Synergy between Ocean Colour & Biogeochemical/ Ecosystem Models

REPORT NUMBER 19

iCCG

In the IOCCG Report Series:

Reports and Monographs of the International Ocean Colour Coordinating Group

An Affiliated Programme of the Scientific Committee on Oceanic Research (SCOR)
An Associated Member of the Committee on Earth Observation Satellites (CEOS)

IOCCG Report Number 19, 2020

Synergy between Ocean Colour and Biogeochemical/ Ecosystem Models

Edited by: Stephanie Dutkiewicz

Report of the IOCCG working group on the Role of Ocean Colour in Biogeochemical, Ecosystem and Climate Modelling, chaired by Stephanie Dutkiewicz, and based on contributions from (in alphabetical order):

Mark Baird	Commonwealth Scientific and Industrial Research Organisation, Australia
Fei Chai	University of Maine, USA/Second Institute of Oceanography, China
Stefano Ciavatta	Plymouth Marine Laboratory/National Centre for Earth Observation, UK
Stephanie Dutkiewicz	Massachusetts Institute of Technology, USA
Christopher A. Edwards	University of California, Santa Cruz, USA
Hayley Evers-King	Plymouth Marine Laboratory, UK
Marjorie A. M. Friedrichs	Virginia Institute of Marine Science, William & Mary, USA
Sergey Frolov	General Dynamics Information Technology, USA
Marion Gehlen	Laboratoire des Sciences du Climat et de l'Environnement, France
Stephanie Henson	National Oceanography Center, UK
Anna Hickman	University of Southampton, UK
Amir Ibrahim	NASA Goddard Space Flight Center, USA
Oliver Jahn	Massachusetts Institute of Technology, USA
Emlyn Jones	Commonwealth Scientific and Industrial Research Organisation, Australia
Daniel E. Kaufman	Chesapeake Research Consortium, Annapolis, MD, USA
Frédéric Mélin	European Commission, Joint Research Centre, Italy
Colleen Mouw	University of Rhode Island, USA
Barbara Muhling	University of California Santa Cruz/NOAA Southwest Fisheries, USA
Cecile Rousseaux	USRA/NASA Goddard Space Flight Center, USA
Igor Shulman	US Naval Research Laboratory, USA
Charles A. Stock	NOAA Geophysical Fluid Dynamics Laboratory, USA
P. Jeremy Werdell	Ocean Ecology Laboratory, NASA Goddard Space Flight Center, USA
Jerry D. Wiggert	University of Southern Mississippi, USA

The International Ocean Colour Coordinating Group (IOCCG) is an international group of experts in the field of satellite ocean colour, acting as a liaison and communication channel between users, managers and agencies in the ocean colour arena.

http: //www.ioccg.org

Contents

Chapter 1

Bridging Satellite Ocean Colour Remote Sensing and Biogeochemical/Ecosystem Modelling

Stephanie Dutkiewicz, Mark Baird, Stefano Ciavatta, Stephanie Henson, Anna Hickman, Cecile Rousseaux and Charles Stock

This report is intended as part of the important dialogue between the ocean colour and the biogeochemical/ecosystem/climate modelling communities. Numerical modellers are frequent users of ocean colour products, but many modellers remain unsure of the best way to use these products, and are often unaware of the uncertainties associated with them. On the other hand, the ocean colour community often are unsure on how models work, their usefulness and their limitations.

The colour of the ocean is set by incident light interacting with constituents (both dissolved, living and non-living particles) in the water. "Ocean colour", as referred to in this report, is the science and products developed from satellite remote sensing of the light reflected from the ocean. These satellites provide data in select visible wavebands of the light spectrum that tell us about the "colour" of the ocean, and thus also about the types of constituents (including phytoplankton) in the water. Chapter 2 provides a non-experts introduction to ocean colour.

In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales. We discuss the process of constructing models and uses of models in Chapter 3. "Biogeochemical" and "ecosystem" modelling is used in the title to signify that we are encompassing models with different interests. Biogeochemical models address questions that are related to nutrient and carbon cycling. The focus of ecosystem models is on the ecology of the ocean. However, the difference between the two is not clear-cut and there is significant overlap in the two types of modelling.

