axis of a spherical coordinate system. Then

$$\begin{aligned}
 \hat{e}_A &= \hat{e}_x \sin c \cos \phi_A + \hat{e}_y \sin c \sin \phi_A + \hat{e}_z \cos c \\
 \hat{e}_B &= \hat{e}_x \sin b \cos \phi_B + \hat{e}_y \sin b \sin \phi_B + \hat{e}_z \cos b \\
 \hat{e}_A \bullet \hat{e}_B &= \sin c \cos \phi_A \sin b \cos \phi_B + \sin c \sin \phi_A \sin b \sin \phi + \cos c \cos b \\
 &= \sin c \sin b \cos(\phi_B - \phi_A) + \cos c \cos b \\
 &= \sin c \sin b \cos A + \cos c \cos b,
 \end{aligned}$$

where the last step derives from the fact that in this particular coordinate system $\phi_B - \phi_A = A$. This is the law of cosines for spherical triangles:

$$\cos a = \cos b \cos c + \sin b \sin c \cos A.$$

But, since the spherical triangle is independent of the coordinate system used to describe the three vectors, we actually have three relationships referred to as the *law of cosines*:

$$\begin{aligned}
 \cos a &= \cos b \cos c + \sin b \sin c \cos A, \\
 \cos b &= \cos c \cos a + \sin c \sin a \cos B, \\
 \cos c &= \cos a \cos b + \sin a \sin b \cos C.
 \end{aligned} \tag{14.17}$$

We can establish another important set of relationships by converting $\cos A$ to $\sin A$, etc., as follows. Manipulate the top equation in Eq. (14.17)

$$\begin{aligned}
 \cos A &= \frac{\cos a - \cos b \cos c}{\sin b \sin c} \\
 \cos^2 A &= \frac{(\cos a - \cos b \cos c)^2}{\sin^2 b \sin^2 c} \\
 \sin^2 A &= 1 - \frac{(\cos a - \cos b \cos c)^2}{\sin^2 b \sin^2 c}.
 \end{aligned}$$

From the second equation in Eq. (14.17)

$$\sin^2 B = 1 - \frac{(\cos b - \cos c \cos a)^2}{\sin^2 c \sin^2 a}$$

Forming $\sin^2 A / \sin^2 B$, it is then a matter of algebra to show that

$$\frac{\sin^2 A}{\sin^2 B} = \frac{\sin^2 a}{\sin^2 b} \quad \text{or} \quad \frac{\sin A}{\sin a} = \frac{\sin B}{\sin b},$$

where the positive square root is taken because all the angles are $\leq \pi$. There are actually a total of three such relationships that can be formed from Eq. (14.17). They are

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}.$$

These constitute the *law of sines* for spherical triangles.

14.5.2 Least-Squares Analysis

Least-squares analysis was used in Chapter 6 to fit the results of detailed radiative transfer computations to simple relationships suggested by the QSSA. It has also been used extensively to develop empirical relationships between radiometric quantities and the concentration of Chlorophyll *a* or the pigment concentration. Here we describe and develop the mathematics of least-squares analysis.

Least-Squares is a method of fitting measurements to a given functional form. When the functional form is linear, the method is also called *linear regression*. For example, suppose that we are given a set of N measurements of a variable y , i.e., y_i , $i = 1 \rightarrow N$, corresponding respectively to *known* values of the variable x , x_i . We assume here that the x_i values are known exactly (but this is not always the case). If we believe there is a linear relationship between x and y , e.g., $y = a + bx$, how do we find the “best” values of a and b ? Of course we cannot answer this until we have a definition of “best.” In the method of least-squares, best means that we minimize the average of $(y_i - y_i^{(e)})^2$, where $y_i^{(e)}$ is the “expected” value of y given $x = x_i$, over the data set, i.e., minimize

$$S_2 \triangleq \frac{1}{N} \sum_{i=1}^N (y_i - y_i^{(e)})^2 = \sum_{i=1}^N [y_i - (a + bx_i)]^2$$

How do we do this? We note S_2 is a function of a and b , i.e., $S_2 = S_2(a, b)$, so we minimize S in the way we normally minimize a function:²⁴ we require

$$\frac{\partial S_2}{\partial a} = 0 \quad \text{and} \quad \frac{\partial S_2}{\partial b} = 0.$$

²⁴The subscript “2” on S is to remind us that we are minimizing the *square* of $y_i - (a + bx_i)$, rather than some other measure of the deviation.

Carrying out the indicated differentiation results in

$$\begin{aligned} \sum_{i=1}^N [y_i - (a + bx_i)] &= 0, \text{ or } Na + b \sum_{i=1}^N x_i = \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i [y_i - (a + bx_i)] &= 0, \text{ or } a \sum_{i=1}^N x_i + b \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i \end{aligned}$$

Let

$$\begin{aligned} \sum_{i=1}^N x_i &= N\langle x \rangle, & \sum_{i=1}^N y_i &= N\langle y \rangle, \\ \sum_{i=1}^N x_i^2 &= N\langle x^2 \rangle, & \sum_{i=1}^N x_i y_i &= N\langle xy \rangle. \end{aligned}$$

Then, the solution for a and b is

$$a = \frac{\langle x^2 \rangle \langle y \rangle - \langle x \rangle \langle xy \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \quad \text{and} \quad b = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}.$$

These equations also apply directly to several other kinds of fitting functions. For example, suppose we believe that x and y should be related by $y = Ax^B$. Then taking logarithms, $\log(y) = \log(A) + B \log(x)$. So if we carry out the linear analysis on the pairs $(\log(y_i), \log(x_i))$, we have $\log(A) = a$ and $B = b$. Similarly, if we believe $y = A \exp(Bx)$, then $\ln(y) = \ln(A) + Bx$, and using the pairing $(\ln(y_i), x_i)$, yields $\ln(A) = a$, and $B = b$. Generally, $(a + bx_i)$ can be replaced by any function $y = f(y_i, a, b, \dots)$, e.g., $y = a + bx + cx^2$, etc., and the procedure can be carried out as above; however, solving for a, b, \dots , may be challenging. When the linear function $y_i = (a + bx_i)$ is assumed, the resulting linear relationship is often referred to as the *regression line*. Here, we refer to this version of the least squares method as *ordinary least squares* (OLS).

What if we assumed that the y_i 's were exact, but not the x_i 's? Letting $x_i^{(e)}$ be the expected value of x_i given y_i , using the same model $y = a' + b'x$, we could minimize $\sum_i (x_i - x_i^{(e)})^2$ or

$$S'_2 \triangleq \frac{1}{N} \sum_{i=1}^N (x_i - x_i^{(e)})^2 = \sum_{i=1}^N \left[x_i - \left(\frac{y_i - a'}{b'} \right) \right]^2.$$

Carrying out the minimization of S'_2 as we did for S_2 above yields

$$a' = \langle y \rangle - b' \langle x \rangle \quad \text{and} \quad b' = \frac{\langle y^2 \rangle - \langle y \rangle^2}{\langle xy \rangle - \langle x \rangle \langle y \rangle}.$$

We call this version of the least squares method, *backward* least squares (BLS).

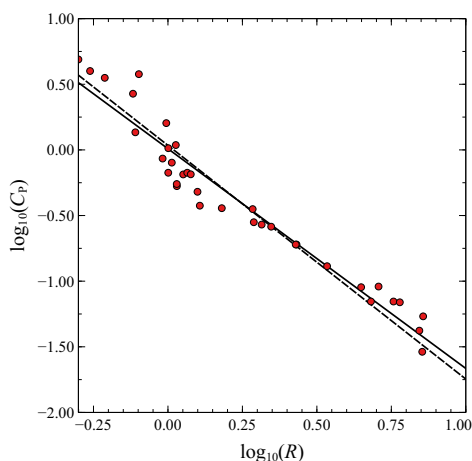
We can also consider cases where both x_i and y_i are in error, which is typical of empirical ocean color data, e.g., the pigment concentration as a function of radiance ratios (see Figure 6.64 and 6.65). One of the more popular regression method is called *reduced major axis* regression (RMA). In this case one minimizes

$$S_{\text{RMA}} \triangleq \frac{1}{N} \sum_{i=1}^N (x_i - x_i^{(e)})(y_i - y_i^{(e)}), \text{ with the constraint } y_i^{(e)} = a_{\text{RMA}} + b_{\text{RMA}}x_i^{(e)}.$$

Although tricky to carry out,²⁵ the result is

$$a_{\text{RMA}} = \langle y \rangle - b_{\text{RMA}} \langle x \rangle \quad \text{and} \quad b_{\text{RMA}} = \pm \sqrt{bb'},$$

where the sign of the square root is the same as the sign of $\langle xy \rangle - \langle x \rangle \langle y \rangle$, b is the result of the OLS and b' , the BLS. An example of the difference between various least squares approaches is shown below, where we fit the CZCS blue-green pigment algorithm in Figure



6.64 with both the OLS (solid line) and the BLS (dashed line) methods. Clearly, *large* differences in the resulting fit are *not* to be expected among the various approaches.

In the analysis of experimental data using the ordinary least squares, the statistical theory that is underlying the method assumes that the x_i values are exact, and the y_i values have errors that are normally distributed with a standard deviation σ , i.e., $N(\langle y \rangle, \sigma)$, with σ assumed here to be the same for all values of y . However, often some values of y_i are more accurately measured than others. If we know the standard deviation σ_i for each

²⁵One must use the method of *Lagrange multipliers* to effect the minimization of S_{RMA} .

of the individual data points, then we can take into account the accuracy variation by minimizing

$$S_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - (a + bx_i)}{\sigma_i} \right)^2.$$

In this case, points with smaller standard deviations are weighted more heavily than those with larger σ 's.

Finally, the case where both the x_i 's and the y_i 's have normally-distributed errors characterized by σ_x and σ_y , respectively (assumed to be independent of i), has been worked out and is called Deming regression. The resulting a_D and b_D in $y = a_D + b_D x$ are then functions of both σ_x and σ_y . Letting $\delta = \sigma_y/\sigma_x$, and

$$\begin{aligned} S_{xx} &\triangleq \langle x^2 \rangle - \langle x \rangle^2, \\ S_{yy} &\triangleq \langle y^2 \rangle - \langle y \rangle^2, \\ S_{xy} &\triangleq \langle xy \rangle - \langle x \rangle \langle y \rangle, \end{aligned}$$

it is found that

$$a_D = \langle y \rangle - b_D \langle x \rangle$$

where

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{(S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2}}{2S_{xy}}.$$

One can easily verify that for large δ (large error in the y_i 's compared to the x_i 's), $b_D \rightarrow b$, while for small δ (small error in the y_i 's compared to the x_i 's), $b_D \rightarrow b'$.

Anyone with experience with linear regression will recall cases in which a few data points (or even one point) that appear to be outliers (i.e., do not seem to follow the “trend” of the data) will strongly influence the regression slope (the parameter b). In the ordinary least squares the assumption that the errors in y_i are normally distributed implies that deviations of several σ 's, from their likely values are *highly improbable*, so they are treated by the method as being correct (i.e., valid). This causes data outliers to significantly affect the resulting a and b . In contrast, choosing to minimize

$$S_{\text{Abs}} \triangleq \frac{1}{N} \sum_{i=1}^N |y_i - (a + bx_i)|,$$

implicitly assumes that the error in the y_i 's is *exponentially* distributed, which implies larger deviations are more likely than for normally-distributed errors with the same σ . Minimizing S_{Abs} provides values of a and b that are more tolerant to outliers. Clearly, in data analysis it would be valuable to have some understanding of the distribution of the error.

14.5.3 Eigenvalues and Eigenfunctions of a Square Matrix

The representation we choose for writing a vector, e.g., in two dimensions we write an arbitrary vector \vec{V} as $\vec{V} = V_1\hat{e}_1 + V_2\hat{e}_2$, is called a *basis* and the unit vectors \hat{e}_1 and \hat{e}_2 are called the *basis* vectors. In matrix form the basis vectors are written

$$\hat{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \hat{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ and the vector } \mathbf{V} = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}.$$

Usually, one chooses basis vectors so that $\hat{e}_1 \bullet \hat{e}_2 = 0$, i.e., an *orthogonal* basis, but this is not necessary, it is only a convenience.

Consider a matrix \mathbf{M} defined to operate on vectors. (This matrix will typically represent some physical process, for example \vec{V} might be a displacement, and $\mathbf{M}\vec{V}$ a force.) To understand the action of the matrix on a given vector, it is useful to first examine its action on the basis vectors. Writing

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix},$$

It is clear that

$$\mathbf{M}\hat{e}_1 = m_{11}\hat{e}_1 + m_{21}\hat{e}_2, \text{ and } \mathbf{M}\hat{e}_2 = m_{12}\hat{e}_1 + m_{22}\hat{e}_2. \quad (14.18)$$

so

$$\begin{aligned} \mathbf{M}\vec{V} &= \mathbf{M}(V_1\hat{e}_1 + V_2\hat{e}_2) \\ &= V_1(m_{11}\hat{e}_1 + m_{21}\hat{e}_2) + V_2(m_{12}\hat{e}_1 + m_{22}\hat{e}_2) \\ &= (m_{11}V_1 + m_{12}V_2)\hat{e}_1 + (m_{21}V_1 + m_{22}V_2)\hat{e}_2, \text{ or} \\ \mathbf{M}\vec{V} &= \begin{pmatrix} m_{11}V_1 + m_{12}V_2 \\ m_{21}V_1 + m_{22}V_2 \end{pmatrix}. \end{aligned} \quad (14.19)$$

Of course the last step could have been carried out by direct multiplication of the matrix \mathbf{M} and the \vec{V} using its matrix representation \mathbf{V} ; however, we prefer to use Eq. (14.18) to explicitly demonstrate that knowing the action of \mathbf{M} on the basis vectors, provides its action on any vector in the space.

We now ask an important question: are there special directions (or vectors) such that when the matrix operates on a vector pointing in one of these special directions, the resulting vector is the same direction, with only its magnitude changed? Or restated mathematically, can we find a unit vector in our space of two-dimensional vectors, \hat{e}_a , such that

$$\mathbf{M}\hat{e}_a = \lambda_a\hat{e}_a$$

where λ_a is a constant? Note that simplicity of this relationship compared to that in Eqs. (14.18). Such vectors are easy to find. First expand the unknown vector using the basis vectors, i.e., $\hat{e}_a = \alpha\hat{e}_1 + \beta\hat{e}_2$, where $\alpha^2 + \beta^2 = 1$ because \hat{e}_a is a unit vector. Then, find the action of \mathbf{M} on \hat{e}_a :

$$\begin{aligned}\mathbf{M}\hat{e}_a &= \mathbf{M}(\alpha\hat{e}_1 + \beta\hat{e}_2) \\ &= \alpha(m_{11}\hat{e}_1 + m_{21}\hat{e}_2) + \beta(m_{12}\hat{e}_1 + m_{22}\hat{e}_2) \\ &= \lambda_a\hat{e}_a = \lambda_a(\alpha\hat{e}_1 + \beta\hat{e}_2),\end{aligned}$$

so,

$$\alpha[(m_{11} - \lambda_a)\hat{e}_1 + m_{21}\hat{e}_2] + \beta[m_{12}\hat{e}_1 + (m_{22} - \lambda_a)\hat{e}_2] = 0.$$

Taking the scalar product of this last equation with \hat{e}_1 and \hat{e}_2 , yields the set of simultaneous equations,

$$\begin{aligned}(m_{11} - \lambda_a)\alpha + m_{12}\beta &= 0 \\ m_{21}\alpha + (m_{22} - \lambda_a)\beta &= 0\end{aligned}\tag{14.20}$$

This set of equations has a solution, other than $\alpha = \beta = 0$, only if the determinate of the coefficients of α and β vanishes,²⁶ i.e.,

$$\begin{vmatrix} (m_{11} - \lambda_a) & m_{12} \\ m_{21} & (m_{22} - \lambda_a) \end{vmatrix} = 0.\tag{14.21}$$

This determinant is a quadratic equation, which (usually) yields two distinct values of λ_a . We will continue to call one of these λ_a and will call the other λ_b . The two values λ_a and λ_b are called the *eigenvalues* of \mathbf{M} . Given the eigenvalues, we can find α and β (there are two sets, one for \hat{e}_a corresponding to λ_a , and another \hat{e}_b corresponding to λ_b), by inserting the λ 's into Eq. (14.20).²⁷ For example, for λ_a we have

$$\beta = -\frac{(m_{11} - \lambda_a)}{m_{12}}\alpha,$$

with a similar equation with λ_b . This relationship is combined with $\alpha^2 + \beta^2 = 1$ to provide α and β . The resulting vectors \hat{e}_a and \hat{e}_b are called the *eigenvectors* of \mathbf{M} . If

²⁶Recall from algebra that the solution to the set of equations

$$\begin{aligned}a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2\end{aligned}\quad \text{is} \quad x = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \quad \text{and} \quad y = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}.$$

If $b_1 = b_2 = 0$, yielding zero for the numerators in the solution, in order that there be a solution, other than $x = y = 0$, the denominators must also vanish, i.e., the determinate of the coefficients of x and y must vanish.

²⁷Of course, only the ratio of the components is needed to establish the direction of \hat{e}_a .

the matrix \mathbf{M} is real and symmetric, the eigenvalues and eigenvectors have some special properties: (1) the eigenvalues and eigenvectors are real; and (2) the eigenvectors are mutually perpendicular.²⁸

Once we have derived the eigenvectors \hat{e}_a and \hat{e}_b , we can write vectors in terms of these rather than \hat{e}_1 and \hat{e}_2 , i.e., $\vec{V} = V_a\hat{e}_a + V_b\hat{e}_b$, then the action of \mathbf{M} on \vec{V} is

$$\mathbf{M}\vec{V} = \mathbf{M}(V_a\hat{e}_a + V_b\hat{e}_b) = V_a\lambda_a\hat{e}_a + V_b\lambda_b\hat{e}_b. \quad (14.22)$$

Clearly, Eq. (14.22) is much simpler than Eq. (14.19). In fact, if we represent the new basis vectors as column vectors in the usual manner, i.e.,

$$\hat{e}_a = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \hat{e}_b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{so in this system } \vec{V} = \begin{pmatrix} V_a \\ V_b \end{pmatrix},$$

then

$$\mathbf{M}\vec{V} = \begin{pmatrix} \lambda_a V_a \\ \lambda_b V_b \end{pmatrix} = \boldsymbol{\lambda} \begin{pmatrix} V_a \\ V_b \end{pmatrix}, \quad \text{where } \boldsymbol{\lambda} = \begin{pmatrix} \lambda_a & 0 \\ 0 & \lambda_b \end{pmatrix} \quad (14.23)$$

Thus, when the new basis vectors are chosen to be the eigenfunctions of \mathbf{M} , the action of \mathbf{M} on a vector \vec{V} , expanded in terms of the eigenfunctions, is the same as the action of a *diagonal* matrix $\boldsymbol{\lambda}$ which has the eigenvalues along the diagonal.²⁹

²⁸These properties are easy to demonstrate.

$$\begin{aligned} \mathbf{M}\hat{e}_a = \lambda_a\hat{e}_a &\implies \widetilde{\hat{e}_b^*}\mathbf{M}\hat{e}_a = \lambda_a\widetilde{\hat{e}_b^*}\hat{e}_a \\ \mathbf{M}\hat{e}_b = \lambda_b\hat{e}_b &\implies \widetilde{\hat{e}_a^*}\mathbf{M}\hat{e}_b = \lambda_b\widetilde{\hat{e}_a^*}\hat{e}_b \implies \widetilde{\hat{e}_a^*}\mathbf{M}^*\hat{e}_b^* = \lambda_b^*\widetilde{\hat{e}_a^*}\hat{e}_b^* \implies \widetilde{\hat{e}_b^*}\mathbf{M}^*\hat{e}_a = \lambda_b^*\widetilde{\hat{e}_b^*}\hat{e}_a \end{aligned}$$

Subtract the last of the upper equations from the last of the lower, and the result is

$$\widetilde{\hat{e}_b^*}\mathbf{M}^*\hat{e}_a - \widetilde{\hat{e}_b^*}\mathbf{M}\hat{e}_a = (\lambda_b^* - \lambda_a)\widetilde{\hat{e}_b^*}\hat{e}_a.$$

If \mathbf{M} is real and symmetric, $\widetilde{\mathbf{M}^*} = \mathbf{M}$, and the left hand side of the above equation vanishes. If we choose \hat{e}_b to be \hat{e}_a , then the last equation reads $(\lambda_a^* - \lambda_a)\widetilde{\hat{e}_a^*}\hat{e}_a = 0$, but $\widetilde{\hat{e}_a^*}\hat{e}_a$ cannot vanish (its equal to 1), so $(\lambda_a^* = \lambda_a)$, i.e., λ_a is real. But, given the way \hat{e}_a and \hat{e}_b are determined, if λ_a is real, then \hat{e}_a must be also be real, i.e., $\hat{e}_a = \hat{e}_a^*$, etc. But in this case, as long as $(\lambda_b \neq \lambda_a)$, $\widetilde{\hat{e}_b^*}\hat{e}_a = 0$, showing that $\hat{e}_b \bullet \hat{e}_a = 0$, and \hat{e}_b is perpendicular to \hat{e}_a . Aside: the reader can easily verify by direct calculation with Eq. (14.21), that if $m_{12} = m_{21}$, and the m_{ij} 's are real, then λ_a is real.