This report is not intended to be comprehensive. We focus particularly on large scale, three dimensional models, and often consider open ocean (rather than coastal) processes. We have also included mostly phytoplankton-centric research. This is a result of the working group's main interests and we emphasis that there are many more types of models and relevant work, beyond what is discussed here. We have attempted to add numerous references for further exploration by an interested reader.

1.1 Goals of the Report

The overall goal of this report is to achieve more synergy between ocean colour and models. To do this we:

❖ Provide non-experts in ocean colour with non-jargon understanding of uses, as well as uncertainties, and limitations of ocean colour products;

❖ Provide the ocean colour community with an understanding of types, uses and limitations of ecosystem/biogeochemical/climate models, including data assimilation;

❖ Explore the similarities/difference between similarly named variables in the two communities;

❖ Provide recommendations on the use of ocean colour products for model skill assessment;

❖ Introduce new developments in the parameterization of optics and radiative transfer in models that could provide better links between the two communities;

❖ Provide examples of studies which have integrated ocean colour and models to better understand processes and trends in the ocean's ecosystem and biogeochemistry, as well as feedback to the climate;

❖ Provide examples where models can help inform on ocean colour, with the goal of fostering further use of models as laboratories for ocean colour studies, understanding uncertainties and algorithm development;

❖ Highlight gaps in research and understanding.

1.2 Layout of the Report

The remaining chapters in the report are summarized here briefly to aid the reader in identifying those that are most useful to them. Each chapter is written so that it stands on its own; thus a reader does not need to read the entire report.

Chapter 2: Ocean Colour Remote Sensing Overview — In Chapter 2 we provide an overview of satellite remote sensing products for a modelling audience to help identify the strengths and limitations of various existing products, as well as potential future products resulting from the anticipated capability of next generation ocean colour sensors.

Chapter 3: Biogeochemical and Ecosystem Models: What are They and How can They be Used? — This chapter introduces scientists (particularly the ocean colour community) to the ideas, concepts, and basic building blocks of biogeochemical/ecosystem models. We provide a section on how modellers have treated light in models, and some of the new model developments to include radiative transport and spectral light. We also highlight some of the types and uses of biogeochemical/ecosystem models. Appendix 4 provides definitions of some unavoidable jargon and other model terminology.

Chapter 4: The Mismatch between Model Output and Ocean Colour Products — In this chapter we highlight how ocean colour products, *in situ* observations and biogeochemical model output do not compare cleanly. There are discrepancies about what is actually being

captured (a mismatch in name), uncertainties in *in situ* observation and ocean colour products that are not well understood or quantified, and biases linked to missing data. These mismatches are often not well understood and provide a hindrance to the best use of ocean colour (and models).

Chapter 5: Ocean Colour for Model Skill Assessment — In Chapter 5, we introduce commonly used model skill metrics, and with Chl as an example, show how these metrics are used. The chapter highlights some of the issues that arise from using satellite products (especially as uncertainties increase with more derived products) and in trying to compare point-to-point in a heterogeneous ocean. Emphasis is put on using skill assessment "fit to purpose" of the space and time scales of interest.

Chapter 6: Assimilation of Ocean Colour — The assimilation of ocean colour into biogeochemical models provides a rigorous method to include both observations and models into one unified output that incorporates the information and advantages from both sources. Chapter 6 provides a brief introduction to data assimilation, the different methods that can be employed, and the potential outputs, including state estimates and parameter estimations. The chapter provides several case studies of using ocean colour products and models together formally, to provide best estimates of ocean biogeochemical fields.

Chapter 7: Synergistic Use of Ocean Colour Data and Models to Understand Marine Biogeochemical Processes — The increasing number of satellite ocean colour missions and products, combined with the continuing development of complex numerical models, allows for new and exciting multi-disciplinary approaches to tackling ecological and biogeochemical questions. Chapter 7 provides case studies that have used both ocean colour and models to significantly enhance our understanding of global-, basin-, and mesoscale time-series and elucidates details of regional processes and phytoplankton physiological states.

Chapter 8: Using Models to Inform Ocean Colour Science — Numerical models can be used as laboratories to help understand some of the limitations and uncertainties of ocean colour products, to help understand future needs of ocean colour missions, and to aid in algorithm development. Additionally, models can be subsampled to match the spatial and temporal distribution of satellite observations to investigate issues with missing data. In Chapter 8 we provide several case studies that explore the ways that models can be used to help inform ocean colour output, and planning for the future.

Chapter 9: Summary and Final Recommendations — In this final chapter we provide some recommendations for the further linking of ocean colour products and models.

1.3 Definitions, Symbols, Jargon and Acronyms

According to one definition: "A model of a system or process is a theoretical description that can help you understand how the system or process works, or how it might work" (https://www.collinsdictionary.com/dictionary/english/model).

In the marine biogeochemical/ecosystem numerical modelling world, a model is a set of equations ("theoretical description") of marine physical, biogeochemical and ecological

processes ("the system"), that are translated into computer code that then provides, as output, how the "system" changes with time. By using a computer, many components and timescales can be included that would otherwise not be possible.

However, the word "model" encompasses many things to many people, and it is important here to be careful about what we mean when we use the word. Theoreticians will call the equations a "model" by themselves. There are statistical techniques to analysis data that are called "models". In the ocean colour community a "model" might mean the method of taking an ocean colour measurement (such as reflectance) and using an algorithm or semi-empirical method to produce a derived product such as Chl-a. In this report we will usually differentiate this type of model by using the word "algorithm" instead.

We will remind the reader of the different uses of the word "model", and in particular what we mean by the word, at the beginning of each chapter:

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

Both communities (ocean colour and modellers) have their own set of symbols, jargon and acronyms. It is thus often difficult to communicate as, in some way, we are talking different languages. In this report we try to maintain a consistent set of symbols, based mostly on those used by the ocean colour community (Appendix 1). Appendix 2 provides some commonly used acronyms, including those for the past, current and future sensors. Appendix 3 provides some of the terminology used in the ocean colour and space agency communities. We also provide a table with jargon and some of the acronyms that are frequently used by the modelling community (Appendix 4). None of these tables are exhaustive, but hopefully provide enough information to help the reader negotiate the often complex new language of an unfamiliar field.

Chapter 2

Ocean Colour Remote Sensing Overview

**Colleen Mouw, Cecile Rousseaux, Frédéric Mélin, Hayley Evers-King,
Amir Ibrahim and Jeremy Werdell**

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeo-chemical/ecosystem computer models at large regional or global scales.

2.1 Scope

Since the launch of the first mission (Coastal Zone Colour Scanner, CZCS) in 1978, satellite remote sensing of ocean colour has provided an unprecedented view of biogeochemical processes of the ocean surface layer. Given the spatial coverage and repeat frequency, ocean colour imagery has been an important data source for model assessment. The extent of these comparisons and expansion into model assimilation grew significantly with the continuous, global, ocean colour record, beginning with the launch of the Sea-viewing Wide-Field-of-view Sensor (SeaWiFS) in 1997 and the following missions that continue through to the present. There are a variety of satellite products and associated algorithms; some are intended for global use, while others are for regionally-specific applications. Satellite products are often used to assess model performance, however, often the satellite products are not a direct match to the state variables (see Appendix 3, Chapter 3) in models (see Chapter 4). Most biogeochemical models have not been developed with optical interests — rather they have been structured to follow pathways of matter and energy, and capture basic groups of marine organisms to investigate carbon cycling and ecology. Thus, the link to optical and ocean colour products are not always clear. Here we provide an overview of satellite remote sensing and associated products for a modelling community to help identify the strengths and limitations of various commonly utilized products. The products addressed here are not meant to be an exhaustive list, rather those commonly utilized by the modelling community.