²⁹This can be shown directly. Consider a matrix \mathbf{T} , in which the columns are the eigenvectors, i.e.,

$$\mathbf{T} = \begin{pmatrix} \alpha_a & \alpha_b \\ \beta_a & \beta_b \end{pmatrix}, \quad \text{which has the property, } \widetilde{\mathbf{T}}\mathbf{T} = \mathbf{T}\widetilde{\mathbf{T}} = \mathbf{1}.$$

Direct calculation shows that

$$\widetilde{\mathbf{T}}\mathbf{M}\mathbf{T} = \boldsymbol{\lambda} \quad \text{where } \boldsymbol{\lambda} = \begin{pmatrix} \lambda_a & 0 \\ 0 & \lambda_b \end{pmatrix},$$

and so

$$\widetilde{\mathbf{T}}\vec{V} = \begin{pmatrix} \alpha_a & \alpha_b \\ \beta_a & \beta_b \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} V_1\alpha_a + V_2\beta_a \\ V_1\alpha_b + V_2\beta_b \end{pmatrix} = \begin{pmatrix} \hat{e}_a \bullet \vec{V} \\ \hat{e}_b \bullet \vec{V} \end{pmatrix} = \begin{pmatrix} V_a \\ V_b \end{pmatrix}.$$

The procedure we have carried out here can be extended to any number of dimensions. In higher dimensions, the number of eigenvalues will equal the dimensionality of the space, i.e., in the generalization to N dimensions, Eq. (14.21) will be an $N \times N$ determinant that when expanded will be an N^{th} order polynomial in λ . The algebra becomes more tedious as N increases, but the procedure is the same as outlined here.

Although little used in the text, nevertheless this procedure was implicit in our discussion of the *polarizability tensor* in Section 1.7.5. There, we wrote the molecular polarizability of an anisotropic molecule, e.g., a diatomic molecule, as

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & \alpha_z \end{pmatrix} = \begin{pmatrix} \alpha^\perp & 0 & 0 \\ 0 & \alpha^\perp & 0 \\ 0 & 0 & \alpha^\parallel \end{pmatrix},$$

whereas in general we would have

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{pmatrix}.$$

Thus, we *assumed* that we had already performed the procedure to find a set of coordinate axes for which $\boldsymbol{\alpha}$ was rendered diagonal. (We didn't actually carry out the procedure, but knowing that such a set of axes had to exist, we surmised that they had to be symmetry axes of the molecule.)

14.5.4 Random Variables and the Central Limit Theorem

In order to understand the development of Monte Carlo methods for solving the radiative transfer equation we need some basics concerning the description of random variables. Random variables are basically variables whose values cannot be determined exactly from one sample to the next. For example, if one places a macroscopic ball in a vat of (still) water, and takes a sequence of photographs of it, its position will remain fixed in time, i.e., it will not move and a prediction of its position is highly accurate. However, if the ball is microscopic and the same experiment is preformed, the position of the ball will be different in each photograph (due to Brownian motion) — its position will appear to follow an erratic path. In the latter case, the position coordinates (x, y) are both random variables, i.e., they cannot be accurately predicted.

Written this way, $(\tilde{\mathbf{T}}\mathbf{M}\mathbf{T})(\tilde{\mathbf{T}}\vec{V}) = \boldsymbol{\lambda}\tilde{\mathbf{T}}\vec{V}$, and the transformation $\tilde{\mathbf{T}}\mathbf{M}\mathbf{T}$ turns \mathbf{M} into the diagonal matrix $\boldsymbol{\lambda}$. General vectors \vec{V} before the transformation are transformed into $\tilde{\mathbf{T}}\vec{V}$.

14.5.4.1 Random Variables

We begin our discussion of random variables by looking at an example we have seen before in the text, and for which we know the answer. Consider a non-scattering medium, i.e., one that only has absorption. Let photons enter the medium. Individual photons travel a distance X before being absorbed, with X , of course, being different for each photon. We want to describe the distribution of distances traveled in the medium before absorption. How do we do it? We perform the experiment and simply tally the values of X for each of a large number (N) of photons, counting the number that are absorbed within various intervals of distance into the medium, i.e., we determine n_x , the number for which $X \leq x \leq X + \Delta X$, for the full range of X from $0 \rightarrow \infty$. Then, the *probability* that $X \leq x \leq X + \Delta X$ is defined by

$$\Delta \mathcal{P}_X(x) \triangleq \lim_{N \rightarrow \infty} \frac{n_x}{N},$$

and the *probability density* of X as

$$p_X(x) \triangleq \frac{\Delta \mathcal{P}_X(x)}{\Delta X}.$$

As we let the intervals become small, i.e., $\Delta X \rightarrow 0$,

$$p_X(x) \rightarrow \frac{d\mathcal{P}_X(x)}{dX}.$$

This is the probability density of paths that terminate at x . Since all paths must terminate somewhere, a requirement for \mathcal{P} is

$$\int_0^\infty d\mathcal{P}_X(x) = \int_0^\infty p_X(x) dx = 1,$$

where a probability of unity represents absolute certainty. If the above integral is satisfied, $p_X(x)$ is said to be normalized, and normalization is a requirement for all probability densities. In probability theory one defines the expected value or *expectation* of x to be

$$\mathcal{E}[x] \triangleq \int_0^\infty x d\mathcal{P}_x(x) = \int_0^\infty x p_x(x) dx,$$

and the *variance* of X as

$$\mathcal{V}[x] \triangleq \mathcal{E}[(x - \mathcal{E}[x])^2].$$

The expectation of x is often designated by μ_x , so $\mathcal{V}[x] = \mathcal{E}[(x - \mu_x)^2]$

What distance on the average do photons travel in the medium before absorption? Given all of the x 's in our experiment, the average of the measurements is

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i.$$

Consider the expectation of the average, $\mathcal{E}[\langle x \rangle]$,

$$\mathcal{E}[\langle x \rangle] = \frac{1}{N} \sum_{i=1}^N \mathcal{E}[x_i] = \frac{1}{N} N \mathcal{E}[x] = \mathcal{E}[x],$$

because the expected values of any of the samples (labeled by i) are identical. When the expectation value of an estimate for a variable is equal to the actual expectation value of the variable, the estimate is said to be an *unbiased* estimate. Thus, $\langle x \rangle$ is an unbiased estimate of $\mathcal{E}[x]$ and, because of this, the expectation of x is often referred to as the average of x . Likewise, consider

$$S^2 \triangleq \frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2.$$

A straightforward but tedious calculation shows that

$$\mathcal{E}[S^2] = \left(1 - \frac{1}{N}\right) \mathcal{V}[x],$$

so S^2 is *not* an unbiased estimate of the variance of x , but

$$\frac{N}{N-1} S^2$$

is an unbiased estimate, i.e., its expectation is equal to the variance. The square root of the variance is called the *standard deviation* denoted by σ , i.e., $\sigma \triangleq \sqrt{\mathcal{V}[x]}$.

In general, we would have to perform an experiment to determine $p_X(x)$; however, in the example being used for discussion here, we can determine it from the definition of the absorption coefficient. From the definition of the absorption coefficient in Chapter 2, in a medium, if there are n photons present in a beam, the change in that number due to absorption dn in traversing a distance dx is $dn = -na \, dx$. Thus, if N photons are introduced at $x = 0$, the number remaining at any value of x is

$$\int_N^n \frac{dn}{n} = - \int_0^x a \, dx \quad \text{or} \quad n(x) = N \exp(-ax).$$

We want the probability of absorption between x and $x + dx$. If we introduce N photons at $x = 0$, this is just

$$d\mathcal{P}_X = \frac{\text{number absorbed in } dx}{N} = \frac{n(x) - n(x + dx)}{N} = a \exp(-ax) \, dx,$$

so $p_X(x) = a \exp(-ax)$.³⁰ Now, we can calculate some expectation values. First, let's calculate the expectation of x . This is just

$$\mathcal{E}[x] = \int_0^\infty x \exp(-ax) \, dx = \frac{1}{a},$$

³⁰Note that $\int_0^\infty p_X(x) \, dx = \int_0^\infty a \exp(-ax) \, dx = 1$, so the density function automatically satisfies the normalization requirement.

or loosely, the average distance traveled by a photon in the medium before absorption is just $1/a$. This is sometimes called the *mean free path for absorption*.³¹ What about the variance? We will need

$$\mathcal{E}[x^2] = \int_0^\infty x^2 \exp(-ax) dx = \frac{2}{a^2},$$

because

$$\mathcal{V}[x] = \mathcal{E}[(x - \mathcal{E}[x])^2] = \mathcal{E}[x^2 - 2x\mathcal{E}[x] + \mathcal{E}[x]^2].$$

Since $\mathcal{E}[x]$ is just a number, $\mathcal{E}[2x\mathcal{E}[x]] = 2\mathcal{E}[x]^2$, so

$$\mathcal{V}[x] = \mathcal{E}[x^2] - \mathcal{E}[x]^2 = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = \frac{1}{a^2}, \quad \text{and} \quad \sigma = \frac{1}{a}.$$

14.5.4.2 The Central Limit Theorem

Assume we are given random variables x_1, x_2, \dots, x_N distributed according to the probability densities $p_{x_1}(x_1), p_{x_2}(x_2), \dots, p_{x_N}(x_N)$. We wish to find the probability density $p_{x_1+x_2+\dots+x_N}$ for the sum $x_1 + x_2 + \dots + x_N$. In particular, we are most interested in the special case where all of the x_i 's have the same probability density, $p_x(x)$. To effect this we consider the expectation of $\exp(ikx)$, i.e.,

$$\mathcal{E}[\exp(ikx)] = \int_{-\infty}^{+\infty} p_x(x) \exp(ikx) dx \triangleq \Phi_x(k). \quad (14.24)$$

The function $\Phi_x(k)$ is called the *characteristic function*. One immediately recognizes that $\Phi_x(k)$ and $p_x(x)$ are *Fourier transform pairs*, i.e.,³²

$$p_x(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_x(k) \exp(-ikx) dk.$$

³¹Likewise, in a scattering medium the mean free path between scattering events is $1/b$. When both absorption and scattering are present, the mean free path between interactions (absorption or scattering) is $1/(a+b) = 1/c$.

³²In the time domain, Fourier transform pairs $h(t)$ and $\hat{h}(f)$ are defined through

$$\hat{h}(f) = \int_{-\infty}^{+\infty} h(t) \exp(2\pi i f t) dt \quad \text{and} \quad h(t) = \int_{-\infty}^{+\infty} \hat{h}(f) \exp(-2\pi i f t) df.$$

In the usual convention, where one uses angular frequency $\omega = 2\pi f$, these become

$$\hat{h}(\omega) = \int_{-\infty}^{+\infty} h(t) \exp(i\omega t) dt \quad \text{and} \quad h(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{h}(\omega) \exp(-\omega t) d\omega.$$

If the factors of 2π are left out of the arguments of the exponentials, then a factor of $1/(2\pi)$ must be introduced in one to the integrals of the pair. Alternatively, the factor $1/\sqrt{2\pi}$ could appear in both. All of these conventions are used making for a somewhat confused literature on Fourier transforms. Make sure you know the convention employed in any source used as reference.

Often it is easier to work with Φ compared to p . Consider the sum of two random variables

$$\begin{aligned}\Phi_{x_1+x_2}(k) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{x_1}(x_1)p_{x_2}(x_2) \exp[ik(x_1+x_2)] dx_1 dx_2 \\ &= \left(\int_{-\infty}^{+\infty} p_{x_1}(x_1) \exp(ikx_1) dx_1 \right) \left(\int_{-\infty}^{+\infty} p_{x_2}(x_2) \exp(ikx_2) dx_2 \right) \\ &= \Phi_{x_1}(k)\Phi_{x_2}(k),\end{aligned}$$

and if x_1 and x_2 have the same density $p_x(x)$, then $\Phi_{x_1+x_2}(k) = [\Phi_x(k)]^2$.

Another feature of the characteristic function is that it generates *moments* of the distribution. Expand the exponential in Eq. (14.24), noting that k is a constant, and we get

$$\begin{aligned}\mathcal{E}[\exp(ikx)] &= 1 + ik\mathcal{E}[x] - \frac{k^2}{2}\mathcal{E}[x^2] - i\frac{k^3}{6}\mathcal{E}[x^3] \cdots \\ &= 1 + ik\langle x \rangle - \frac{k^2}{2}\langle x^2 \rangle - i\frac{k^3}{6}\langle x^3 \rangle \cdots\end{aligned}$$

Thus,

$$\begin{aligned}\Phi_x(0) &= 1 \\ \Phi'_x(0) &= i\langle x \rangle \\ \Phi''_x(0) &= -\langle x^2 \rangle \\ \Phi'''_x(0) &= -i\langle x^3 \rangle \\ &\vdots\end{aligned}\tag{14.25}$$

where the prime indicates differentiation with respect to k . So repeated differentiation of the characteristic function n times with respect to k generates the moment $\langle x^n \rangle$ of the probability density.

Now we are ready to answer the question posed at the beginning of this section: We wish to find the probability density $p_{x_1+x_2+\cdots+x_N}$ for the sum $S = x_1 + x_2 + \cdots + x_N$, where all of the x_i 's have the same probability density, $p_x(x)$. But, rather than this let's find the density for the sum

$$\begin{aligned}y &= \frac{x_1 + x_2 + x_3 + \cdots + x_N - N\langle x \rangle}{\sqrt{N\sigma^2}} = \frac{S - N\langle x \rangle}{\sqrt{N\sigma^2}} \\ &= \sum_{i=1}^N \frac{x_i - \langle x_i \rangle}{\sqrt{N\sigma^2}} \\ &= \sum_{i=1}^N \frac{y_i}{\sqrt{N}}, \quad \text{where} \quad y_i = \frac{x_i - \langle x_i \rangle}{\sigma} \quad \text{and} \quad \sigma^2 = \langle x_i^2 \rangle - \langle x_i \rangle^2.\end{aligned}$$

Note, the variable y_i has zero mean, and unit variance. With this substitution, $\Phi_y(k) = [\Phi_{y_i/\sqrt{N}}(k)]^N$, but using Eq. (14.25)

$$\begin{aligned}\Phi_y(k) &= [\Phi_{y_i/\sqrt{N}}(k)]^N = \left[1 + ik \frac{\langle y_i \rangle}{\sqrt{N}} - \frac{k^2}{2} \frac{\langle y_i^2 \rangle}{N} - \frac{k^3}{6} \frac{\langle x_i^3 \rangle}{N^{3/2}} \dots \right]^N \\ &= \left[1 - \frac{k^2}{2} \frac{\langle y_i^2 \rangle}{N} - \frac{k^3}{6} \frac{\langle x_i^3 \rangle}{N^{3/2}} \dots \right]^N.\end{aligned}$$

To find the limiting distribution for large values of N , we need to evaluate the last expression as $N \rightarrow \infty$, i.e., to find $\lim_{N \rightarrow \infty} [\Phi_{y_i/\sqrt{N}}(k)]^N$. This can be done most easily by taking the natural logarithm of the equation, i.e.,

$$\begin{aligned}\ell n[\Phi_{y_i/\sqrt{N}}(k)]^N &= N \ell n[\Phi_{y_i/\sqrt{N}}(k)] \\ &= N \ell n \left[1 - \frac{k^2}{2} \frac{\langle y_i^2 \rangle}{N} - \frac{k^3}{6} \frac{\langle x_i^3 \rangle}{N^{3/2}} \dots \right].\end{aligned}$$

Taking the direct limit of this yields the indeterminate $0 \times \infty$, so we rewrite it in a form to be able to use L'Hospital's rule for limits:

$$\ell n[\Phi_{y_i/\sqrt{N}}(k)]^N = \frac{1}{1/N} \ell n \left[1 - \frac{k^2}{2} \frac{\langle y_i^2 \rangle}{N} - \frac{k^3}{6} \frac{\langle x_i^3 \rangle}{N^{3/2}} \dots \right]$$

and the indeterminate is now $0/0$, and we can use the rule directly. Carrying this out we find

$$\lim_{N \rightarrow \infty} \ell n[\Phi_{y_i/\sqrt{N}}(k)]^N \rightarrow -\frac{k^2}{2} \quad \text{or} \quad \Phi_y(k) = [\Phi_{y_i/\sqrt{N}}(k)]^N \rightarrow \exp \left[-\frac{k^2}{2} \right]$$

Now

$$\begin{aligned}p_y(y) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_y(k) \exp(-ikx) dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \left(-\frac{k^2}{2} \right) \exp(-ikx) dk \\ &= \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{y^2}{2} \right)\end{aligned}$$

Noting that $p_S(S) dS = p_y(y) dy$,

$$p_S(S) = \frac{1}{\sqrt{2\pi N \sigma^2}} \exp \left(-\frac{(S - N\langle x \rangle)^2}{2N\sigma^2} \right).$$

Thus, S has a normal distribution³³ with mean $N\langle x \rangle$ and variance $N\sigma^2$. This is the *Central Limit Theorem*. Note that $\langle S \rangle = N\langle x \rangle \implies \langle x \rangle = \langle S \rangle / N$. Also, $\sigma_{S/N}^2 \triangleq \langle (S/N -$

³³Writing the normal probability density as $p_x(x) dx = (1/\sqrt{2\pi s^2}) \exp[-(x - \mu)^2/(2s^2)] dx$, one can show

$\langle S \rangle / N)^2 = (\langle S^2 \rangle - \langle S \rangle^2) / N^2 = N\sigma^2 / N^2 = \sigma^2 / N$, so the variance in $S/N = \langle x \rangle$ is σ^2 / N and the standard deviation is σ / \sqrt{N} .

This theorem has wide application in physics and mathematics. For example, if one measures a quantity x with an error that is distributed according to some distribution having a standard deviation σ , then the mean of a large number N of measurements of x is an unbiased estimator of x and the standard deviation of the mean is σ / \sqrt{N} . We used this fact in assessing the variance of integrals computed using the Monte Carlo method.

Another interesting application is the generation of random numbers with a normal distribution. Assume that we can generate random numbers x with a uniform distribution on the interval $[0, r]$, i.e., $p(x) = 1/r$. Then the mean x is $r/2$, and the variance in x is $\sigma^2 = \int_0^r x^2 dx - \langle x \rangle^2 = r^2/3 - r^2/4 = r^2/12$. According to the Central Limit Theorem, the sum $\sum_{i=1}^N x_i - N\langle x \rangle = \sum_{i=1}^N x_i - Nr/2 = \sum_{i=1}^N (x_i - r/2)$ approaches a normal distribution with mean zero and variance $N\sigma^2 = Nr^2/12$. Say we want the variance of the normal distribution σ_N^2 to be a given number. Then $\sigma_N^2 = Nr^2/12$, and this fixes r . How do we generate random numbers on the interval $0 \rightarrow r$? Just use a random number generator that is $U[0,1]$ providing uniform ρ 's on the interval 0 to 1 (there are plenty of these) and multiply each ρ by r . Do this N times, sum the result to give S , which will be $N[(r/2), \sigma_N^2]$. To get $N[0, \sigma_N^2]$, just subtract $r/2$ from each value of $r\rho$. To generate $N[0,1]$, we need $r = \sqrt{12/N}$. So one generates ρ_i from $U[0,1]$, takes

$$\rho_N = \sqrt{\left(\frac{12}{N}\right)} \sum_{i=1}^N \left(\rho_i - \frac{1}{2}\right),$$

and the resulting ρ_N 's will be $N[0,1]$. This is the method used to generate the noise models in Chapter 11.

14.5.5 Monte Carlo Evaluation of Integrals

It was shown in Chapter 2 that the solution to the radiative transfer equation can be written as a multi-dimensional integral. The integral was evaluated there using Monte Carlo methods and provided a basis for solution of the radiative transfer equation by Monte Carlo techniques. In this section we will show how one can perform an integration

by direct integration that

$$\langle x \rangle = \int_{-\infty}^{+\infty} x p_x(x) dx = \mu \quad \text{and} \quad \langle x^2 \rangle = \int_{-\infty}^{+\infty} x^2 p_x(x) dx = s^2 + \mu^2,$$

so $\sigma_x = \langle x^2 \rangle - \mu^2 = s^2$.

numerically using random variables (the Monte Carlo method), determine the variance of the Monte Carlo estimate of an integral, and discuss the possibility of reducing the variance.

14.5.5.1 Basic Ideas

Assume that we want to evaluate the integral $I = \int_0^1 f(x) dx$. Recall that the integral is defined as follows. Divide the x axis between 0 and 1 into N intervals and let $\Delta x = 1/N$, then

$$I \triangleq \lim_{N \rightarrow \infty} \sum_{i=1}^N f(\xi_i) \Delta x,$$

where ξ_i is a point on the x axis *anywhere* in the i^{th} interval, i.e., $x_{i-1} \leq \xi_i \leq x_i$.³⁴ Thus, a straightforward procedure for evaluating the integral numerically is

$$\int_0^1 f(x) dx = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(\xi_i),$$

i.e., it is just the value of f in each interval *averaged* over all of the intervals.³⁵ The choice of the position of ξ_i within the intervals is arbitrary, but certain choices yield familiar approximations. For example, if ξ_i is in the center of the interval, this is just the trapezoidal rule for integration.

How do we use random variables to evaluate the integral? It is simple. Let X be a random variable on the interval $0 \rightarrow 1$. Assume further that X is uniformly distributed on the interval, i.e., $p_X(x) = 1$. Then,

$$I = \int_0^1 f(x) dx = \int_0^1 p_X(x) f(x) dx = \mathcal{E}[f].$$

So, we get the value of the integral by computing the expectation of $f(x)$ when x is a uniformly distributed random variable. How do we do this in practice? There are many freely-available random number generators³⁶ for computers that provide a random sequence on the interval $0 \rightarrow 1$ with a uniform distribution (denoted $U[0,1]$). Using one of these, a

³⁴This is not the most general definition of the integral. More generally, the intervals (Δx 's) do not have to all be the same size, but then in the limiting procedure one must require that the largest interval goes to zero as $N \rightarrow \infty$.

³⁵Note, the "1" in the equation has the same units as x .

³⁶Really *pseudo* random number generators, as they always produce the same sequence of random numbers starting from a given random number.

sequence of N random values of X , i.e., x_1, x_2, \dots, x_N , are selected, and used to compute

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^N f(x_i),$$

which is an *unbiased* estimate of $\mathcal{E}[f]$ and thus of I .

Now, what is the variance in I ? Here we need to appeal to the Central Limit Theorem, which roughly states that if $g_i, i = 1, N$ are random variables all with the same distribution, then the variance of their sum, σ_{Sum}^2 , is the individual variance of the g 's, σ_g^2 , divided by N . Thus, the variance in the estimate of I above is

$$\mathcal{V}[I] = \frac{1}{N} \mathcal{V}[f] \quad \text{or} \quad \sigma_I^2 = \frac{1}{N} \sigma_f^2.$$

This shows that the “spread” of I values from multiple Monte Carlo evaluations of I ,³⁷ roughly $\sqrt{\mathcal{V}[I]}$, is proportional to $1/\sqrt{N}$. Loosely speaking, this is the error in any single evaluation of I . Thus, to get an order of magnitude improvement in the accuracy of the value of the integral, i.e., one decimal point, N must increase by a factor of 100.³⁸ An unbiased estimate of $\mathcal{V}[f]$ is simply $S_f^2/(N-1)$, i.e.,

$$\mathcal{V}[f] = \sigma_f^2 \approx \frac{N}{N-1} S_f^2 = \frac{1}{N-1} \sum_{i=1}^N (f(x_i) - \langle f \rangle)^2 = \frac{1}{N-1} \left[\left(\sum_{i=1}^N f(x_i)^2 \right) - \langle f \rangle^2 \right],$$

so an unbiased estimate of the variance of the Monte Carlo estimate of the integral is

$$\sigma_I^2 = \left(\frac{1}{N} \right) \left(\frac{1}{N-1} \right) \left[\left(\sum_{i=1}^N f(x_i)^2 \right) - \langle f \rangle^2 \right].$$

There are other approaches that can be employed in Monte Carlo evaluations. For example, what if we wanted to evaluate

$$I = \int_0^1 f(x)w(x) dx \quad ?$$

Following the above procedure, we would evaluate it as follows:

$$I = \int_0^1 f(x)w(x) dx = \int_0^1 f(x)w(x)p_X(x) dx = \mathcal{E}[fw] \approx \frac{1}{N} \sum_{i=1}^N f(x_i)w(x_i)$$

³⁷Of course, with different sets of random values of x .

³⁸For reference, the error in the trapezoidal rule for this integral is $f''(\zeta)/(12N^2)$, where ζ is some (unknown) x between 0 and 1, and N is the number of intervals. Noting that the number of function evaluations for Monte Carlo and trapezoidal integration is N (roughly the “cost” of doing the integral), the trapezoidal rule is clearly far more accurate for the same number of function evaluations.

where the N values of x (x_i , $i = 1, N$) are sampled from $p_X(x)$, the uniform distribution. But there is another way (actually many other ways). Assuming that the integral $I_w \triangleq \int_0^1 w(x) dx < \infty$ (it better be, or the desired integral will probably diverge as well), we could define a new probability density

$$p_w(x) \triangleq \frac{w(x)}{I_w},$$

and write the desired integral as

$$I = I_w \int_0^1 f(x) p_w(x) dx = I_w \mathcal{E}_{p_w}[f] \approx \frac{I_w}{N} \sum_{i=1}^N f(x_i),$$

where the N values of x (x_i , $i = 1, N$) are sampled from $p_w(x)$.³⁹ But, in practice, how does one sample from the distribution $p_w(x)$? It's easy. We already have a random number generator $U[0,1]$ generating numbers with a uniform density (call it $p_U(y)$). Start with the uniform distribution and demand that

$$p_w(x) dx = p_U(y) dy = dy,$$

or in words, we want the probability that x is between x and $x + dx$ to be the same as the probability that y is between y and $y + dy$. Integrate this from $y = 0 \rightarrow \rho_i$ and $x = 0 \rightarrow x_i$:

$$\rho_i = \int_0^{x_i} p_w(x) dx.$$

This relates x_i sampled from $p_w(x)$ to ρ_i sampled from $p_U(y)$. For our example, if $w(x) = x$, then, $I_w = 1/2$, $p_w(x) = 2x$ and

$$\rho_i = \int_0^{x_i} 2x dx = x_i^2 \quad \text{or} \quad x_i = \sqrt{\rho_i}.$$

So, selecting ρ_i , $i = 1 \rightarrow N$, from $U[0,1]$ and taking x_i to be $\sqrt{\rho_i}$ we have

$$I = \int_0^1 x f(x) dx \approx \frac{1}{2N} \sum_{i=1}^N f(x_i).$$

A second example relevant to radiative transfer is

$$I = \int_0^\infty f(x) \exp(-ax) dx,$$

³⁹Here we add the subscript “ p_w ” to \mathcal{E} to indicate that the expectation is with respect to the density p_w .

where $w(x) = \exp(-ax)$. Here, $I_w = 1/a$, $p_w(x) = a \exp(-ax)$, so

$$\rho_i = \int_0^{x_i} a \exp(-ax) dx = 1 - \exp(-ax_i) \quad \text{or} \quad x_i = -\frac{1}{a} \ln(1 - \rho_i),$$

and

$$I = \int_0^\infty f(x) \exp(-ax) dx \approx \frac{1}{aN} \sum_{i=1}^N f(x_i),$$

where ρ_i is sampled from $U[0,1]$.⁴⁰

14.5.5.2 Example of an Evaluation

As an example of the evaluation of an integral consider

$$I = \int_0^\infty x \exp(-x) dx.$$

Here, we take $f(x) = x$ and use $p(x) = \exp(-x)$. Then $I = \mathcal{E}_p[x]$ and $\sigma_f^2 = \mathcal{E}_p[x^2] - (\mathcal{E}_p[x])^2$. For this case we know the answers already by direct integration: $I = 1$ and $\sigma_f^2 = 1$. The random numbers ρ_i are chosen from $U[0,1]$ and the x_i -values are given by $x_i = -\ln(\rho_i)$. The integral is then approximated by

$$I = \mathcal{E}_p[x] \approx \sum_{i=1}^N x_i \quad \text{and} \quad \mathcal{E}_p[x^2] \approx \sum_{i=1}^N x_i^2.$$

We now perform the following numerical experiment. Choose a value for N and compute an estimate of I and σ_f^2 . Carry this out 100 times yielding 100 estimates for I . Compute the mean I (I_{Avg}), its variance (σ_I^2), the minimum I (I_{Min}) and the maximum (I_{Max}) values of I over the 100 evaluations. The following table provides the results of this exercise.⁴¹ One

	$N = 10^1$	$N = 10^2$	$N = 10^3$	$N = 10^4$
I_{Avg}	0.986	1.022	1.001	1.000
I_{Min}	0.375	0.777	0.899	0.976
I_{Max}	2.052	1.332	1.083	1.026
σ_f	0.848	0.998	0.996	1.000
σ_I	0.341	0.101	0.032	0.009

⁴⁰Notice that ρ and $1 - \rho$ have the *same* distribution, therefore we could let $x_i = -\frac{1}{a} \ln(\rho_i)$ as well.

⁴¹The value of σ_f^2 in the table is the average σ_f^2 over the 100 evaluations and provides an *unbiased* estimate of $\int_0^\infty x^2 \exp(-x) dx = \sigma_f^2 + I^2$.

can make several conclusions from this experiment: first, using a small N can result in large fluctuations in the result as indicated by I_{Min} and I_{Max} ; second, to a good approximation $\sigma_I = \sigma_f/\sqrt{N}$; and finally very accurate results can be obtained for sufficiently large N .

14.5.5.3 Reduction of Variance

Sometimes it is possible to reduce the variance by sampling from a different probability density. Consider again, $I = \int_0^1 f(x) dx = \int_0^1 f(x)p(x) dx = \mathcal{E}_p[f]$, where, as earlier, the “ p ” subscript on \mathcal{E} (and σ_I^2 below) denotes expectation when $p(x)$ is the density function (here uniform, but we still include it for clarity). The variance in the integral is

$$\sigma_{I_p}^2 = \frac{\mathcal{E}_p[f^2] - (\mathcal{E}_p[f])^2}{N}.$$

Now let’s change the sampling density from $p(x)$ to $\tilde{p}(x)$ as follows:

$$I = \int_0^1 f(x)p(x) dx = \int_0^1 \left[\frac{p(x)}{\tilde{p}(x)} \right] f(x)\tilde{p}(x) dx \triangleq \int_0^1 w(x)f(x)\tilde{p}(x) dx = \mathcal{E}_{\tilde{p}}[wf].$$

The variance in I in this case is

$$\sigma_{I_{\tilde{p}}}^2 = \frac{\mathcal{E}_{\tilde{p}}[w^2 f^2] - (\mathcal{E}_{\tilde{p}}[wf])^2}{N}.$$

Now,

$$\mathcal{E}_{\tilde{p}}[w^2 f^2] = \int_0^1 [w(x)f(x)]^2 \tilde{p}(x) dx = \int_0^1 \frac{1}{\tilde{p}(x)} [f(x)]^2 dx = \int_0^1 \frac{1}{\tilde{p}(x)} [f(x)]^2 p(x) dx,$$

but if $\tilde{p}(x)$ is large where $f(x)$ is large,

$$\int_0^1 \frac{1}{\tilde{p}(x)} [f(x)]^2 p(x) dx < \int_0^1 [f(x)]^2 p(x) dx = \mathcal{E}_p[f^2].$$

In this case,

$$\mathcal{E}_{\tilde{p}}[w^2 f^2] < \mathcal{E}_p[f^2] \quad \text{but} \quad \mathcal{E}_{\tilde{p}}[wf] = \mathcal{E}_p[f]$$

so $\sigma_{I_{\tilde{p}}}^2 < \sigma_{I_p}^2$, i.e., the variance in the estimate is reduced if the sampling density $p(x)$ is replaced by one ($\tilde{p}(x)$) that has a shape similar to $f(x)$.⁴² Thus, we can reduce the variance by choosing a \tilde{p} that is large where $f(x)$ is large, yielding many samples there, and small where $f(x)$ is small, providing sparse sampling in those regions. The strategy of using such a \tilde{p} is often referred to as *importance sampling*.

⁴²This would suggest that the best \tilde{p} would be f itself. Unfortunately, the probability density must be normalized, and if \tilde{p} were taken to be f itself, that would require knowing $\int_0^1 f(x) dx$, just the integral we are trying to evaluate.

14.5.5.4 Singular Integrands

Although one might not expect it, the Monte Carlo method can also be used even when the integrand is singular. In such cases, one can try to shift the singularity to the sampling function, e.g., consider

$$I = \int_0^1 \frac{f(x)}{\sqrt{x}} dx.$$

If $f(x)$ is finite on the interval $[0, 1]$, because the singularity is so mild, the integral is finite even though the integrand is infinite at $x = 0$. If we try to integrate using choosing x_i from $U[0,1]$, i.e.,

$$I \int_0^1 \frac{f(x)}{\sqrt{x}} p(x) dx \approx \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{\sqrt{x_i}},$$

then, the variance will be large as values of x_i too close to $x = 0$ make an abnormally large contribution to the sum. However, if we can find a probability density that is $\propto \sqrt{x}^{-1}$, we can use that rather than the uniform distribution. We note that

$$p_1(x) = \frac{1}{2\sqrt{x}}$$

is such a density.⁴³ Thus,

$$I = \int_0^1 \frac{f(x)}{\sqrt{x}} dx = \int_0^1 2f(x)p_1(x) dx \approx \frac{1}{N} \sum_{i=1}^N 2f(x_i),$$

where the x_i 's are chosen from $p_1(x)$. How do we sample from $p_1(x)$? Using the technique developed earlier, if ρ_i is sampled from $U[0,1]$, then the associated x_i , given by

$$\rho_i = \int_0^{x_i} p_1(x) dx \quad \text{or} \quad x_i = \rho_i^2,$$

will be distributed according to $p_1(x)$.

For an example of evaluating an integral with a singular integrand, consider the case where $f(x) = \exp(x)$. Then, sampling from $p_1(x)$ using the method described above,

$$I = \int_0^1 \frac{\exp(x)}{\sqrt{x}} dx \approx \frac{1}{N} \sum_{i=1}^N 2\exp(x_i).$$

For 1000 realizations, each with $N = 100$, we find $I \approx 2.9293$, compared to the exact result $I = 2.925304 \dots$.⁴⁴

⁴³It is imperative that the integral of the new density can be computed, as it is required to effect normalization.

⁴⁴This integral could also be evaluated using the direct Monte Carlo method after the substitution $x = y^2$, which removes the singularity.

14.5.6 Fourier Series

In Section 2.7.1.2 of Chapter 2, we expanded the azimuthal dependence of the scattering phase function $P(\hat{\xi} \rightarrow \hat{\xi}')$ in a Fourier series in $\phi - \phi'$. Here, we describe Fourier series and how to derive them.

14.5.6.1 Derivation of Fourier Series

The expansion of a function in a series of sines and/or cosines is called a Fourier series. Assume we have a function $f(x)$ defined on the interval $[-\pi, \pi]$. Can we represent the function in the following form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) ? \quad (14.26)$$

Let's assume we can. How do we find the constants a_n and b_n ? The procedure is quite simple. First, we note the following integrals:

$$\begin{aligned} \int_{-\pi}^{\pi} \sin(mx) \sin(nx) dx &= \pi \delta_{mn}, \\ \int_{-\pi}^{\pi} \cos(mx) \cos(nx) dx &= \pi \delta_{mn}, \\ \int_{-\pi}^{\pi} \cos(mx) \sin(nx) dx &= 0, \end{aligned} \quad (14.27)$$

where, m and n are integers, neither of which is zero. The function $\delta_{mn} = 0$, if $m \neq n$, and $\delta_{mn} = 1$, if $m = n$. It is called the *Kronecker delta*. Clearly, in the first integral, if $m = 0$ or $n = 0$, the integral is zero, and in the second integral, if $m = 0$ and $n = 0$, the integral is 2π . The integrals above are usually referred to as the *orthogonality* of the sines and cosines or the *orthogonality conditions* for the sines and cosines. Given these integrals, multiplying the series in Eq. (14.26) by $\sin(\ell x)$, where ℓ is an integer, integrating from $-\pi \rightarrow \pi$, and using the integrals above, we get

$$\begin{aligned} \int_{-\pi}^{\pi} f(x) \sin(\ell x) dx &= \frac{a_0}{2} \int_{-\pi}^{\pi} \sin(\ell x) dx \\ &+ \sum_{n=1}^{\infty} a_n \int_{-\pi}^{\pi} \cos(nx) \sin(\ell x) dx \\ &+ \sum_{n=1}^{\infty} b_n \int_{-\pi}^{\pi} \sin(nx) \sin(\ell x) dx \\ &= \pi b_{\ell}. \end{aligned}$$

In a similar manner,

$$\int_{-\pi}^{\pi} f(x) \cos(\ell x) dx = \pi a_{\ell},$$

$$\int_{-\pi}^{\pi} f(x) dx = \pi a_0.$$

Thus, the evaluation of the a_n 's and b_n 's is remarkably simple. The Fourier series will converge *point wise* if the function $f(x)$ is continuous on the interval; however, even if it is discontinuous at a finite number of points within the interval the series still converges as long as the discontinuities are finite. Where $f(x)$ is discontinuous, the series converges to the midpoint of the discontinuity.⁴⁵ Note that the function $f(x)$ is only defined for the interval $[-\pi, \pi]$; however, its Fourier series representation exists for all x and is periodic with period 2π . Furthermore, since $\cos(nx)$ is an *even* function of x and $\sin(nx)$ is an *odd* function of x , if $f(x)$ is an even function, $b_{\ell} = 0$, while if $f(x)$ is an odd function $a_{\ell} = 0$.⁴⁶ If the function $f(x)$ is only defined for part of the interval $[-\pi, \pi]$, e.g., only for $[0, \pi]$, one can still derive a Fourier series simply by defining f in the range for which it is undefined to have whatever (finite) values one wishes. For example, one could define $f(x)$ to be zero for $[-\pi, 0]$. The Fourier series would then converge to $f(x)$ for $0 \leq x \leq \pi$ and to zero for $-\pi \leq x \leq 0$, i.e., it would converge to the correct f for the range of x for which f was originally defined. However, it is important to note that the speed of convergence — the number of terms in the series required for a given accuracy — of the series to f in the defined range can depend significantly on what is chosen for f in the undefined part of the range.

What if we have a more general interval than $[-\pi, \pi]$, i.e., what if $f(x)$ is defined on the interval $[-L, L]$? Noting the strong relationship between Eqs. (14.26) and (14.27), if we make the substitution

$$y = \frac{\pi x}{L}, \text{ so } x = \frac{L}{\pi} y \text{ and } dx = \frac{L}{\pi} dy,$$

we find

$$\begin{aligned} \int_{-L}^L \sin\left(\frac{m\pi y}{L}\right) \sin\left(\frac{n\pi y}{L}\right) dy &= L\delta_{mn}, \\ \int_{-L}^L \cos\left(\frac{m\pi y}{L}\right) \cos\left(\frac{n\pi y}{L}\right) dy &= L\delta_{mn}, \\ \int_{-L}^L \cos\left(\frac{m\pi y}{L}\right) \sin\left(\frac{n\pi y}{L}\right) dy &= 0. \end{aligned} \tag{14.28}$$

⁴⁵If $f(x)$ has a finite discontinuity at x_0 , then $\lim_{x \rightarrow x_0^+} f(x) = f^+(x_0)$ and $\lim_{x \rightarrow x_0^-} f(x) = f^-(x_0)$, where $x \rightarrow x_0^{\pm}$ means x approaches x_0 from $x < x_0$ (– i.e., *smaller* values of x), or from $x > x_0$, (+ i.e., *larger* values of x). At x_0 the Fourier series converges to $\frac{1}{2}(f^+(x_0) + f^-(x_0))$.

⁴⁶Recall, $f(x)$ is even if $f(-x) = f(x)$, while $f(x)$ is odd if $f(-x) = -f(x)$. So $f(x) = x$ is odd, and $f(x) = x^2$ is even, but $f(x) = x + x^2$ or $f(x) = 1 + x$ are neither even nor odd.

Thus if we write

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right), \quad (14.29)$$

then

$$\begin{aligned} a_\ell &= \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{\ell\pi x}{L}\right) dx, \\ b_\ell &= \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{\ell\pi x}{L}\right) dx. \end{aligned} \quad (14.30)$$

In complex analysis, the sines and cosines are related to exponentials through

$$\cos \theta + i \sin \theta = \exp(i\theta).$$

This suggests we can write Fourier series as a series of complex exponentials by noting that

$$\begin{aligned} \cos\left(\frac{\ell\pi x}{L}\right) &= \frac{1}{2} \left[\exp\left(i\frac{\ell\pi x}{L}\right) + \exp\left(-i\frac{\ell\pi x}{L}\right) \right] \\ \sin\left(\frac{\ell\pi x}{L}\right) &= \frac{1}{2i} \left[\exp\left(i\frac{\ell\pi x}{L}\right) - \exp\left(-i\frac{\ell\pi x}{L}\right) \right] \end{aligned}$$

Inserting these into (14.29) yields

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp\left(i\frac{n\pi x}{L}\right)$$

where

$$\begin{aligned} c_n &= \frac{1}{2}(a_n - ib_n), \quad n > 0 \\ c_n &= \frac{1}{2}(a_n + ib_n), \quad n < 0 \\ c_n &= \frac{1}{2}a_0, \quad n = 0 \end{aligned}$$

Noting that

$$\int_{-L}^L \exp\left(-i\frac{n\pi x}{L}\right) \exp\left(+i\frac{m\pi x}{L}\right) dx = 2L \delta_{mn},$$

we can find c_ℓ directly by multiplying the series by $\exp\left(-i\frac{\ell\pi x}{L}\right)$ and integrating over the range of x :

$$c_\ell = \frac{1}{2L} \int_{-L}^L f(x) \exp\left(-i\frac{\ell\pi x}{L}\right) dx \quad (14.31)$$

14.5.6.2 The Fourier Transform

It is interesting to note that if we let $\ell\pi/L = k_\ell$, then $\Delta k_\ell = k_{\ell+1} - k_\ell = \pi/L$, and

$$f(x) = \sum_{\ell=-\infty}^{\infty} c_\ell \exp\left(i\frac{\ell\pi x}{L}\right) = \sum_{k_\ell=-\infty}^{\infty} \left(\frac{L}{\pi}c_\ell\right) \exp\left(ik_\ell x\right) \Delta k_\ell. \quad (14.32)$$

Also,

$$\left(\frac{L}{\pi}\right) c_\ell = \frac{1}{2\pi} \int_{-L}^L f(x) \exp(-ik_\ell x) dx. \quad (14.33)$$

These suggest, but by no means prove, that when $L \rightarrow \infty$,

$$g(k) \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \exp(-ikx) dx,$$

and

$$f(x) \triangleq \int_{-\infty}^{\infty} g(k) \exp(ikx) dk.$$

These can be rigorously established. The functions $f(x)$ and $g(k)$ are called Fourier transform pairs.⁴⁷ The function $g(k)$ is called the Fourier transform of $f(x)$ and is sometimes written $g(k) = \mathcal{F}[f(x)]$, with $f(x) = \mathcal{F}^{-1}[g(k)]$, where \mathcal{F} indicates Fourier transformation and \mathcal{F}^{-1} the inverse transformation. The Fourier transform can be helpful in solving differential equations through their property that

$$\mathcal{F}\left[\frac{df(x)}{dx}\right] = ik\mathcal{F}[f(x)],$$

if $\mathcal{F}[f(x)]$ exists.

14.5.6.3 Generalized Fourier Series

Assume that we have a set of functions $u_n(x)$, $n = 1, 2, \dots, \infty$, where u may be complex, that have the property

$$\int_a^b u_n^*(x) u_m(x) dx = \delta_{mn}.$$

⁴⁷A requirement for these to exist and be valid is

$$\int_{-\infty}^{\infty} f(x)^2 dx < \infty,$$

which means at a minimum $f(x) \rightarrow 0$ as $x \rightarrow \pm\infty$.

We desire to find a_n such that, for any function $f(x)$ defined on the interval $[a, b]$,

$$f(x) = \sum_{n=1}^{\infty} a_n u_n(x).$$

This is simple using the same procedure as with Fourier series. First multiply the series by $u_\ell^*(x)$, then integrate the result from $a \rightarrow b$. This yields,

$$a_\ell = \int_a^b u_\ell^*(x) f(x) dx$$

The requirements on f are similar to those for a Fourier series. The proof that the series converges can be difficult. It hinges on the *completeness* of the functions $u_n(x)$, which has been established for many of the so-called special functions. An example of a set such functions is the Legendre polynomials associated with Mie scattering theory presented in Section 3.3.1 of Chapter 3.

14.5.6.4 Relationship Between Fourier series and Least-Squares

An interesting question is, what are the best choices for a_n and b_n if we only use a finite number N of terms in Eq. (14.26)? Let's try a least-squares estimate for the coefficients. Let

$$S_N(x) = \frac{a_0}{2} + \sum_{n=1}^N a_n \cos(nx) + \sum_{n=1}^N b_n \sin(nx),$$

i.e., the *finite* sum of the series. Let us choose the a_n 's and b_n 's in a manner that minimizes

$$\Gamma = \int_{-\pi}^{\pi} (f(x) - S_N(x))^2 dx,$$

i.e., a quantity proportional to the square of the difference between $f(x)$ and $S_N(x)$ averaged over all x in the interval. Thus, we require

$$\frac{\partial \Gamma}{\partial a_\ell} = 0 \quad \text{and} \quad \frac{\partial \Gamma}{\partial b_\ell} = 0.$$

Now,

$$\begin{aligned} \frac{\partial \Gamma}{\partial a_\ell} &= 2 \int_{-\pi}^{\pi} (f(x) - S_N(x)) \frac{\partial S_N(x)}{\partial a_\ell} dx \\ &= 2 \int_{-\pi}^{\pi} (f(x) - S_N(x)) \cos(\ell x) dx \\ &= 0. \end{aligned}$$

Evaluating the integral yields,

$$a_\ell = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(\ell x) dx,$$

exactly the same as a_ℓ evaluated for Fourier series using the method described earlier. Similar results are obtained for b_ℓ . Thus, the Fourier coefficients are the least-squares fit of $f(x)$ to $S_N(x)$ averaged over the interval. In exactly the same way, it can be shown that the coefficients a_n in a generalized Fourier series $S_N(x) = \sum_{n=1}^N a_n u_n(x)$, represent the least-squares fit of $f(x)$ to $S_N(x)$ for the given N .

14.5.7 Orientational Averaging of Direction Cosines

In Section 14.4 we briefly considered in two dimensions what are known as *direction cosines*. Here we consider three dimensions and, in particular, the transformation between coordinate systems sharing a common origin, but rotated through an angle about some axis with respect to one another. We then compute the orientational average of various products of these direction cosines. One application is the computation of the orientational average of the induced dipole moment of an anisotropic molecule placed in an electric field (Chapter 1).

14.5.7.1 Direction Cosines

Consider a fixed coordinate system with cartesian axes denoted by X , Y , and Z , and a second system with cartesian axes denoted by x , y , and z . The second system shares the same origin as the fixed system but is rotated through an arbitrary angle about an arbitrary axis with respect to it. Let the unit vectors along the three axes in the (X, Y, Z) system be \hat{e}_X , \hat{e}_Y and \hat{e}_Z and along the (x, y, z) system be \hat{e}_x , \hat{e}_y and \hat{e}_z . The two sets of unit vectors are related by

$$\begin{aligned}\hat{e}_x &= (\hat{e}_x \bullet \hat{e}_X) \hat{e}_X + (\hat{e}_x \bullet \hat{e}_Y) \hat{e}_Y + (\hat{e}_x \bullet \hat{e}_Z) \hat{e}_Z, \\ \hat{e}_y &= (\hat{e}_y \bullet \hat{e}_X) \hat{e}_X + (\hat{e}_y \bullet \hat{e}_Y) \hat{e}_Y + (\hat{e}_y \bullet \hat{e}_Z) \hat{e}_Z, \\ \hat{e}_z &= (\hat{e}_z \bullet \hat{e}_X) \hat{e}_X + (\hat{e}_z \bullet \hat{e}_Y) \hat{e}_Y + (\hat{e}_z \bullet \hat{e}_Z) \hat{e}_Z,\end{aligned}\tag{14.34}$$

which in matrix form becomes

$$\begin{pmatrix} \hat{e}_x \\ \hat{e}_y \\ \hat{e}_z \end{pmatrix} = \begin{pmatrix} a_{xX} & a_{xY} & a_{xZ} \\ a_{yX} & a_{yY} & a_{yZ} \\ a_{zX} & a_{zY} & a_{zZ} \end{pmatrix} \begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix} \triangleq \mathbf{a} \begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix},\tag{14.35}$$

where $a_{xX} = \hat{e}_x \bullet \hat{e}_X$, $a_{xY} = \hat{e}_x \bullet \hat{e}_Y$, etc., are called *direction cosines*. The reverse transformation is

$$\begin{aligned}\hat{e}_X &= (\hat{e}_x \bullet \hat{e}_X)\hat{e}_x + (\hat{e}_y \bullet \hat{e}_X)\hat{e}_y + (\hat{e}_z \bullet \hat{e}_X)\hat{e}_z, \\ \hat{e}_Y &= (\hat{e}_x \bullet \hat{e}_Y)\hat{e}_x + (\hat{e}_y \bullet \hat{e}_Y)\hat{e}_y + (\hat{e}_z \bullet \hat{e}_Y)\hat{e}_z, \\ \hat{e}_Z &= (\hat{e}_x \bullet \hat{e}_Z)\hat{e}_x + (\hat{e}_y \bullet \hat{e}_Z)\hat{e}_y + (\hat{e}_z \bullet \hat{e}_Z)\hat{e}_z,\end{aligned}\tag{14.36}$$

and in matrix form is written

$$\begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix} = \begin{pmatrix} a_{xX} & a_{yX} & a_{zX} \\ a_{xY} & a_{yY} & a_{zY} \\ a_{xZ} & a_{yZ} & a_{zZ} \end{pmatrix} = \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \\ \hat{e}_z \end{pmatrix} = \tilde{\mathbf{a}} \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \\ \hat{e}_z \end{pmatrix},\tag{14.37}$$

Thus, if we use the notation

$$\hat{e}_s = \begin{pmatrix} \hat{e}_x \\ \hat{e}_y \\ \hat{e}_z \end{pmatrix}, \quad \text{and} \quad \hat{e}_L = \begin{pmatrix} \hat{e}_X \\ \hat{e}_Y \\ \hat{e}_Z \end{pmatrix},$$

where the subscripts “s” and “L” stand for unit vectors with “small” and “large” subscripts, respectively, then simply

$$\hat{e}_s = \mathbf{a} \hat{e}_L \quad \text{and} \quad \hat{e}_L = \tilde{\mathbf{a}} \hat{e}_s$$

This shows that $\tilde{\mathbf{a}} = \mathbf{a}^{-1}$, i.e., the inverse of the matrix \mathbf{a} is the transpose of \mathbf{a} .

Now, any vector \vec{V} can be written in terms of its components in the X, Y, Z or the x, y, z systems. To distinguish these two representations let

$$\begin{aligned}\vec{V} &= V_X \hat{e}_X + V_Y \hat{e}_Y + V_Z \hat{e}_Z \\ \vec{v} &= v_x \hat{e}_x + v_y \hat{e}_y + v_z \hat{e}_z.\end{aligned}$$

where \vec{V} and \vec{v} are the *same* vector, but expanded in terms of components in different ways. Different symbols are used for this vector for the purpose of clarity. How are the components in the two systems related? To find out, use Eqs. (14.34) and (14.35) to replace the \hat{e}_x, \hat{e}_y , etc. with the \hat{e}_X, \hat{e}_Y , etc:

$$\begin{aligned}\vec{v} &= v_x \hat{e}_x + v_y \hat{e}_y + v_z \hat{e}_z \\ &= a_{xX} v_x \hat{e}_X + a_{xY} v_x \hat{e}_Y + a_{xZ} v_x \hat{e}_Z + \cdots + a_{zZ} v_z \hat{e}_Z \\ &= (a_{xX} v_x + a_{yX} v_y + a_{zX} v_z) \hat{e}_X + \cdots + a_{zZ} v_z \hat{e}_Z\end{aligned}$$

or

$$V_x = a_{xX} v_x + a_{yX} v_y + a_{zX} v_z, \quad \text{etc.}$$

In matrix form

$$\mathbf{V} = \begin{pmatrix} a_{xX} & a_{yX} & a_{zX} \\ a_{xY} & a_{yY} & a_{zY} \\ a_{xZ} & a_{yZ} & a_{zZ} \end{pmatrix} \mathbf{v} = \tilde{\mathbf{a}} \mathbf{v}.$$

Similarly, $\mathbf{v} = \mathbf{a} \mathbf{V}$. Now, although the components of the vector \vec{V} change when the coordinate system is rotated, its magnitude should not, i.e., $\tilde{\mathbf{V}} \mathbf{V} = \tilde{\mathbf{v}} \mathbf{v}$. Is this satisfied? Let's check

$$\tilde{\mathbf{V}} \mathbf{V} = (\widetilde{\tilde{\mathbf{a}} \mathbf{v}}) \tilde{\mathbf{a}} \mathbf{v} = \tilde{\mathbf{v}} \tilde{\mathbf{a}} \mathbf{v} = \tilde{\mathbf{v}} (\mathbf{a} \tilde{\mathbf{a}}) \mathbf{v} = \tilde{\mathbf{v}} \mathbf{v},$$

where the last step derives from $\tilde{\mathbf{a}} = \mathbf{a}^{-1}$, i.e., $\mathbf{a} \tilde{\mathbf{a}} = \mathbf{1}$.

At this point it is useful to simplify the notation. Let's use upper case indices to denote the unit vectors in the fixed system, e.g., $\hat{e}_I = \hat{e}_X$, for $I = 1$, etc., and lower case indices to denote the e.g., $\hat{e}_i = \hat{e}_x$, for $i = 1$, etc. So, more compactly, Eq. (14.34) can be written

$$\hat{e}_i = \sum_{I=1}^3 (\hat{e}_i \bullet \hat{e}_I) \hat{e}_I = \sum_{I=1}^3 a_{iI} \hat{e}_I, \quad (14.38)$$

and the reverse (inverse) equations;

$$\hat{e}_I = \sum_{i=1}^3 (\hat{e}_I \bullet \hat{e}_i) \hat{e}_i = \sum_{i=1}^3 a_{iI} \hat{e}_i. \quad (14.39)$$

Since the cartesian axes are all mutually perpendicular, $\hat{e}_i \bullet \hat{e}_j = \delta_{ij}$, where $\delta_{ij} = 1$, if $i = j$, and 0, if $i \neq j$, and similarly $\hat{e}_I \bullet \hat{e}_J = \delta_{IJ}$. Because $\hat{e}_i \bullet \hat{e}_j = \delta_{ij}$,

$$\hat{e}_i \bullet \hat{e}_j = \delta_{ij} = \sum_{I=1}^3 \sum_{J=1}^3 a_{iI} a_{jJ} \hat{e}_I \bullet \hat{e}_J = \sum_{I=1}^3 \sum_{J=1}^3 a_{iI} a_{jJ} \delta_{IJ} = \sum_{I=1}^3 a_{iI} a_{jI}.$$

Therefore,

$$\sum_{I=1}^3 a_{iI}^2 = 1 \quad \text{and} \quad \sum_{I=1}^3 a_{iI} a_{jI} = 0 \quad \text{for } i \neq j. \quad (14.40)$$

Similarly,

$$\sum_{i=1}^3 a_{iI}^2 = 1 \quad \text{and} \quad \sum_{i=1}^3 a_{iI} a_{iJ} = 0 \quad \text{for } I \neq J. \quad (14.41)$$

Thus, there are a total of *six* relationships among the nine a_{iJ} 's, so only three are independent. This reflects that fact that any position of x, y, z relative to X, Y, Z can be achieved by rotating x, y, z relative to X, Y, Z through a single angle (requiring one parameter to specify) about a given direction (requiring two parameters to specify).

14.5.7.2 Orientational Averaging

In Chapter 1, Section 1.7.5, we needed to compute the orientational average of the dipole moment of an anisotropic molecule subjected to an electric field with a direction fixed relative to a laboratory coordinate system. This required averages of quantities like

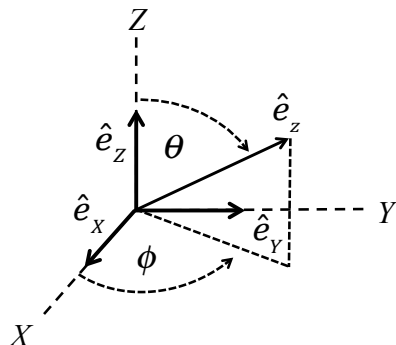
$$a_{iI} a_{iJ} \quad \text{and} \quad a_{iI} a_{iJ} a_{jI} a_{jM},$$

where some or all of the indices may be the same, over all orientations of the molecular-fixed (rotated) system relative to the laboratory (fixed) system. By such an *orientational average*, denoted by angle brackets $\langle \cdots \rangle$, we mean

$$\langle f \rangle = \frac{1}{4\pi} \int f d\Omega,$$

where f is a function of direction (e.g., θ and ϕ in spherical coordinates), Ω is the solid angle in the *fixed* system, and the integration is to be taken over the entire range of solid angles (4π).

Although somewhat tedious, the averages of quantities such as given above are not difficult to find. For example, consider the case where all the indices are all the same, i.e.,



a_{iI}^2 . The figure above shows that

$$\hat{e}_z = \sin \theta \cos \phi \hat{e}_X + \sin \theta \sin \phi \hat{e}_Y + \cos \theta \hat{e}_Z.$$

Using this it is easy to find $\langle a_{zZ}^2 \rangle$ since $a_{zZ} = \hat{e}_z \cdot \hat{e}_Z = \cos \theta$:

$$\langle a_{zZ}^2 \rangle = \frac{1}{4\pi} \int (\hat{e}_z \cdot \hat{e}_Z)^2 d\Omega = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \cos^2 \theta \sin \theta d\theta = \frac{1}{3}.$$

Now, by symmetry, the average of the cosine to any power of the angle between \hat{e}_z and \hat{e}_Z must be the same as the average of the cosine to the same power of the angle between \hat{e}_x and \hat{e}_X or between \hat{e}_y and \hat{e}_Y . Thus,

$$\langle a_{xX}^2 \rangle = \langle a_{yY}^2 \rangle = \langle a_{zZ}^2 \rangle.$$

Note that any of these three terms can be transformed into any other by a simple relabeling of the axes. For example, letting $x \rightarrow y$, $y \rightarrow z$, $z \rightarrow x$, $X \rightarrow Y$, $Y \rightarrow Z$ and $Z \rightarrow X$, the first is transformed into the second, etc. In what follows we shall employ the rule that any terms that can be transformed into one another by a relabeling of one or both sets of axes must have identical orientational averages, i.e., a relabeling of the axes *cannot* change the averages. Therefore

$$\langle a_{iI}^2 \rangle = \frac{1}{3}.$$

Now, taking the orientational average of Eq. (14.41) provides

$$\sum_{i=1}^3 \langle a_{iI} a_{iJ} \rangle = 0 \quad \text{for } I \neq J,$$

or

$$\langle a_{xI} a_{xJ} \rangle + \langle a_{yI} a_{yJ} \rangle + \langle a_{zI} a_{zJ} \rangle = 0 \quad \text{for } I \neq J.$$

Relabeling the x , y , and z axes simply transforms the various terms into one another, so they are all identical. Since they sum to zero, $\langle a_{iI} a_{iJ} \rangle = 0$ for $I \neq J$. Thus,

$$\langle a_{iI} a_{iJ} \rangle = \frac{1}{3} \delta_{IJ}.$$

Using a similar procedure orientational averages of the product of four direction cosines are easily found. Letting all of the indices be the same, we have

$$\langle a_{zZ}^4 \rangle = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \cos^4 \theta \sin \theta d\theta = \frac{1}{5}.$$

By the same reasoning as for $\langle a_{zZ}^2 \rangle$,

$$\langle a_{xX}^4 \rangle = \langle a_{yY}^4 \rangle = \langle a_{zZ}^4 \rangle,$$

therefore

$$\langle a_{iI}^4 \rangle = \frac{1}{5}.$$

Now, if we square the relationship $\sum_{I=1}^3 a_{iI}^2 = 1$, and take the orientational average of the result we get

$$\langle a_{xX}^4 \rangle + \langle a_{yY}^4 \rangle + \langle a_{zZ}^4 \rangle + 2\langle a_{xX}^2 a_{yY}^2 \rangle + 2\langle a_{yY}^2 a_{zZ}^2 \rangle + 2\langle a_{xX}^2 a_{zZ}^2 \rangle = 1$$

or

$$\sum_{I=1}^3 \langle a_{iI}^4 \rangle + 2 \sum_{I=1}^3 \sum_{J>I=1}^3 \langle a_{iI}^2 a_{iJ}^2 \rangle = 1$$

Consider the terms multiplied by two in the above equation. If we make the transformation $x \rightarrow y, y \rightarrow z, z \rightarrow x, X \rightarrow Y, Y \rightarrow Z$ and $Z \rightarrow X$, then the first term becomes the second, the second becomes the third and the third becomes the first, so each term in the each of the individual sums can be transformed into another term in the sum by a relabeling of the axes. As stated above the average of any terms that can be transformed to each other by a relabeling of the axes must be equal, therefore

$$\frac{3}{5} + 6\langle a_{iI}^2 a_{iJ}^2 \rangle = 1 \quad \text{so} \quad \langle a_{iI}^2 a_{iJ}^2 \rangle = \frac{1}{15}, \quad I \neq J.$$

Note that when writing $\langle a_{iI}^2 a_{iJ}^2 \rangle$ in the equation above we implicitly used the fact that $I \neq J$ in the second term (if $I = J$ the answer is $1/5$ not $1/15$).

Now, do the same for $i \neq j$, i.e., average $\left(\sum_{I=1}^3 a_{iI} a_{jI}\right)^2 = 0$. Combining terms that are simple renumbering of the axes, we have

$$3\langle a_{iI}^2 a_{jI}^2 \rangle + 6\langle a_{iI} a_{iJ} a_{jI} a_{jJ} \rangle = 0$$

or

$$\langle a_{iI} a_{iJ} a_{jI} a_{jJ} \rangle = -\frac{1}{30}, \quad i \neq j, \quad I \neq J.$$

Now consider $\langle a_{iI} a_{iI} a_{jI} a_{jJ} \rangle$ with $i \neq j$ and $I \neq J$. This kind of product can be examined by multiplying $\sum_{i=1}^3 a_{iI} a_{iI} = 1$ by $\sum_{j=1}^3 a_{jI} a_{jJ} = 0$. Then as before,

$$3\langle a_{iI}^3 a_{iJ} \rangle + 6\langle a_{iI}^2 a_{jI} a_{jJ} \rangle = 0.$$

The first term can be evaluated in a manner similar to the evaluation of $\langle a_{zZ}^4 \rangle$: take $i = z$, $I = X$, and $J = Y$, then

$$\langle a_{iI}^3 a_{iJ} \rangle = \langle (\hat{e}_z \cdot \hat{e}_X)^3 (\hat{e}_z \cdot \hat{e}_Y) \rangle = \langle \sin^4 \theta \cos^3 \phi \sin \phi \rangle = 0,$$

where the last equality results from a direct evaluation of the integral. This shows that

$$\langle a_{iI}^3 a_{iJ} \rangle = 0, \quad i \neq j \quad \text{and} \quad \langle a_{iI}^2 a_{jI} a_{jJ} \rangle = 0, \quad i \neq j, \quad I \neq J.$$

Finally, consider $\langle a_{iI} a_{iJ} a_{jI} a_{jM} \rangle$, where $i \neq j$ and $I \neq J \neq M$, by examining

$$\left\langle \left(\sum_{i=1}^3 a_{iI} a_{iJ} \right) \left(\sum_{j=1}^3 a_{jI} a_{jM} \right) \right\rangle = 0,$$

which yields

$$3\langle a_{iJ} a_{iM} a_{iI}^2 \rangle + 6\langle a_{iI} a_{iJ} a_{jI} a_{jM} \rangle = 0.$$

The first term can be evaluated directly (let $i = x$, $I = Z$, $J = X$, and $M = Y$) to give zero, so

$$\langle a_{iJ} a_{iM} a_{iI}^2 \rangle = 0, \quad I \neq J \neq M, \quad \text{and} \quad \langle a_{iI} a_{iJ} a_{jI} a_{jM} \rangle = 0, \quad i \neq j, \quad I \neq J \neq M.$$

These are all of the averages that are needed in the text and are conveniently summarized in the table below.

Orientation averages of direction cosines.

Average	Result	Conditions
$\langle a_{iI}^2 \rangle$	1/3	
$\langle a_{iI} a_{iJ} \rangle$	0	$I \neq J$
$\langle a_{iI}^4 \rangle$	1/5	
$\langle a_{iI}^2 a_{iJ}^2 \rangle$	1/15	$I \neq J$
$\langle a_{iI} a_{iJ} a_{jI} a_{jJ} \rangle$	-1/30	$i \neq j, I \neq J$
$\langle a_{iI}^3 a_{iJ} \rangle$	0	$i \neq j$
$\langle a_{iI}^2 a_{jI} a_{jJ} \rangle$	0	$i \neq j, I \neq J$
$\langle a_{iI}^2 a_{iJ} a_{iM} \rangle$	0	$I \neq J \neq M$
$\langle a_{iI} a_{iJ} a_{jI} a_{jM} \rangle$	0	$i \neq j, I \neq J \neq M$

14.6 Bibliographic Notes

The material here can be found in any number of books on mathematical methods in physics. They are available at both the undergraduate and graduate level. Having taught mathematical methods at both the undergraduate and graduate levels, the author believes the best at the undergraduate level is [Nearing \[2010\]](#). It can be found online and downloaded in several forms at “<http://web2.physics.miami.edu/~nearing/mathmethods/>”. At the graduate level, two excellent books that treat most of the topics discussed here are [Arfken \[1970\]](#) and [Butkov \[1968\]](#). For numerical analysis and curve fitting, i.e., least squares, an essential reference is [Press et al. \[1992\]](#).

Bibliography

- Acker, J. G. (2015). *The Color of the Atmosphere with the Ocean Below*. Greenbelt, MD: J. G. Acker.
- Ahmad, Z., B. A. Franz, C. R. McClain, E. J. Kwiatkowska, J. Werdell, E. P. Shettle, and B. N. Holben (2010). New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans. *Applied Optics* 49, 5545–5560.
- Anderson, G. P., J. H. Chetwynd, S. A. Clough, E. P. Shettle, and F. X. Kneizys (1986). Atmospheric Constituent Profiles. Technical Report AFGL-TR-86-0110, Air Force Geophysics Laboratory, Hanscom AFB, MA 01731.
- Arfken, G. (1970). *Mathematical Methods for Physicists*. New York: Academic Press.
- Bailey, S. W., B. A. Franz, and P. J. Werdell (2010). Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. *Optics Express* 18, 7521–7527.
- Baker, K. S. and R. C. Smith (1982). Bio-optical classification and model of natural waters. 2. *Limnology and Oceanography* 27, 500–509.
- Ball Aerospace Division (1979). Development of the Coastal Zone Color Scanner for Nimbus-7: Volume 1 - Mission Objectives and Instrument Description. Final Report F78-11, Rev. A NASA Contract NAS5-20900.
- Bandeem, W. R. and R. S. Fraser (1982). Radiative effects of the El Chichón volcanic eruption: Preliminary results concerning remote sensing. NASA Technical Memorandum 84959, Goddard Space Flight Center, Greenbelt, MD.
- Banzon, V. F., R. E. Evans, H. R. Gordon, and R. M. Chomko (2004). SeaWiFS observations of the Arabian Sea southwest monsoon bloom for the year 2000. *Deep-Sea Research II* 51, 189–208.
- Barger, V. and M. G. Olssen (1986). *Classical Electricity and Magnetism*. Prentice Hall.

- Barger, V. and M. G. Olssen (1995). *Classical Mechanics: A Modern Perspective*. McGraw-Hill.
- Barnes, R. A., R. E. Eplee Jr., F. S. Patt, H. H. Kieffer, T. C. Stone, G. Meister, J. J. Butler, and C. R. McClain (2004). Comparison of SeaWiFS measurements of the Moon with the U.S Geological Survey Lunar model. *Applied Optics* 43, 5838–5854.
- Barnes, R. A., R. E. Eplee Jr., F. S. Patt, and C. R. McClain (1999). Changes in the radiometric sensitivity of Seawifs determined from Lunar and Solar-based measurements. *Applied Optics* 38, 4649–4664.
- Barnes, R. A., A. W. Holmes, W. L. Barnes, W. E. Esaias, C. R. McClain, and T. Svitek (1994a). Seawifs prelaunch radiometric calibration and spectral characterization. SeaWiFS Technical Report Series: Volume 23, NASA, Greenbelt, MD, Technical Memorandum 104566.
- Barnes, R. A., A. W. Holmes, W. L. Barnes, W. E. Esaias, C. R. McClain, and T. Svitek (1994b). SeaWiFS prelaunch radiometric calibration and spectral characterization. SeaWiFS Technical Report Series: Volume 23, NASA Technical Memorandum 104566, Goddard Space Flight Center, Greenbelt, MD.
- Bartlett, J. S. (1996). The influence of Raman scattering by seawater and fluorescence by phytoplankton on ocean color. Master's thesis, Dalhousie University.
- Bartlett, J. S., K. J. Voss, S. Sathyendranath, and A. Vodacek (1998). Raman scattering by pure water and seawater. *Applied Optics* 37, 3324–3332.
- Bates, D. R. (1984). Rayleigh scattering by air. *Planetary and Space Science* 32, 785–790.
- Bauer, D. and A. Ivanoff (1965). Au sujet de la mesure du coefficient de diffusion de la lumiere par les eaux de mer pour des angles compris entre 14° et $1^\circ 30'$. *Comptes Rendus de L'Academie Sciences, Paris* 260, 631–634.
- Bauer, D. and A. Morel (1967). Etude aus petit angles de l'indicatrice de la lumiere par eaux de mer. *Annales de Geophysiques* 23, 109–123.
- Beardsley, G. F. and J. R. V. Zaneveld (1969). Theoretical dependence of the near-asymptotic apparent optical properties on the inherent optical properties of sea water. *Journal of the Optical Society of America* 59, 373–377.
- Bodhaine, B. A., N. B. Wood, E. G. Dutton, and J. R. Susser (1999). On Rayleigh optical depth calculations. *Journal of Atmospheric and Oceanic Technology* 16, 1854–1861.
- Bogumil, K., J. Orphal, T. Homann, S. Voigt, P. Spietz, O. C. Fleischman, A. Vogel, M. Hartman, H. Bovensmann, J. Frerick, and J. P. Burrows (2003). Measurements

- of molecular absorption spectra with the SCIAMACHY pre-flight model: Instrument characterization and reference data for atmospheric remote sensing in the 230-2380 nm region. *Journal of Photochemistry and Photobiology A: Chemistry* 157, 167–184.
- Bohren, C. F. and D. R. Huffman (1983). *Absorption and Scattering of Light by Small Particles*. New York, NY: Wiley.
- Born, M. and E. Wolf (1975). *Principles of Optics*. Oxford, U.K.: Pergamon Press.
- Bortkovskii, R. S. (1987). *Air-Sea Exchange of Heat and Moisture During Storms*. Dordrecht: Reidel.
- Boss, E., W. H. Slade, M. J. Behrenfeld, and G. Dall’Olmo (2009). Acceptance angle effects on the beam attenuation coefficient in the ocean. *Optics Express* 17(3), 1535–1550.
- Boydton, G. C. and H. R. Gordon (2000). Irradiance inversion algorithm for estimating the absorption and backscattering coefficients of natural waters: Raman scattering effects. *Applied Optics* 39, 3012–3022.
- Boydton, G. C. and H. R. Gordon (2002). An irradiance inversion algorithm for absorption and backscattering profiles in natural waters: Improvement for very clear waters. *Applied Optics* 41, 2224–2227.
- Bricaud, A. and A. Morel (1986). Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. *Applied Optics* 25, 571–580.
- Bricaud, A., A. Morel, M. Babin, K. Allali, and H. Claustre (1998). Variations of light absorption by suspended particles with chlorophyll *a* concentration in oceanic (Case 1) waters: Analysis and implications for bio-optical models. *Journal of Geophysical Research* 103(C13), 31033–31044.
- Bricaud, A., A. Morel, and L. Prieur (1983). Optical efficiency factors of some phytoplankters. *Limnology and Oceanography* 28, 816–832.
- Brown, O. B. and H. R. Gordon (1973). Two component Mie scattering models of Sargasso Sea particles. *Applied Optics* 12, 2461–2465.
- Brown, O. B. and H. R. Gordon (1974). The size-refractive index distribution of clear coastal water particulates from light scattering. *Applied Optics* 13, 2874–2881.
- Butkov, E. (1968). *Mathematical Physics*. Reading, MA: Addison-Wesley.
- Case, K. M. (1957). Transfer problems and the reciprocity principle. *Reviews of Modern Physics* 29, 651–663.

- Case, K. M. and P. F. Zweifel (1967). *Linear Transport Theory*. Reading MA: Addison-Wesley.
- Cattrall, C., K. L. Carder, K. J. Thome, and H. R. Gordon (2002). Solar-reflectance-based calibration of spectral radiometers. *Geophysical Research Letters* 29, (2–1)–(2–4).
- Chandrasekhar, S. (1950). *Radiative Transfer*. Oxford, U.K.: Oxford University Press.
- Chang, C. H. and L. A. Young (1974). Sea water temperature measurement from Raman spectra. Avco Everett Research Laboratory, Inc., 2385 Revere Beach Parkway, Everett, MA, Research Note 960, N62269-73-C-0073, sponsored by Advanced Research Projects Agency, ARPA Order No. 2194.
- Chomko, R. and H. R. Gordon (1998). Atmospheric correction of ocean color imagery: Use of the Junge power-law aerosol size distribution with variable refractive index to handle aerosol absorption. *Applied Optics* 37, 5560–5572.
- Chomko, R. M. and H. R. Gordon (2001). Atmospheric correction of ocean color imagery: Test of the spectral optimization algorithm with SeaWiFS. *Applied Optics* 40, 2973–2984.
- Chomko, R. M., H. R. Gordon, S. Maritorena, and D. A. Siegel (2003). Simultaneous determination of oceanic and atmospheric parameters for ocean color imagery by spectral optimization: A validation. *Remote Sensing of Environment* 84, 208–220.
- Clark, D. K., E. T. Baker, and A. E. Strong (1980). Upwelled spectral radiance distribution in relation to particulate matter in sea water. *Boundary-Layer Meteorology* 18, 287–298.
- Clark, D. K., H. R. Gordon, K. J. Voss, Y. Ge, W. Broenkow, and C. Trees (1997). Validation of atmospheric correction over the oceans. *Journal of Geophysical Research* 102(D), 17209–17217.
- Clark, D. K., M. A. Yarbrough, M. E. Feinholz, S. Flora, W. Broenkow, Y. S. Kim, B. C. Johnson, S. W. Brown, M. Yuen, and J. L. Mueller (2003). MOBY, a radiometric buoy for performance monitoring and vicarious calibration of satellite ocean color sensors: Measurement and data analysis protocols. Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision Volume VI: Special Topics in Ocean Optics Protocols and Appendices, NASA Technical Memorandum 2003-21621, Goddard Space Flight Center.
- Clarke, G. L., G. C. Ewing, and C. J. Lorenzen (1970). Spectra of backscattered light from the sea obtained from aircraft as a measurement of chlorophyll concentration. *Science* 167, 1119–1121.
- Clavano, W. R., E. Boss, and L. Kamp-Boss (2009). Inherent optical properties of non-spherical marine-like particles — From theory to observation. In R. N. Gibson, R. J. A.

- Atkinson, and J. D. M. Gibson (Eds.), *Oceanography and Marine Biology. An Annual Review, Volume 45*, pp. 1–38. Boca Raton, FL: CBC Press.
- Cox, C. and W. Munk (1954). Measurements of the roughness of the sea surface from photographs of the sun's glitter. *Journal of the Optical Society of America* 44, 838–850.
- Cruz, R. A., A. Marcano, C. Jacinto, and T. Catunda (2009). Ultrasensitive thermal lens spectroscopy of water. *Optics Letters* 34, 1882–1884.
- Dall'Olmo, G., T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade (2009). Significant contribution of large particles to optical backscattering in the open ocean. *Biogeosciences* 6, 947–967.
- Deirmendjian, D. (1969). *Electromagnetic Scattering on Spherical Polydispersions*. New York, NY: Elsevier.
- Delbouille, L., L. Neven, and G. Roland (1973). Photometric atlas of the Solar spectrum from $\lambda 3000$ to $\lambda 10000$. Institut d'Astrophysique de l'Universit de Liège Report, (Belgium).
- Dereniak, E. L. and D. G. Crowe (1984). *Optical Radiation Detectors*. New York, NY: Wiley.
- Ding, K. and H. R. Gordon (1994). Atmospheric correction of ocean color sensors: Effects of Earth curvature. *Applied Optics* 33, 7096–7106.
- Ding, K. and H. R. Gordon (1995). Analysis of the influence of O₂ A-band absorption on atmospheric correction of ocean color imagery. *Applied Optics* 34, 2068–2080.
- DiToro, D. M. (1978). Optics of turbid estuarine waters: Approximations and applications. *Water Research* 12, 1059–1068.
- Draine, B. T. (1988). The discrete-dipole approximation and its application to interstellar graphite grains. *Astrophysical Journal* 333, 848–872.
- Duntley, S. Q. (1954). Measurement of the distribution of water wave slopes. *Journal of the Optical Society of America* 44, 574–575.
- Ebuchi, N. and S. Kizu (2002). Probability distribution of surface wave slope derived using sun glitter images from geostationary meteorological satellite and surface vector winds from scatterometers. *Journal of Oceanography* 58, 477–486.
- Eplee Jr., R. E. and S. W. Bailey (2014). On orbit calibration of ocean color reflective Solar bands. In G. Zibordi, C. J. Donlon, and A. C. Parr (Eds.), *Optical Radiometry for Ocean Climate Measurements*, pp. 121–152. Academic Press.

- Eplee Jr., R. E., W. D. Robinson, S. W. Bailey, D. K. Clark, P. J. Werdell, M. Wang, R. A. Barnes, and C. R. McClain (2001). Calibration of SeaWiFS. II. Vicarious techniques. *Applied Optics* 40, 6701–6718.
- Evans, R. H. and H. R. Gordon (1994). CZCS system calibration: A retrospective examination. *Journal of Geophysical Research* 99(C), 7293–7307.
- Eyring, H., J. Walter, and G. E. Kimball (1944). *Quantum Chemistry*. London: Wiley.
- Farinato, R. S. and R. L. Rowell (1976). New values of the light scattering depolarization and anisotropy of water. *Journal of Chemical Physics* 65, 593–595.
- Fessenkov, U. G. (1962). Photometry of the Moon. In Z. Kopal (Ed.), *Physics and Astronomy of the Moon*, pp. 99–130. New York, N.Y: Academic.
- Feynman, R. P. (1985). *QED: The strange theory of light and matter*. Princeton, NJ: Princeton University Press.
- Feynman, R. P., R. B. Leighton, and M. Sands (1964). *The Feynman Lectures on Physics (Vol. 2)*. Reading, MA: Addison-Wesley.
- Fournier, G. R. and J. L. Forand (1994). Analytic phase function for ocean water. *Society of Photo-Optical Instrumentation Engineers, Ocean Optics XII* 2258, 194–201.
- Franz, B. A., S. W. Bailey, P. J. Werdell, and C. R. McClain (2007). Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. *Applied Optics* 46, 5068–5082.
- Fraser, R. S. (1976). Satellite measurement of mass of Sahara dust in the atmosphere. *Applied Optics* 15, 2471–2479.
- Frouin, R., M. Schwindling, and P. Y. Deschamps (1996). Spectral reflectance of sea foam in the visible and near-infrared: In-situ measurements and implications for remote sensing of ocean color and aerosols. *Journal of Geophysical Research* 101(C), 14361–14371.
- Fry, E. S., G. W. Kattawar, and R. M. Pope (1992). Integrating cavity absorption meter. *Applied Optics* 31, 2055–2065.
- Fry, E. S. and K. J. Voss (1985). Measurement of the mueller matrix for phytoplankton. *Limnology and Oceanography* 30(6), 1322–1326.
- Gao, B.-C., A. F. H. Goetz, and W. J. Wiscombe (1993). Cirrus cloud detection from airborne imaging spectrometer data using the 1.38 micron water vapor band. *Geophysical Research Letters* 20, 301–304.

- Gardner, W. D., I. D. Walsh, and M. J. Richardson (1993). Biophysical forcing of particle production and distribution during a spring bloom in the North Atlantic. *Deep-Sea Research* 40, 171–195.
- Ge, Y., K. J. Voss, and H. R. Gordon (1995). In situ measurements of inelastic scattering in monterey bay using Solar Fraunhofer lines. *Journal of Geophysical Research* 100, 13227–13236.
- Goldstein, H. (1980). *Classical Mechanics (2nd Ed.)*. Reading, MA: Addison-Wesley.
- Goody, R. M. and Y. L. Yung (1989). *Atmospheric Radiation, Theoretical Basis*. Oxford, U.K.: Oxford University Press.
- Gordon, H. (2007a). Rayleigh-gans scattering approximation: Surprisingly useful for understanding backscattering from disk-like particles. *Optics Express* 15(9), 5572–5588.
- Gordon, H. R. (1973). Simple calculation of the diffuse reflectance of the ocean. *Applied Optics* 12, 2803–2804.
- Gordon, H. R. (1976). Radiative transfer: a technique for simulating the ocean in satellite remote sensing calculations. *Applied Optics* 15, 1974–1979.
- Gordon, H. R. (1978a). Remote sensing of optical properties in continuously stratified waters. *Applied Optics* 17, 1893–1897.
- Gordon, H. R. (1978b). Removal of atmospheric effects from satellite imagery of the oceans. *Applied Optics* 17, 1631–1636.
- Gordon, H. R. (1979). The diffuse reflectance of the ocean: The theory of its augmentation by chlorophyll *a* fluorescence at 685 nm. *Applied Optics* 18, 1161–1166.
- Gordon, H. R. (1980). Irradiance attenuation coefficient in a stratified ocean: A local property of the medium. *Applied Optics* 19, 2092–2094.
- Gordon, H. R. (1981). Reduction of error introduced in the processing of Coastal Zone Color Scanner-type imagery resulting from sensor calibration and Solar irradiance uncertainty. *Applied Optics* 20, 207–210.
- Gordon, H. R. (1987). Calibration requirements and methodology for remote sensors viewing the oceans in the visible. *Remote Sensing of Environment* 22, 103–126.
- Gordon, H. R. (1988). Ocean color remote sensing systems: Radiometric requirements. *Society of Photo-Optical Instrumentation Engineers, Recent Advances in Sensors, Radiometry and Data Processing for Remote Sensing* 924, 151–167.

- Gordon, H. R. (1990). Radiometric considerations for ocean color remote sensors. *Applied Optics* 29, 3228–3236.
- Gordon, H. R. (1992). Diffuse reflectance of the ocean: Influence of nonuniform phytoplankton pigment profile. *Applied Optics* 31, 2116–2129.
- Gordon, H. R. (1993). The sensitivity of radiative transfer to small-angle scattering in the ocean: A quantitative assessment. *Applied Optics* 32, 7505–7511.
- Gordon, H. R. (1995). Remote sensing of ocean color: A methodology for dealing with broad spectral bands and significant out-of-band response. *Applied Optics* 34, 8363–8374.
- Gordon, H. R. (1997). Atmospheric correction of ocean color imagery in the Earth Observing System era. *Journal of Geophysical Research* 102(D), 17081–17106.
- Gordon, H. R. (1998). In-orbit calibration strategy for ocean color sensors. *Remote Sensing of Environment* 63, 265–278.
- Gordon, H. R. (1999). Contribution of Raman scattering to water-leaving radiance: A reexamination. *Applied Optics* 38, 3166–3174.
- Gordon, H. R. (2002). Inverse methods in hydrologic optics. *Oceanologia* 44, 9–58.
- Gordon, H. R. (2005). Normalized water-leaving radiance: Revisiting the influence of surface roughness. *Applied Optics* 44, 241–248.
- Gordon, H. R. (2006). Backscattering of light from disk-like particles: Is fine-scale structure or gross morphology more important? *Applied Optics* 45, 7166–7173.
- Gordon, H. R. (2007b). Backscattering of light from disk-like particles with aperiodic angular fine structure. *Optics Express* 15(25), 16424–16430.
- Gordon, H. R. (2011). Light scattering and absorption by randomly-oriented cylinders: Dependence on aspect ratio for refractive indices applicable for marine particles. *Optics Express* 19, 4673–4691.
- Gordon, H. R. and G. C. Boynton (1997). A radiance - irradiance inversion algorithm for estimating the absorption and backscattering coefficients of natural waters: Homogeneous waters. *Applied Optics* 36, 2636–2641.
- Gordon, H. R. and G. C. Boynton (1998). A radiance - irradiance inversion algorithm for estimating the absorption and backscattering coefficients of natural waters: Vertically stratified water bodies. *Applied Optics* 37, 3886–3896.
- Gordon, H. R., J. W. Brown, O. B. Brown, R. H. Evans, and D. K. Clark (1983). Nimbus-7 Coastal Zone Color Scanner: Reduction of its radiometric sensitivity with time. *Applied Optics* 22, 3929–3931.

- Gordon, H. R. and O. B. Brown (1972). A theoretical model of light scattering by Sargasso Sea particulates. *Limnology and Oceanography* 17, 826–832.
- Gordon, H. R., O. B. Brown, and H. Bader (1972). An experimental and theoretical study of the suspended particulate matter in the Tongue of the Ocean and its influence on underwater visibility. In: Measurement of the volume scattering function of sea water. Ed. L.E. Mertins and D.L. Phillips, Technical Report 334, Range Measurement Laboratory, Patrick Air Force Base, FL, USA.
- Gordon, H. R., O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark (1988). A semi-analytic radiance model of ocean color. *Journal of Geophysical Research* 93(D), 10909–10924.
- Gordon, H. R., O. B. Brown, and M. M. Jacobs (1975). Computed relationships between the inherent and apparent optical properties of a flat homogeneous ocean. *Applied Optics* 14, 417–427.
- Gordon, H. R. and D. K. Clark (1980a). Atmospheric effects in the remote sensing of phytoplankton pigments. *Boundary-Layer Meteorology* 18, 299–313.
- Gordon, H. R. and D. K. Clark (1980b). Remote sensing optical properties of a stratified ocean: An improved interpretation. *Applied Optics* 19, 3428–3430.
- Gordon, H. R. and D. K. Clark (1981). Clear water radiances for atmospheric correction of Coastal Zone Color Scanner imagery. *Applied Optics* 20, 4175–4180.
- Gordon, H. R., D. K. Clark, J. W. Brown, O. B. Brown, R. H. Evans, and W. W. Broenkow (1983). Phytoplankton pigment concentrations in the middle atlantic bight: comparison between ship determinations and Coastal Zone Color Scanner estimates. *Applied Optics* 22, 20–36.
- Gordon, H. R., D. K. Clark, J. L. Mueller, and W. A. Hovis (1980). Phytoplankton pigments derived from the nimbus-7 czcs: Initial comparisons with surface measurements. *Science* 210, 63–66.
- Gordon, H. R. and K. Ding (1992). Self-shading of in-water optical instruments. *Limnology and Oceanography* 37, 491–500.
- Gordon, H. R., T. Du, and T. Zhang (1997a). Atmospheric correction of ocean color sensors: Analysis of the effects of residual instrument polarization sensitivity. *Applied Optics* 36, 6938–6948.
- Gordon, H. R., T. Du, and T. Zhang (1997b). Remote sensing of ocean color and aerosol properties: Resolving the issue of aerosol absorption. *Applied Optics* 36, 8670–8684.

- Gordon, H. R. and B. A. Franz (2008). Remote sensing of ocean color: Assessment of the water-leaving radiance bidirectional effects on the atmospheric diffuse transmittance for SeaWiFS and MODIS intercomparisons. *Remote Sensing of Environment* 112, 2677–2685.
- Gordon, H. R., M. R. Lewis, S. D. McLean, M. S. Twardowski, S. A. Freeman, K. J. Voss, and C. G. Boynton (2009). Spectra of particulate backscattering in natural waters. *Optics Express* 17, 16192–16208.
- Gordon, H. R. and W. R. McCluney (1975). Estimation of the depth of sunlight penetration in the sea for remote sensing. *Applied Optics* 14, 413–416.
- Gordon, H. R. and A. Y. Morel (1983). *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*. New York, NY: Springer-Verlag.
- Gordon, H. R., T. J. Smyth, W. M. Balch, C. G. Boynton, and G. A. Tarran (2009). Light scattering by coccoliths detached from *Emiliania huxleyi*. *Applied Optics* 48, 6059–6073.
- Gordon, H. R., K. J. Voss, and K. A. Kilpatrick (1993). Angular distribution of fluorescence from phytoplankton. *Limnology and Oceanography* 38, 1582–1586.
- Gordon, H. R. and M. Wang (1992). Surface roughness considerations for atmospheric correction of ocean color sensors. 1: The Rayleigh scattering component. *Applied Optics* 31, 4247–4260.
- Gordon, H. R. and M. Wang (1994). Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. *Applied Optics* 33, 443–452.
- Gordon, H. R. and X. Xu (1996). Marine asymptotic daylight field: Effects of inelastic processes. *Applied Optics* 35, 4194–4205.
- Gordon, H. R. and T. Zhang (1996). How well can radiance reflected from the ocean-atmosphere system be predicted from measurements at the sea surface? *Applied Optics* 35, 6527–6543.
- Gordon, H. R., T. Zhang, F. He, and K. Ding (1996). Effects of stratospheric aerosols and thin cirrus clouds on atmospheric correction of ocean color imagery: Simulations. *Applied Optics* 36, 682–697.
- Gower, J. F. R. (Ed.) (1981). *Oceanography from Space*. New York NY: Plenum.
- Griffith, D. J. (1989). *Introduction to Electrodynamics (2nd Ed.)*. Englewood Cliffs, NJ: Prentice Hall.

- Hale, G. M. and M. R. Querry (1973). Optical constants of water in the 200-nm to 200- μ m wavelength region. *Applied Optics* 12, 555–563.
- Hansen, J. E. and L. D. Travis (1974). Light scattering in planetary atmospheres. *Space Science Reviews* 16, 527–610.
- Hapke, B. (1981). Bidirectional reflectance spectroscopy 1: Theory. *Journal of Geophysical Research* 86(B4), 3039–3054.
- Hapke, B. (1993). *Theory of Diffuse Reflectance and Emittance Spectroscopy*. Cambridge: Cambridge University Press.
- Holben, B. N., T. F. Eck, I. Slutsker, D. Tanre, J. P. Buis, A. Setzer, E. Vermote, J. Reagan, Y. Kaufman, T. Nakajima, F. Lavenue, I. Jankowiak, and A. Smirnov (1998). AERONET: A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment* 66, 1–16.
- Hooker, S. B., W. E. Esaias, G. C. Feldman, W. W. Gregg, and C. R. McClain (1992). An overview of SeaWiFS and ocean color. SeaWiFS Technical Report Series: Volume 1, NASA, Greenbelt, MD, Technical Memorandum 104566.
- Hooker, S. B. and C. R. McClain (2000). The calibration and validation of SeaWiFS data. *Progress in Oceanography* 45, 427–465.
- Hu, C., Z. Lee, and B. Franz (2012). Chlorophyll *a* algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research* 117(C), 01011.
- Hwang, P. A. and O. H. Shemdin (1988). The dependence of sea surface slope on atmospheric stability and swell conditions. *Journal of Geophysical Research* 93(C), 13903–13912.
- IOCCG (2013). In-flight calibration of satellite ocean-colour sensors. Reports of the International Ocean-Colour Coordinating Group, No. 14, Ed. R. Frouin.
- Irvine, W. M. (1976). Multiple scattering in planetary atmospheres. *Icarus* 25, 175–204.
- Jackson, J. D. (1975). *Classical Electrodynamics (2nd Ed.)*. New York, NY: Wiley.
- Jerlov, N. G. (1968). *Optical Oceanography*. New York, NY: Elsevier.
- Jerlov, N. G. (1975). *Marine Optics*. New York, NY: Elsevier.
- Johnson, B. C., H. Yoon, J. P. Rice, and A. C. Parr (2014). Principles of optical radiometry and measurement uncertainty. In G. Zibordi, C. J. Donlon, and A. C. Parr (Eds.), *Optical Radiometry for Ocean Climate Measurements*, pp. 13–67. New York, NY: Academic.

- Jonasz, M. and G. R. Fournier (2007). *Light Scattering by Particles in Water*. London: Academic Press.
- Junge, C. (1958). Atmospheric chemistry. In *Advances in Geophysics*, V4, pp. 1–108. London: Academic Press.
- Kenizys, F. X., E. P. Shettle, L. W. Abreu, G. P. Anderson, W. O. G. J. H. Chetwynd, J. E. A. Selby, and S. A. Clough (1988). Users guide to LOWTRAN 7. Air Force Geophysics Laboratory, Hanscomb AFB, MA 01731, AFGL-TR 88-0177.
- Kenizys, F. X., E. P. Shettle, W. O. Gallery, L. W. A. J. H. Chetwynd, J. E. A. Selby, S. A. Clough, and R. W. Fenn (1983). Atmospheric Transmittance/Radiance: The LOWTRAN 6 Model. Air Force Geophysics Laboratory, Hanscomb AFB, MA 01731, AFGL-TR-83-0187, NTIS AD A137796.
- Kieffer, H. H. and T. C. Stone (2005). The spectral irradiance of the Moon. *Astronomical Journal* 129, 2887–2901.
- Kieffer, H. H. and R. L. Wildey (1985). Absolute calibration of LANDSAT instruments viewing the Moon. *Photogrammetric Engineering and Remote Sensing* 51, 1391.
- King, M. D. and R. S. Fraser (1983). The effects of El Chichón on the intensity and polarization of skylight at Mauna Loa. In *Fifth Conference on Atmospheric Radiation*. American Meteorological Society, Boston, MA.
- Koepke, P. (1982). Vicarious satellite calibration in the Solar spectral range by means of calculated radiances and its application to Meteosat. *Applied Optics* 21, 2845–2854.
- Koepke, P. (1984). Effective reflectance of oceanic whitecaps. *Applied Optics* 23, 1816–1824.
- Kopelevich, O. V. (1983). Small-parameter model of optical properties of sea water. In A. S. Monin (Ed.), *Ocean Optics, vol 1: Physical Ocean Optics*. Moscow: Nauka Pub. (in Russian).
- Kopelevich, O. V. and E. M. Mezhericher (1983). Calculation of spectral characteristics of light scattering by sea water. *Izv. Akad. Nauk SSSR, Fiz. Atmos. Okeana*. 19(2), 144.
- Kuchinke, C. P., H. R. Gordon, and B. A. Franz (2009). Spectral optimization for constituent retrieval in case 2 waters I: Implementation and performance. *Remote Sensing of Environment* 113, 571–587.
- Kuchinke, C. P., H. R. Gordon, L. W. Harding, Jr., and K. J. Voss (2009). Spectral optimization for constituent retrieval in case 2 waters II: Validation study in the chesapeake bay. *Remote Sensing of Environment* 113, 610–621.

- Lane, A. P. and W. M. Irvine (1973). Monochromatic phase curves and albedos for the lunar disk. *Astronomical Journal* 78, 267–277.
- Lenoble, J. (1986). *Radiative Transfer in Scattering and Absorbing Atmospheres: Standard Computational Procedures*. A. Deepak.
- Loisel, H. and A. Morel (1998). Light scattering and chlorophyll concentration in case 1 waters: A reexamination. *Limnology and Oceanography* 43, 847–858.
- Marshall, B. R. and R. C. Smith (1990). Raman scattering and in-water ocean optical properties. *Applied Optics* 29, 71–84.
- Martiny, N., R. Frouin, and R. Santer (2005). Radiometric calibration of seawifs in the near infrared. *Applied Optics* 44, 7828–7844.
- Maul, G. A. and H. R. Gordon (1975). On the use of the Earth Resources Technology Satellite (LANDSAT-1) in optical oceanography. *Remote Sensing of Environment* 4, 95–128.
- McClain, C. R., G. Meister, and B. Monosmith (2014). Satellite ocean color sensor design concepts and performance requirements. In G. Zibordi, C. J. Donlon, and A. C. Parr (Eds.), *Optical Radiometry for Ocean Climate Measurements*, pp. 73–119. New York, NY: Academic Press.
- McCormick, N. J. (1996). Analytical transport theory applications in optical oceanography. *Annals of Nuclear Energy* 23, 381–395.
- Meister, G., E. J. Kwiatkowska, B. A. Franz, F. S. Platt, G. C. Feldman, and C. R. McClain (2005). Moderate-Resolution Imaging Spectroradiometer ocean color polarization correction. *Applied Optics* 44, 5524–5535.
- Minnaert, M. (1962). Photometry of the moon. In G. P. Kuiper and B. M. Middlehurst (Eds.), *Planets and Satellites: The Solar System (Volume III)*, pp. 213–248. Chicago, IL: University of Chicago Press.
- Mishchenko, M. I. (2008). Multiple scattering, radiative transfer, and weak localization in discrete random media: Unified microphysical approach. *Reviews of Geophysics* 46, 1–33.
- Mishchenko, M. I. (2013). 125 years of radiative transfer: Enduring triumphs and persisting misconceptions. AIP Conf. Proc. 1531, 11–18. doi 10.1063/1.4804696.
- Mishchenko, M. I. (2014). Directional radiometry and radiative transfer: The convoluted path from centuries-old phenomenology to physical optics. *Journal of Quantitative Spectroscopy and Radiative Transfer* 146, 4–33.

- Mishchenko, M. I., J. M. Dlugach, M. A. Yurkin, L. Bi, B. Cairns, L. Liu, R. L. Panetta, L. D. Travis, P. Yang, and N. Zakharova (2016). First principles modeling of electromagnetic scattering by discrete and discretely heterogeneous random media. *Physics Reports* 632, 1–75.
- Mishchenko, M. I., J. Hovenier, and L. D. Travis (2000). *Light Scattering by Nonspherical Particles*. San Diego, CA: Academic Press.
- Mishchenko, M. I. and L. Travis (1994). Light scattering by polydispersions of randomly oriented spheroids with sizes comparable to wavelengths of observation. *Applied Optics* 33, 7206–7225.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis (2002a). *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge, UK: Cambridge University Press.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis (2002b). *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge, UK: Cambridge University Press.
- Mitchell, B. G. and D. A. Kiefer (1984). Determination of absorption and fluorescence excitation spectra for phytoplankton. In O. Holm-Hansen, L. Bolis, and R. Giles (Eds.), *Marine Phytoplankton and Productivity*, pp. 157–169. New York, NY: Springer-Verlag.
- Mitchell, B. G. and D. A. Kiefer (1988). Chlorophyll *a* specific absorption and fluorescence excitation spectra for light limited phytoplankton. *Deep-Sea Research* 35, 639–664.
- Mobley, C. D. (1989). A numerical model for the computation of radiance distributions in natural waters with wind-roughened surfaces. *Limnology and Oceanography* 34, 1473–1483.
- Mobley, C. D. (1994). *Light and Water; Radiative Transfer in Natural Waters*. New York, NY: Academic Press.
- Mobley, C. D. (2016). Polarized reflectance and transmittance properties of windblown sea surfaces. *Applied Optics* 54, 4828–4849.
- Mobley, C. D. (2018). Chapter 1: Invariant imbedding theory for the vector radiative transfer equation. In A. Kokhanovsky (Ed.), *Springer Series in Light Scattering, Volume 1, Multiple Light Scattering, Radiative Transfer and Remote Sensing*.
- Mobley, C. D., B. Gentili, H. R. Gordon, Z. Jin, G. W. Kattawar, A. Morel, P. Reinersman, K. Stamnes, and R. H. Stavn (1993). Comparison of numerical models for computing underwater light fields. *Applied Optics* 32, 7484–7504.
- Monahan, E. C. (1969). Fresh water whitecaps. *Journal of Atmospheric Science* 26, 1026–1029.

- Monahan, E. C. (1971). Oceanic whitecaps. *Journal of Physical Oceanography* 1, 139–144.
- Monahan, E. C. and I. G. O’Muircheartaigh (1980). Optimal power-law description of oceanic whitecap coverage dependence on wind speed. *Journal of Physical Oceanography* 10, 2094–2099.
- Monahan, E. C. and I. G. O’Muircheartaigh (1986). Whitecaps and the passive remote sensing of the ocean surface. *International Journal of Remote Sensing* 7, 627–642.
- Moore, K. D., K. J. Voss, and H. R. Gordon (1998). Spectral reflectance of whitecaps: Instrumentation, calibration, and performance in coastal waters. *Journal of Atmospheric and Oceanic Technology* 15, 496–509.
- Moore, K. D., K. J. Voss, and H. R. Gordon (2000). Spectral reflectance of whitecaps: Their contribution to water-leaving radiance. *Journal of Geophysical Research* 105C, 6493–6499.
- Morel, A. (1966). Etude experimentale de diffusion de la lumiere par l’eau, les solutions de chlorure de sodium et l’eau de mer optiquement pures. *Journal de Chimie et Physique* 10, 1359–1366.
- Morel, A. (1968). Note au sujet des constants de diffusion de la lumiere par l’eau et l’eau de mer optiquement pures. *Cahiers Oceanographiques* 20, 157–162.
- Morel, A. (1974). Optical properties of pure water and pure sea water. In N. G. Jerlov and E. S. Nielsen (Eds.), *Optical Aspects of Oceanography*. New York, NY: Academic Press.
- Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (Case I waters). *Journal of Geophysical Research* 93(C), 10749–10768.
- Morel, A., D. Antoine, and B. Gentili (2002). Bidirectional reflectance of oceanic waters: Accounting for Raman emission and varying particle scattering phase function. *Applied Optics* 41, 6289–6306.
- Morel, A. and A. Bricaud (1981). Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep-Sea Research* 28(A), 1375–1393.
- Morel, A. and B. Gentili (1996). Diffuse reflectance of oceanic waters. III. Implication of bidirectionality for the remote sensing problem. *Applied Optics* 35, 4850–4862.
- Morel, A., B. Gentili, H. Claustre, M. Babin, A. Bricaud, J. Ras, and F. Tieche (2007). Optical properties of the “clearest” natural waters. *Limnology and Oceanography* 52, 217–229.

- Morel, A. and H. R. Gordon (1980). Report of the Working Group on Ocean Color. *Boundary-Layer Meteorology* 18, 343–355.
- Morel, A. and L. Prieur (1975). Analyse spectrale des coefficients d'atténuation diffuse, de retrodiffusion pour diverses régions marines. Centre Rech. Oceanogr. de Villefranch-sur-Mer, Final Report Contract No. M0-A01-78-00-4092.
- Morel, A. and L. Prieur (1977). Analysis of variations in ocean color. *Limnology and Oceanography* 22, 709–722.
- Moulin, C., H. R. Gordon, V. F. Banzon, and R. H. Evans (2001). Assessment of Saharan dust absorption in the visible from SeaWiFS imagery. *Journal of Geophysical Research* 106(D), 18239–18250.
- Moulin, C., H. R. Gordon, R. M. Chomko, V. F. Banzon, and R. H. Evans (2001). Atmospheric correction of ocean color imagery through thick layers of Saharan dust. *Geophysical Research Letters* 28, 5–8.
- Mueller, J. L., G. S. Fargion, and C. R. McClain (Eds.) (2003a). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation Vol. I: Introduction, Background and Conventions*. NASA Goddard Space flight Center, Technical Memorandum 2003-211621/Rev-Vol I Revision 4.
- Mueller, J. L., G. S. Fargion, and C. R. McClain (Eds.) (2003b). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation Vol. II: Instrument Specifications, Characterization and Calibration*. NASA Goddard Space flight Center, Technical Memorandum 2003-211621/Rev-Vol II Revision 4.
- Mueller, J. L., G. S. Fargion, and C. R. McClain (Eds.) (2003c). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation Vol. III: Radiometric Measurements and Data Analysis Protocols*. NASA Goddard Space flight Center, Technical Memorandum 2003-211621/Rev-Vol III Revision 4.
- Mueller, J. L., G. S. Fargion, and C. R. McClain (Eds.) (2003d). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation Vol. IV: Instruments, Characterizations, Field Measurements and Data Analysis*. NASA Goddard Space flight Center, Technical Memorandum 2003-211621/Rev-Vol IV Revision 4.
- Mueller, J. L., G. S. Fargion, and C. R. McClain (Eds.) (2003e). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation Vol. V: Biogeochemical and Bio-Optical Measurements and Data Analysis Protocols*. NASA Goddard Space flight Center, Technical Memorandum 2003-Rev-Vol V Revision 4.

- Mueller, J. L., G. S. Fargion, and C. R. McClain (Eds.) (2003f). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation Vol. VI: Special Topics in Ocean Optics Protocols and Appendices*. NASA Goddard Space flight Center, Technical Memorandum 2003-211621/Rev-Vol VI Revision 4.
- Muller, M. (1995). Equation of time - problem in astronomy. *Acta Physica Polonica A, Supplement, S-49* 88.
- Nakajima, T. and M. Tanaka (1983). Effect of wind-generated waves on the transfer of solar radiation in the atmosphere-ocean system. *Journal of Quantitative Spectroscopy and Radiative Transfer* 29, 521–537.
- Nearing, J. C. (2010). *Mathematical Tools for Physics*. Mineola, NY: Dover Publications.
- O'Reilly, J. E., S. Maritorena, D. A. Siegel, M. C. O'Brien, D. Toole, B. G. Mitchell, M. Kahru, F. P. Chavez, P. Strutton, G. F. Cota, S. B. Hooker, C. R. McClain, K. L. Carder, F. Muller-Karger, L. Harding, A. Magnuson, D. Phinney, G. F. Moore, J. Aiken, K. R. Arrigo, R. Letelier, and M. Culver (2000). Ocean color chlorophyll *a* algorithms for SeaWiFS, OC2, and OC4: Version 4. NASA Goddard Space Flight Center, Greenbelt, MD, Technical Memorandum 206892.
- O'Reilly, J. E., S. Torena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research* 103(C), 24937–24953.
- Panofsky, W. K. H. and M. Phillips (1964). *Classical Electricity and Magnetism (2nd Ed.)*. Reading, MA: Addison-Wesley.
- Patterson, E. M. (1981). Optical properties of crustal aerosol: Relation to chemical and physical characteristics. *Journal of Geophysical Research* 86(C), 3236–3246.
- Pegau, W. S., D. Gray, and J. R. V. Zaneveld (1997). Absorption and attenuation of visible and near-infrared light in water: Dependence on temperature and salinity. *Applied Optics* 36, 6035–6046.
- Petzold, T. J. (1972). Volume scattering functions for selected natural waters. Scripps Institution of Oceanography, Visibility Laboratory, San Diego, CA, SIO Ref. 72-78.
- Pierce, A. K., R. R. McMath, L. Goldberg, and O. C. Mohler (1950). Observations of Solar limb darkening between 0.5 and 10.2 μm . *Astrophysical Journal* 112, 289–298.
- Placzek, G. (1962). *Rayleigh and Raman Scattering (UCRL-Translation)*. Washington, D.C: U.S. Atomic Energy Commission.
- Plass, G. N., G. W. Kattawar, and F. E. Catchings (1973). Matrix operator theory of radiative transfer 1: Rayleigh scattering. *Applied Optics* 12, 314–328.

- Pope, R. M. and E. S. Fry (1997). Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements. *Applied Optics* 36, 8710–8723.
- Potter, J. F. (1969). The delta function approximation in radiative transfer theory. *Journal of Atmospheric Science* 27, 943–949.
- Preisendorfer, R. W. (1961). Application of radiative transfer theory to light measurements in the sea. *Union Géodésique et Géophysique Internationale* 10, 11–30.
- Preisendorfer, R. W. (1965). *Radiative Transfer on Discrete Spaces*. Oxford: Pergamon.
- Preisendorfer, R. W. (1976a). *Hydrologic Optics V. 1: Introduction*. Springfield: National Technical Information Service PB-259 793/8ST.
- Preisendorfer, R. W. (1976b). *Hydrologic Optics V. 2: Foundations*. Springfield: National Technical Information Service PB-259 794/6ST.
- Preisendorfer, R. W. (1976c). *Hydrologic Optics V. 3: Solutions*. Springfield: National Technical Information Service PB-259 795/3ST.
- Preisendorfer, R. W. (1976d). *Hydrologic Optics V. 4: Imbeddings*. Springfield: National Technical Information Service PB-259 796/1ST.
- Preisendorfer, R. W. (1976e). *Hydrologic Optics V. 5: Properties*. Springfield: National Technical Information Service PB-259 797/9ST.
- Preisendorfer, R. W. (1976f). *Hydrologic Optics V. 6: Surfaces*. Springfield: National Technical Information Service PB-268 704/4ST.
- Preisendorfer, R. W. and C. D. Mobley (1986). Albedos and glitter patterns of a wind-roughened sea surface. *Journal of Physical Oceanography* 16, 1293–1316.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992). *Numerical Recipes in FORTRAN*. Cambridge: Cambridge University Press.
- Purcell, E. M. and C. R. Pennypacker (1973). Scattering and absorption of light by non-spherical dielectric grains. *Astrophysical Journal* 186, 705–714.
- Raman, C. V. (1921). The colour of the sea. *Nature (London)* 108, 367.
- Rayleigh, L. (1910). Colours of the sea and sky. *Nature (London)* 83, 48–50.
- Rougier, M. G. (1934). *L'Astronomie* 48, 224.
- Sathyendranath, S. and T. Platt (1989). Remote sensing of ocean chlorophyll: Consequences of a nonuniform pigment profile. *Applied Optics* 28, 490–495.

- Sathyendranath, S. and T. Platt (1998). Ocean-color model incorporating transspectral processes. *Applied Optics* 37, 2216–2227.
- Saunders, P. M. (1967). Shadowing on the ocean and the existence of the horizon. *Journal of Geophysical Research* 72, 4643–4649.
- Schneider, W., G. K. Moortgat, J. P. Burrows, and G. S. Tyndall (1987). Absorption cross-sections of NO₂ in the UV and visible region (200 - 700 nm) at 298 K. *Journal of Photochemistry and Photobiology* 40, 195–217.
- Shaw, J. A. and J. H. Churnside (1997). Scanning-laser glint measurements of sea-surface slope statistics. *Applied Optics* 36, 4202–4213.
- Shettle, E. P. (1984). Optical and radiative properties of a desert dust model. In G. Fiocco (Ed.), *Proceedings of the Symposium on Radiation in the Atmosphere*, pp. 74–77. Hampton, VA: A. Deepak.
- Shettle, E. P. and R. W. Fenn (1979). Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties. Air Force Geophysics Laboratory, Hanscomb AFB, MA 01731, AFGL-TR-79-0214.
- Shibata, K. (1958). Spectrophotometry of intact biological materials. *Journal of Biochemistry (Tokyo)* 45, 599–604.
- Siegel, D. A., M. Wang, S. Maritorena, and W. Robinson (2000). Atmospheric correction of satellite ocean color imagery: the black pixel assumption. *Applied Optics* 39, 3582–3591.
- Simmons, J. W. and M. J. Guttman (1970). *States Waves and Photons: A Modern Introduction to Light*. Reading, MA: Addison-Wesley.
- Slater, P. N., S. F. Biggar, R. G. Holm, R. D. Jackson, Y. Mao, M. S. Moran, J. M. Palmer, and B. Yuan (1987). Reflectance- and radiance-based methods for the in-flight absolute calibration of multispectral sensors. *Remote Sensing of Environment* 22, 11–37.
- Smirnov, A., B. N. Holben, O. Dubovik, R. Frouin, T. F. Eck, and I. Slutsker (2003). Maritime component in aerosol optical models derived from Aerosol Robotic Network data. *Journal of Geophysical Research* 108(D1), 4033.
- Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker (2003). Effect of wind speed on columnar aerosol optical properties at Midway Island. *Journal of Geophysical Research* 108(D24), D4802.
- Smirnov, A., B. N. Holben, I. Slutsker, D. M. Giles, C. R. McClain, T. F. Eck, S. M. Sakerin, A. Macke, P. Croot, G. Zibordi, P. K. Quinn, J. Sciare, S. Kinne, M. Harvey, T. J. Smyth, S. Piketh, T. Zielinski, A. Proshutinsky, J. I. Goes, N. B. Nelson, P. Larouche, V. F.

- Ra-dionov, P. Goloub, K. K. Moorthy, R. Matarrese, E. J. Robertson, and F. Jourdin (2009). Maritime Aerosol Network as a component of Aerosol Robotic Network. *Journal of Geophysical Research* 114(D06), 204.
- Sogandares, F. M. and E. S. Fry (1997). Absorption spectrum (340-640 nm) of pure water. I. photothermal measurements. *Applied Optics* 36, 8699–8709.
- Spanier, J. and E. M. Gelbard (1969). *Monte Carlo principles and neutron transport problems*. Reading, MA: Addison-Wesley.
- Spencer, J. W. (1971). Fourier series representation of the position of the Sun. *Search* 2, 172.
- Stabeno, P. J. and E. C. Monahan (1986). The influence of whitecaps on the albedo of the sea surface. In E. C. Monahan and G. M. Niocaill (Eds.), *Oceanic Whitecaps*, pp. 261–266. Dordrecht: Reidel.
- Stamnes, K., S.-C. Tsay, W. Wiscombe, and K. Jayaweera (1988). Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics* 27, 2502–2509.
- Stavn, R. H. and A. D. Weidemann (1988). Optical modeling of clear ocean light fields: Raman scattering effects. *Applied Optics* 27, 4002–4011.
- Stramski, D., E. Boss, D. Bogucki, and K. J. Voss (2004). The role of seawater constituents in light backscattering in the ocean. *Progress in Oceanography* 61, 27–56.
- Sugihara, S., M. Kishino, and N. Okami (1984). Contribution of Raman scattering to upward irradiance in the sea. *Journal of the Oceanographical Society of Japan* 40, 397–404.
- Sullivan, J. M. and M. S. Twardowski (2009). Angular shape of the oceanic particulate volume scattering function in the backward direction. *Applied Optics* 48, 6811–6819.
- Sytinskaya, N. N. (1953). *Russian Astronomical Journal* 30, 295.
- Takano, Y. and K. N. Liou (1989). Solar radiative transfer in cirrus clouds Part I: Single-scattering and optical properties of hexagonal ice crystals. *Journal of Atmospheric Science* 46, 224–240.
- Tyler, J. E. and R. C. Smith (1970). *Measurements of Spectral Irradiance Underwater*. New York, NY: Gordon and Breach.
- van de Hulst, H. C. (1957). *Light Scattering by Small Particles*. New York, NY: Wiley.
- van de Hulst, H. C. (1980). *Multiple Light Scattering*. New York, NY: Academic.

- Viollier, M. (1982). Radiance calibration of the Coastal Zone Color Scanner: A proposed adjustment. *Applied Optics* 21, 1142–1145.
- Viollier, M., D. Tanre, and P. Y. Deschamps (1980). An algorithm for remote sensing of water color from space. *Boundary-Layer Meteorology* 18, 247–267.
- Volten, H., J. F. de Haan, J. W. Hovenier, W. V. R. Schreurs, A. G. Dekker, H. J. Hoogenboom, F. Charlton, and R. Wouts (1998). Laboratory measurements of angular distributions of light scattered by phytoplankton and silt. *Limnology and Oceanography* 43(6), 1180–1197.
- Voss, K. J. and E. S. Fry (1984). Measurement of the mueller matrix for ocean water. *Applied Optics* 23, 4427–4439.
- Voss, K. J., H. R. Gordon, S. Flora, B. C. Johnson, M. Yarbrough, M. Feinholtz, and A. T. Houlihan (2017). Method to extrapolate the upwelling radiance attenuation coefficient to the surface as applied to the Marine Optical Buoy (MOBY). *Journal of Atmospheric and Oceanic Technology* 34, 1423–1432.
- Walker, R. E. (1994). *Marine Light Field Statistics*. New York, NY: Wiley.
- Wang, M., B. A. Franz, R. A. Barnes, and C. R. McClain (2001). Effects of spectral bandpass on SeaWiFS-retrieved near-surface optical properties of the ocean. *Applied Optics* 40, 343–348.
- Wang, M. and H. R. Gordon (1993). Retrieval of the columnar aerosol phase function and single scattering albedo from sky radiance over the ocean: Simulations. *Applied Optics* 32, 4598–4609.
- Wang, M. and H. R. Gordon (1994). A simple, moderately accurate, atmospheric correction algorithm for SeaWiFS. *Remote Sensing of Environment* 50, 231–239.
- Wang, M. and H. R. Gordon (2002). Calibration of ocean color scanners: How much error is acceptable in the near infrared? *Remote Sensing of Environment* 82, 497–504.
- Wang, M. and W. Shi (2007). The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. *Optics Express* 15(24), 15722–15733.
- Wang, M., S. Son, and W. Shi (2009). Evaluation of MODIS SWIR and NIR-SWIR atmospheric correction algorithm using SeaBASS data. *Remote Sensing of Environment* 113(3), 635–644.
- Wehrli, C. (1985). Extraterrestrial Solar Spectrum. Publication no. 615, Physikalisch-Meteorologisches Observatorium + World Radiation Center (PMO/WRC) Davos Dorf, Switzerland.

- Whitlock, C. H., D. S. Bartlett, and E. A. Gurganus (1982). Sea foam reflectance and influence on optimum wavelength for remote sensing of ocean aerosols. *Geophysical Research Letters* 7, 719–722.
- Wilson, Jr., E. B., J. C. Decius, and P. C. Cross (1980). *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. Mineola, NY: Dover Publications.
- Xiong, J., G. Toller, J. Sun, B. Wenny, A. Angal, and W. Barnes (2013). MODIS Level 1B Algorithm Theoretical Basis Document (V. 4). NASA Goddard Space Flight Center.
- Yang, H. and H. R. Gordon (1997). Remote sensing of ocean color: Assessment of the water-leaving radiance bidirectional effects on the atmospheric diffuse transmittance. *Applied Optics* 36, 7887–7897.
- Yang, H., H. R. Gordon, and T. Zhang (1995). Island perturbation to the sky radiance over the ocean: Simulations. *Applied Optics* 34, 8354–8362.
- Yariv, A. (1988). *Quantum Electronics*. New York, NY: Wiley.
- Yentsch, C. S. (1962). Measurement of visible light absorption by particulate matter in the ocean. *Limnology and Oceanography* 7, 207–217.
- Zaneveld, J. R. V. (1982). Remotely sensed reflectance and its dependence on vertical structure: A theoretical derivation. *Applied Optics* 21, 4146–4150.
- Zaneveld, J. R. V. (1995). A theoretical derivation of the dependence of the remotely sensed reflectance of the ocean on the inherent optical properties. *Journal of Geophysical Research* 100(C7), 13135–13142.
- Zaneveld, J. R. V., A. H. Barnard, and E. Boss (2005). Theoretical derivation of the depth average of remotely sensed optical parameters. *Optics Express* 13, 9052–9061.
- Zaneveld, J. R. V., J. C. Kitchen, A. Bricaud, and C. Moore (1992). Analysis of *insitu* spectral absorption meter data. *Society of Photo-Optical Instrumentation Engineers, Ocean Optics XI 1750*, 187–200.
- Zhang, J., H. Lu, and L. J. Wang (2008). Precision refractive index measurements of air N₂, O₂, Ar, and CO₂ with a frequency comb. *Applied Optics* 17, 3143–3151.
- Zhang, T. and H. R. Gordon (1997). Retrieval of elements of the columnar aerosol scattering phase matrix from polarized sky radiance over the ocean: Simulations. *Applied Optics* 36, 7948–7959.
- Zhang, X. and L. Hu (2009). Estimating scattering of pure water from density fluctuation of the refractive index. *Optics Express* 17, 1671–1678.

- Zhang, X., L. Hu, and M.-X. He (2009). Scattering by pure seawater: Effect of salinity. *Optics Express* 17, 5698–5710.
- Zibordi, G., B. Holben, I. Slutsker, D. Giles, D. D’Alimonte, F. Mélin, J.-F. Berthon, D. Vandemark, H. Feng, G. Schuster, B. Fabbri, S. Kaitala, and J. Seppälä (2009). AERONET-OC: A network for the validation of ocean color primary radiometric products. *Journal of Atmospheric and Oceanic Technology* 26, 1634–1651.
- Zibordi, G., F. Mélin, and J.-F. Berthon (2006). Comparison of SeaWiFS, MODIS, and MERIS radiometric products at a coastal site. *Geophysical Research Letters* 33, L06617.
- Zibordi, G. and K. J. Voss (2014). In situ optical radiometry in the visible and near infrared. In G. Zibordi, C. J. Donlon, and A. C. Parr (Eds.), *Optical Radiometry for Ocean Climate Measurements*, pp. 247–304. New York, NY: Academic Press.

Index

- H*-functions, 517
- δ -Isotropic approximation
 - modified (MDIA), 517
- absorbing aerosols, 729
 - atmospheric correction for, 731–741
 - effect of vertical distribution, 731
 - the spectral matching algorithm for, 734
- absorptance, 321
- absorption
 - NO₂, 328
 - atmospheric windows in, 328
 - by aerosols, 714, 729
 - by atmospheric gases, 315
 - by CDOM, 353
 - by cylinders, 375
 - by detrital materials, 369
 - by non-algal particles, 353
 - by particles in natural waters, 353
 - by phytoplankton, 353
 - coefficient, 338
 - coefficient of water, 340
 - coefficient, definition of, 118
 - coefficient, intracellular, 365
 - coefficient, relation to refractive index, 47, 365
 - coefficient, specific, 365
 - cross section, 221, 318
 - efficiency, 234
 - index, 739
 - lines, 317
 - mean efficiency, 383
 - mean free path, 936
 - molecular oxygen, 326
 - optical depth, 318
 - optical thickness, 318, 366
 - Ozone, 328, 727
 - Rayleigh-Gans efficiency, 244
 - specific coefficient, 366, 384
 - temperature effects in water, 341
 - vertical distribution in atmosphere, 323
 - water vapor, 327
- AERONET, 305, 706, 716
- aerosol
 - vertical structure, 313
- aerosol models, 295
 - Bermuda, 305
 - candidates for atmospheric correction, 719
 - cirrus clouds, 311
 - Coastal, 300
 - El Chichon, 311
 - Haze C, 296, 714
 - Kaashidoo, 305
 - Lanai, 305
 - Maritime, 300, 717
 - Oceanic, 300
 - physical models, 295–312
 - Saharan dust, 308, 736
 - Shettle and Fenn, 713
 - Smirnov et al., 305
 - stratospheric, 310
 - Tropospheric, 300, 717
 - Urban, 300
 - volcanic, 311
 - windblown dust, 736
- aerosol optical depth

- estimation from ocean color data, 720
- albedo, 113, 671
 - disk equivalent of Moon, 634
- almucantar, 824
- analemma, 643
- ancillary data, 706
- Angstrom's law, 297
- angular field of view, 784
- anomalous diffraction, 391
- apparent optical properties
 - examples from field measurements, 443
 - influence of inelastic processes on, 484–500
- apparent optical properties (AOPs), 435
 - definitions of, 438
 - examples from field measurements, 442
 - examples from simulations, 444–446
 - hybrid model of, 474
 - influence of fluorescence on, 487–490
 - influence of Raman scattering on, 490–500
- Arabian Sea, 740
- astronomical unit, 598, 626, 676, 861
- atmospheric correction, 655
- atmospheric windows, 329, 341
- attenuation
 - coefficient for radiance, 439
 - coefficient for vector irradiance, 445
 - coefficient, definition of, 118
 - coefficient, diffuse, 438
 - coefficient, relation to optical depth, 134
 - coefficients for irradiance, 438
 - cross section, 221
 - cross section in DDA, 257
 - efficiency, 234
 - length, 436
 - Rayleigh-Gans efficiency, 244
- average cosine, 445, 461
 - definition of, 440
- Babinet's principle, 252
- backscattering
 - coefficient, definition of, 358
 - by a coated sphere, 373
 - by cylinders, 378
 - coefficient of natural waters, 359, 360
 - coefficient, relation to Chlorophyll, 360
 - efficiency, definition of, 372
 - influence of the size distribution, 386
 - of spherical particles, 372
 - probability, 359
 - probability, relation to Chlorophyll, 359
- Balmer series, 631
- band head, 326
- band spectra, 325
- basis vectors, 930
- Bessel functions, 233
- bi-directional reflectance distribution function
 - BRDF, definition of, 170
- bio-optical algorithm, 509
- black body, 206, 631
 - radiation, 206
 - spectral radiance, 207
 - Wien's law, 207
- Bohr radius a_0 , 35
- Boltzmann constant, 284
- boundary conditions
 - in matrix operator technique, 159
- bowtie effect, 793
- BRDF, 170, 470, 517, 634
 - effect, 862
 - of reflectance plaque, 819
- BRDF effect, 598
- Brownian motion, 933
- cal/val, 866
- calculus, 897
 - fundamental theorem of, 902
 - multivariable, 909, 911
- calibration
 - constant, 132

- error, 811
- full system, 828
- radiance based, 817
- radiometric, 815
- reflectance based, 817, 818
- sphere, 818
- vicarious, 815, 833, 871
- vicarious(, 820
- candidate aerosol models, 719
- carotenoids, 775
- cascade impactor, 290
- Case 1, 509
- Case 1 waters, 443
 - definition of, 349
- Case 2 waters, 443, 729, 746
 - definition of, 349
- CDOM, 350
- Central Limit Theorem, 936, 941
- characteristic function, 936
- characterization, 833–841
- charge
 - bound, 38
 - density of, 23
 - free, 38
 - on electron, 35
- Chlorophyll, 485
- Chlorophyll *a*, 355, 705, 775, 859
 - fluorescence of, 775
- Chlorophyll *a*
 - absorption spectrum, 367
- cirrus, 311
- Clausius-Mossotti, 42
- clear-water radiance, 678, 869
- coccolith, 356
- coccolithophore, 350
- coccolithophore, 356
- columnar
 - aerosol phase function, 292
 - aerosol size distribution, 292
 - number density, 292
 - number distribution, 305
 - volume distribution, 305
- complex
 - algebra, 895
 - conjugate, 897
 - numbers, 895
 - plane, 896
- conservation
 - of energy, 26
- constrained linear inversion, 289
- continuity
 - equation of, 23
- cosine collector, 132
- cosines
 - law of, 925
- Coulter counter, 367, 376, 380
- Cox-Munk, 444, 460, 462
 - analysis of surface roughness, 575
 - surface slope variance, 444, 576
- critical angle, 58, 458
- cross section
 - for absorption, 221, 222
 - for attenuation, 221
 - for extinction, 221
 - for scattering, 118, 221
 - scattering, Rayleigh with anisotropy, 226
- current, 23
 - bound, 39
 - density of, 23
 - free, 39
- CZCS, 509, 706, 773, 788, 790
 - atmospheric correction for, 712
 - imagery, 712
 - verification of, 868
- declination, of the Sun, 627
- degree of polarization, 684
 - definition of, 86
 - of L_u , 472
 - of scattered light, 234
 - of scattering, effect of size distribution, 389

- depolarization factor
 - definition of, 225
 - of air, 285
- derivative
 - anti, 902
 - definition of, 898
 - directional, 913
 - partial, 912
- detrital material, 349
- dielectric, 35
 - linear-homogeneous-isotropic, 43
 - polarizability, 36
 - polarization of, 36
- dielectric constant
 - and susceptibility, 42
 - complex form of, 45
 - defined, 42
 - mean square fluctuation, 230
- differential equations
 - homogeneous, 905
 - inhomogeneous, 905
 - initial conditions, 908
 - second order, 905
 - steady state solution, 908
 - transient solution, 908
 - with constant coefficients, 905
- diffraction
 - contribution to scattering, 250–253
 - Huygens-Fresnel principle, 250
- diffuse attenuation coefficient, 438
- diffuse transmittance, 656, 663–667
 - rigorous computation of, 722–726
- dipole, 36
 - field of, 36
 - mean moment, 225
 - moment, 36
 - polarization of a dielectric sphere, 37
 - polarization of a medium, 37
- Dirac delta function
 - definition of, 902
 - in multiple dimensions, 917
- direct transmittance, 663
- direction cosines, 144
 - definition of, 952
- Discrete-Dipole Approximation (DDA), 253–257
- dispersion, 48
 - anomalous, 48
 - Kronig-Kramers relation, 369
 - relation, 369
- dissolved organic material, 350
- distribution function, 440
 - relation to average cosine, 440
- divergence theorem, 23
- Dobson Unit, 319
- Dobson units (DU), 727
- Doppler effect, 321
- E. Huxleyi, 350
- eccentricity, of Earth's orbit, 626
- ecliptic, 627, 637
- effective radius, 295
- effective variance, 295
- eigenvalues, 931
- eigenvectors, 931
- Einstein, 96, 441
- El Chichon, 311
- electric
 - susceptibility, 40
- electric field, 22
 - of a dipole, 34
 - of a spherical ball, 33
 - of a spherical cavity, 33
 - boundary conditions for, 52
 - Gauss' law for, 33
 - in matter, 32
 - of a dielectric sphere, 72
 - of a dipole (total), 64
 - of a moving charge, 60
 - of a point charge, 33
- electromagnetic field, 26
- electromagnetic radiation

- damping due to, 44
- detection of, 95, 96
- detectors, 96
- from accelerated charge, 60
- from oscillating dipole, 62
- scattered, 60
- spectrum of, 98
- electromagnetic wave, 26
 - absorption coefficient of, 47
 - analysis of polarization of, 75
 - analysis of polarization of, 92
 - boundary conditions for, 52
 - energy density, 29
 - from a dipole, 63
 - harmonic in time, 26
 - in a dielectric, 42, 46
 - linear polarization of, 75
 - one-dimensional, 27
 - plane, 28
 - polarization of, 54
 - Poynting vector, 30
 - radiation of, 62
 - reflected at a boundary, 54
 - reflection at a boundary, 51
 - scattering of, 64
 - speed of, in a dielectric, 43
 - speed of, in vacuum, 27
 - Stokes representation, 77
 - three-dimensional, 26
 - total internal reflection, 59
 - transmitted at a boundary, 54
 - transverse, 27
 - wave period, 28
 - wave speed, 27
 - wavelength, 28
 - wavelength in a dielectric, 48
 - wavenumber, 27
- emissivity, 211
 - directional spectral, 211
 - spectral, 211
- energy density
 - of electric field, 26
 - of magnetic field, 26
 - relation to scalar irradiance, 130
- equation of time, 629, 637–643
- equinox, 627
- equipartition theorem, 320
- equivalent homogeneous medium, 508
- Euler angles, 393
- euphotic depth, 441
- extinction
 - by cylinders, 376
 - cross section, 221
 - cross section in DDA, 257
 - efficiency, 234
 - mean efficiency, 383
 - Rayleigh-Gans efficiency, 244
- extraterrestrial solar irradiance
 - at mean Earth-Sun distance, 597
- first-order solution to RTE
 - derivation of, 174
- fluorescence, 443, 485
 - of Chlorophyll *a*, 775
 - quantum efficiency of, 489
- flux
 - definition of, 23
- footprint, 784
- Fourier series, 946
 - generalized, 949
 - relation to least squares, 950
- Fourier transform, 936
- Fournier-Forand phase function, 358, 390
- Fraunhofer lines, 631
- Fresnel
 - reflection coefficient, 57
 - transmission coefficient, 57
- full width at half maximum
 - FWHM, 319, 321
- Gaussian quadratures, 153
 - division points, 154

- weights, 154
- gelbstoff, 350
- Gershun tube, 104
- Gershun's equation, 171, 445
- Gershun's law
 - derivation of, 171
 - interpretation of, 173
- glitter pattern, 570
- global carbon cycle, 356, 705
- GMT, 629
- GOCI, 795
- Greenwich Mean Time, 629

- Hankel functions, 233
- Heisenberg uncertainty principle, 320
- Henye-Greenstein, 358
- Hermitian conjugate, 90
- hour angle, 628
- HPLC, 355
- hybrid model, 474
- HydroLight, 522
- hyperspectral, 789

- ideal gas, 230, 284
- IFOV, 791
 - instantaneous field of view, 784
- ill conditioned, 289
- immerison effect, 132
- incoherent, 99
- inelastic, 484
 - processes, 525
- inelastic processes, 436
- inelastic scattering, 485
- inherent optical properties (IOPs), 113, 436
 - effect of particle size distribution, 380
 - estimation from AOPs, 503–509
 - for Junge-distributed particles in water, 384
 - measurement of, 339
 - of natural waters, 337–364
 - of the atmosphere, 329
 - of water, 340–349
 - summability over components, 119, 338
- instantaneous field of view, 784
- integrating cavity absorption meter, 341
- integration
 - by parts, 904
 - definition of, 901
 - multivariable, 915
- integration time, 776, 785
- irradiance, 110
 - attenuation coefficient for, 438
 - attenuation coefficient for vector, 445
 - attenuation of PAR, K_{PAR} , 441
 - depth of penetration, 477–480
 - downwelling, 128, 437
 - extraterrestrial solar, 441, 630
 - in one-dimensional media, 128
 - mean extraterrestrial solar, 631
 - measurement of, 131
 - meter, 132
 - of the sun, 112
 - photosynthetically available (PAR), 441
 - plane, 128
 - quantum, 133, 440
 - ratio, 439
 - reflectance, 439
 - scalar, 129, 437
 - transmittance of, 597
 - upwelling, 128, 437
 - vector, 129, 437

- Jacobian determinant, 573
- Jacobian determinate, 919
- Junge, 868
- Junge distribution
 - for aerosols, 292
 - for particles in water, 381
 - particles in water, IOPs for, 384

- King factor, 285
- Kishino, 352

- Kronecker delta, 946
- L'hospital's rule, 901
- lambertian
 - emitter, 105
 - reflector, 165, 168, 170, 341
 - reflector, albedo of, 113
- Lanai, Hawaii, 442
- Laplacian operator, 914
- law of cosines, 925
- law of sines, 926
- least squares
 - backward, 928
 - Deming, 929
 - method of, 926
 - ordinary, 927
 - reduced major axis regression, 928
- Legendré polynomials, 153
- Legendre polynomials, 233
- Leibnitz integral rule, 918
- LIDAR, 313, 741
- limb darkening, 633
 - of the Sun, 112
- line width, 319
 - natural, 320
- linear independence, 905
- linear regression, 926
- LITE, 313, 741
- local property, 484
- Lorentz force, 22
- Lorentz-Lorenz, 42
- lunar
 - opposition effect, 651
 - surge, 651
- magnetic field, 22
 - boundary conditions for, 52
- marine boundary layer, 313
- MASCOT, 358
- matrix, 920
 - definition of, 920
 - diagonal, 921
 - eigenvalues of, 931
 - eigenvectors of, 931
 - inverse, 921
 - Mueller, 924
 - multiplication, 920
 - multiplication by a scalar, 922
 - orthogonal, 921
 - symmetric, 921
 - trace of, 922
- Mauna Loa, 311
- Maxwell equations, 22
 - linearity of, 24
- mean free path, 936
- meter-atmosphere, 318
- micron, 288
- Mie theory, 232
 - attenuation cross section, 234
 - attenuation efficiency, 234
 - scattering cross section, 234
 - scattering efficiency, 234
- MOBY, 874
- MODIS, 709, 712, 773, 788, 791, 819
- modulation, 786
- Monte Carlo, 185, 446
 - naive, 190
 - backward, 198
 - evaluation of integrals by, 940
 - for irradiance, 197
 - for wind-ruffled surface, 592–596
 - forward, 191
 - heuristic introduction, 139
 - importance sampling, 944
 - integration by, 939
 - variance reduction, 147
 - variance reduction in, 944
- Moon, 633
- Mueller matrix, 82
 - of ocean color sensor, 834
 - and coherency matrix, 89
 - combinations of optical elements, 83

- definition of, 82
 - for dielectric reflection, 89
 - for linear polarizer, 82
 - for Rayleigh scattering, 181
 - for scattering, 87
 - for scattering in Case 1 waters, 360
 - in terms of spin matrices, 91
 - non-commutative nature, 83
- $N[0, \sigma_N^2]$, 939
- $N[0, 1]$, 939
- natural fluorescence, 485
- Newton's law, 44
- Newton's laws, 784
- noise
 - induced, 779
 - intrinsic, 779
 - Johnson, 776
 - photon, 776
 - quantization, 781
 - shot, 776
- non-algal particles, 350
- normalized water-leaving radiance, 596–599, 675, 743
 - definition of, 597
- normalized water-leaving reflectance, 708, 743
 - definition of, 599
- OC4, 509
- omnidirectional wind, 594
 - definition of, 579
- optical depth
 - absorption, 318
 - definition of, 134
 - of Ozone, 727
 - of scattering, 717
 - Rayleigh scattering by air, 286
- optical theorem, 222, 268
 - derivation of, 264–270
- orbit
 - geosynchronous, 795
 - polar, 794
- orthogonality condition, 946
- oscillator strength, 50
- out-of-band
 - sensor response to, 836
- package effect, 367, 369
 - and detrital absorption, 370
 - and plankton absorption, 369
- PAR, 441
- particulate inorganic Carbon (PIC), 356
- particulate organic carbon (POC), 352, 356
- path length amplification factor, 352
- Pauli spin matrices, 89
- penetration depth, 477–480
- perihelion, definition of, 626
- permeability of free space, μ_0 , 24
- permittivity of free space, ϵ_0 , 24
- perpendicular plane, definition of, 671
- Petzmas, 358, 455, 471, 476, 726
- Petzold, 357, 444, 471
 - Phase-T, 357, 726
- phase angle, 634
- phase function
 - phase matrix — Rayleigh scattering, 226
 - for Rayleigh scattering, 134
 - for single particle, 374
 - Fourier expansion in azimuth of, 149
 - Fournier-Forand, 390
 - Henye-Greenstein, 358
 - influence of the size distribution, 389
 - Mueller matrix for Case 1 waters, 360
 - Mueller matrix for Raman scattering, 273
 - Mueller matrix for Rayleigh scattering, 181
 - of cylinders, 379
 - Petzmas, 358
 - Petzold Phase-T, 357
 - phase matrix – Rayleigh scattering, 284
 - Rayleigh scattering by air, 286
 - symmetries of, 149

- WetLab Average, 358
- Phase T, 357
- Phase-T, 444, 452, 456, 476, 506
- photocurrent, 97
- photosynthesis, 356, 440
- phototube, 96
- phytoplankton, 349, 485
- PIC, 356
- pigment concentration, 443, 705, 859
 - definition of, 355
- pixel, 783
 - variation in size, 785
- Planck function, 207
- Planck's constant, 50
- plane of incidence, 54
- POC, 356
 - relation to scattering coefficient, 356
- polarizability, 36
 - tensor, 933
- polarization
 - and incoherence, 86
 - degree of, 87, 184, 684
 - elliptical, 59, 77
 - left circular, 77
 - linear, 78
 - of a dielectric, 37
 - of an electromagnetic wave, 54
 - partial, 86
 - pass direction, 78
 - quarter-wave plate, 79
 - right circular, 77
 - sensor sensitivity to, 833
 - Stokes vector representation of, 80
 - unpolarized light, 83
- power-law, 292, 381
- Poynting's
 - theorem, 25
 - vector, 25
- primary production, 705
- principal plane, 823
- probability, 934
 - density of, 934
 - normalization of, 934
- probability density, 141
- pure rotational spectrum, 326
- Q-factor, 446, 507
 - definition of, 439
- QSSA
 - application to diffuse transmittance, 726
 - development of, 447–451
 - effect of atmosphere on, 462
 - effect of environment on, 457–467
 - effect of refracting interface on, 458
 - effect of scattering by water on, 455
 - hybrid model, 474
 - modeling with, 500
 - range of validity of, 451–455
 - vertical structure in IOPs, 480–484
- quantization, 781
- quantum efficiency of fluorescence, 489
- quantum irradiance, 440
- quasi-single scattering approximation (QSSA), 436, 447
- radiance, 102
 - across a refracting interface, 109
 - aperture, 106
 - apparent, 106
 - as a conserved quantity, 103
 - asymptotic distribution, 158, 525
 - attenuation coefficient for, 439
 - average cosine of, 440
 - band, 817
 - band integrated, 817
 - clear-water concept of, 678, 869
 - definition of, 102
 - distribution functions of, 440
 - field, 106
 - multidimensional integral for, 185
 - nadir viewing, 439
 - normalized water-leaving, 596–599, 743

- of a beam, 111
 - of a surface, 105
 - of the Moon, 113, 634
 - of the Sun, 112
 - path, 707
 - polarized, 108, 684
 - reflected by an interface, 568
 - reflected from a rough surface, 573
 - reflection of sky, 588
 - saturation, 781, 842
 - transmittance of sky, 589
 - transmitted by a rough surface, 582
 - transmitted through an interface, 568
 - water-leaving, 439, 509, 521, 656
- radiant emittance, 111
- radiant intensity, 100
- radiant power)spectral, 99
- radiative transfer equation, 120
 - complete, 123
 - derivation of, 120
 - expansion in azimuth, 151
 - in medium with varying refractive index, 125
 - in one-dimensional media, 127
 - including polarization, 123, 185
 - reflecting and transmitting boundaries, 164
 - solution methods, 135–163
 - superposition of solutions, 124
 - uniqueness of solutions, 124
 - with internal sources, 121
- radiometry, 100
- Raman scattering, 270–273, 442, 485
 - by water, 344–349
 - inelastic volume scattering function, 348
 - Mueller matrix for, 273
 - wavelength dependence in water, 348
- random numbers
 - pseudo, 141
 - U[0,1], 141
 - uniform distribution, 141, 260
- random variable, 934
 - unbiased* estimate of, 935
 - expectation of, 934
 - probability density of, 934
 - standard deviation of, 935
 - uniform distribution of, 941
 - variance of, 934, 935
- Rayleigh scattering, 222
 - by air, 286
 - Mueller matrix for, 222
- Rayleigh-Gans approximation, 238–244
- reciprocal centimeters, 325
- reciprocity principle, 125, 149, 677, 722
 - across a refracting interface, 586
 - in single scattering solution, 180
- Redfield ratios, 356
- reflectance
 - definition of, 170
 - diffuse, 518
 - normalized water-leaving, 510, 599, 708, 743
 - of lambertian reflector, 170
 - plaque, 818
 - relation to BRDF, 170
 - remote sensing, 439, 509, 599
- reflection
 - Brewster's angle for, 57
 - Fresnel equations, 57
 - Fresnel equations for, 57
 - function of a boundary, 164
 - law of, 55
 - law of in vector form, 570
 - of diffuse radiance by a rough surface, 585
 - of radiance from a rough surface, 572, 573
 - of sky radiance, 588
 - of Sun light from a rough surface, 574
 - total internal, 59
- refraction

- of diffuse radiance by a rough surface, 585
 - of radiance by a rough surface, 580
 - Snell's law in vector form, 580
- refractive index
 - complex form of, 46
 - defined, 43
- refractivity, 285
- regularization, 289
- remote sensing augmented reflectance
 - definition of, 601
- remote sensing problem, 656, 668
- remote sensing ratio, 439, 597
 - QSSA result for, 453
- remote sensing reflectance, 439, 462, 599
 - augmented by whitecaps, 601
- resistive pulse particle counter, 290, 367
- resolution
 - spatial, 784
- resonance, 908
- retarded time, 61
- rotation matrix, 680
- RTE solution methods, 135–163
 - adding, 160
 - boundary conditions, 135
 - discrete-ordinate, 154
 - doubling, 160
 - first-order solution, 174
 - first-order with polarization, 180
 - inclusion of a wind-ruffled surface, 590–596
 - invariant imbedding, 161, 478
 - matrix operator, 159
 - Monte Carlo, 139
 - single scattering solution, 174, 177
 - successive order of scattering, 135, 486
- Saharan dust
 - absorption of, 309
 - models, 308
 - size distribution of, 308
- sampling time, 785
- saturation radiance, 781
- scalar irradiances, 437
- scalars, 909
- scattering
 - albedo of single, 134
 - amplitude, 218
 - approximations for, 238–263
 - average number, 137
 - by a coated sphere, 373
 - by a homogeneous, 232
 - by a molecule, 67
 - by an atom, 64–72
 - by cylinders, 376
 - by irregularly-shaped particles, 253
 - by particles in natural waters, 354
 - by sea water, effect of salinity, 343
 - by small sphere, 72
 - by water, effect of temperature, 344
 - coefficient, 221
 - coefficient of, 115
 - coefficient, specific, 372
 - cross section, 221
 - depolarization factor for, 225
 - differential cross section of, 256
 - efficiency, 234
 - elastic, 118, 338
 - geometrical optics approximation of, 257–263
 - inelastic, 118, 338, 484
 - influence of particle shape, 375
 - influence of the size distribution, 386
 - mean efficiency, 383
 - Mie theory, 232
 - Mueller matrix for in Case 1 waters, 360
 - multiple, 826
 - multiple in atmospheric correction, 718–720
 - of water, 341
 - phase function, 134
 - phase function of, 134

- plane of, 64
- polarization of in Case 1 waters, 360
- Raman, 444
- Raman, by water, 344–349
- Rayleigh, 222
- Rayleigh by liquids, 226
- Rayleigh with molecular anisotropy, 226
- single scattering approximation, 177
- specific coefficient, 384
- table of for pure water, 342
- table of for sea water, 343
- volume scattering function (VSF), 113
- scattering coefficient, 221
 - definition of, 115
 - for atomic scattering, 117
 - for inelastic scattering, 119
 - measurement of in natural waters, 354
 - Rayleigh with molecular anisotropy, 225
 - relation to pigment concentration, 355
- scattering cross section
 - Rayleigh with molecular anisotropy, 223
- SeaWiFS, 712, 773, 788, 819
 - test of SMA with, 736
- selenographic, 634
- series
 - Fourier, 946
 - Taylor, 899
- seston, 355, 356
- shadowing factor, 585
- shape factor, 512
- Shettle and Fenn, 299
- sidereal day, 637
- sines
 - law of, 926
- single-scattering approximation, 655, 706
- single-scattering solution
 - accuracy of, 671
 - application in QSSA, 448
 - application to atmospheric correction, 710–713
 - derivation of, 177
- size distribution, 238
 - columnar volume, 305
 - influence on IOPs in water, 382
 - Junge, 292, 381
 - Junge distribution, IOPs for, 384
 - log normal, 293
 - log-normal, 381
 - modified gamma, 311
 - number, 292
 - of aerosols, 289
 - of volume, 292
 - power-law, 381
- Snell's law, 55
 - vector form of, 56
- SNR
 - quantization, 781
 - signal-to-noise ratio, 776
 - usable, 782
- solar constant, 631
- solar day
 - mean, 629
- solid angle, 100
- solstice, 627
- spatial frequency, 786
- spatial period, 786
- spatial resolution, 784
- spectral line broadening
 - Doppler broadening, 321
 - pressure broadening, 320
- spectral matching algorithm (SMA), 734–740
- spherical coordinates
 - definition of, 918
- spherical triangle, 924
 - law of cosines, 926
 - law of sines, 926
- standard lamp, 818
- Stokes
 - Mueller matrix, 82
 - rotation matrix, 81, 182
 - vector, 77–80, 828
 - vector for intensity, 107

- vector for radiance, 107
- Stokes vector
 - for irradiance (of a beam), 111
- stratosphere, 310
- stray light, 840
- Sun dogs, 312
- Sun glint, 570, 662
- Sun glitter, 470, 570, 662
- Sun halos, 312
- superposition
 - in a dielectric, 41
 - of balls of charge, 34
 - principle of, 24
- SWIR, 341
 - for atmospheric correction of MODIS, 729
- synodic period, 633
- T-Matrix, 331
- Taylor series, 899
- Total Ozone Mapping Spectrometer (TOMS), 727
- transmission
 - function of a boundary, 164
- transmittance
 - diffuse, 656, 663–667, 722
 - direct, 287, 663
 - of irradiance, 597
 - of sky radiance, 589
 - Rayleigh, of atmosphere, 287
- trapezoidal rule, 940
- troposphere, 310
- validation
 - definition of, 859
- van de Hulst approximation, 244–248
 - for absorption, 365
 - for extinction, 370
- vectors, 909
 - basis, 930
 - components of, 910
 - cross product of, 910
 - curl of, 914
 - divergence of, 914
 - divergence theorem, 916
 - dot product of, 910
 - gradient operator, 913
 - gradient vector, 913
 - identities for, 914
 - Laplacian operator, 914
 - magnitude of, 909
 - multiplication by scalars, 910
 - negative of, 910
 - position vector, 912
 - representation as column matrix, 922
 - scalar product of, 910
 - Stokes theorem, 916
 - sum of two, 910
 - unit, 909
 - vector product of, 910
- verification, 869
- vibration-rotation spectrum, 326
- vicarious gain, 876
- volume scattering function
 - definition of, 113
 - diffraction contribution to, 252
 - for atomic scattering, 117
 - for inelastic scattering, 119
 - for Rayleigh scattering, 222
 - in the DDA, 256
 - inelastic for Raman scattering, 348
 - of particles in Case 1 waters, 357
 - of pure liquids, 231
 - phase function of, 134
 - Rayleigh with molecular anisotropy, 225
- volume scattering matrix
 - definition of, 114
 - for anisotropic Rayleigh scattering, 226
 - for particle scattering, 221
 - for Raman scattering, 273
 - for Rayleigh scattering, 181, 222
 - Mueller matrix for atomic scattering, 117

- of Case 1 waters, 360
 - of pure water, 342
- Voss-Fry, 472
- Wadsworth mounting, 791
- water-leaving radiance, 439, 509, 521, 656
- wave equation
 - one-dimensional, 27
 - three-dimensional, 26
- wave function (in quantum mechanics), 50
- wave number
 - definition, 325
- wave packet, 84
- WetLab Average, 357
- whitecaps, 567, 600–602
- work
 - done by fields, 24
- yellow substance, 350