2.2 What Does a Satellite Radiometer "See"?

The radiance observed by a satellite spectroradiometer contains information about the optically significant constituents in the atmosphere and ocean. Ocean colour algorithms are used to derive the optical properties of the ocean's constituents from the visible light emerging from below the water surface, which can, in turn, be used to infer their concentrations. In cloud-free

conditions, the atmospheric constituents account for the largest contribution to the top-of-atmosphere (TOA) radiance measured by a satellite, typically 90%, but it varies depending on wavelength and water brightness (IOCCG 2010) (Figure 2.1). The ocean accounts for the small residual signal necessary for deriving ocean colour products. This inherently requires a stringent process to remove the radiometric contribution of the atmosphere and ocean surface known as the "atmospheric correction", which is discussed below.
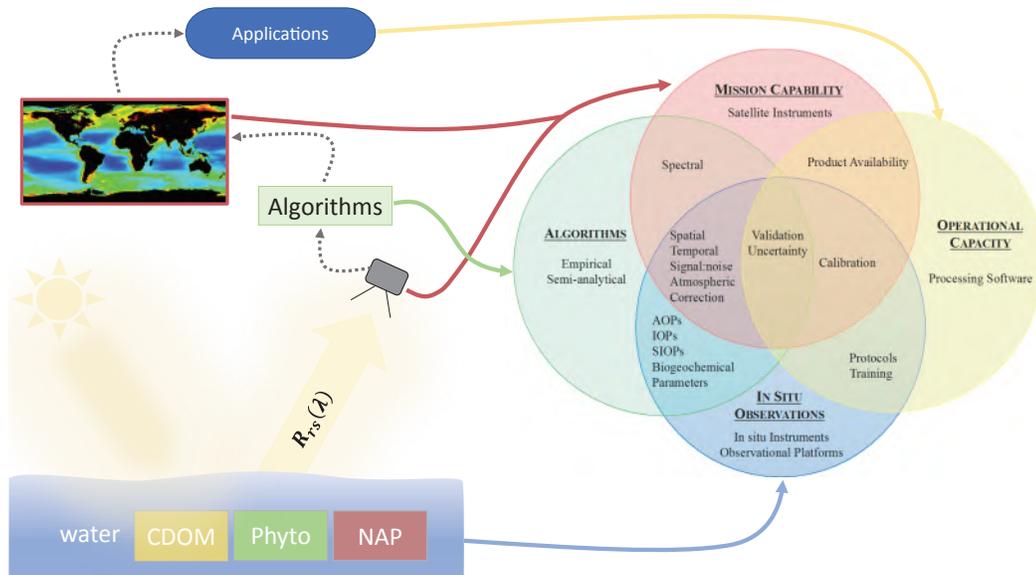
Figure 2.1 Schematic of path radiance from the Sun (bottom left), through the photic zone of the ocean, and through the atmosphere to a satellite radiometer. The optical constituents that influence ocean colour include pure water, chromophoric dissolved organic matter (CDOM), phytoplankton, and non-algal particles (NAP). Satellite imagery is produced from atmospherically-corrected spectral remote sensing reflectance, $R_{rs}(\lambda)$, with algorithms that connect satellite observations to optical, biogeochemical and water quality parameters. The inlaid image (top left) is mean chlorophyll-a concentration obtained from the MODIS-Aqua mission mean (2002–2017, https://oceancolour.gsfc.nasa.gov/). Right: Venn diagram of the fundamental elements of ocean colour remote sensing (from Mouw et al. (2015), reproduced with permission from Elsevier).

2.2.1   Atmosphere

Atmospheric constituents include aerosols and gas molecules (such as ozone, water vapor, oxygen, etc.) that diversely contribute to absorption and scattering of sunlight. Air molecules contribute to Rayleigh scattering, accounting for the largest component of the atmospheric signal, particularly in shorter wavelengths. The scattering efficiency of these molecules decreases with increasing wavelength. Some atmospheric gas molecules absorb the visible light across a broad spectrum (such as ozone and nitrogen dioxide), while other gases have strong distinct absorption features, such as water vapor and oxygen (Figure 2.2). Multispectral satellite ocean colour bands are usually positioned at atmospheric window bands away from strongly absorbing gases such as the oxygen and water vapor bands. Atmospheric aerosols

both absorb and scatter light depending on the aerosol composition and size distribution. Examples of global aerosols distribution, including sea salt, dust and pollution, are shown in Figure 2.3.
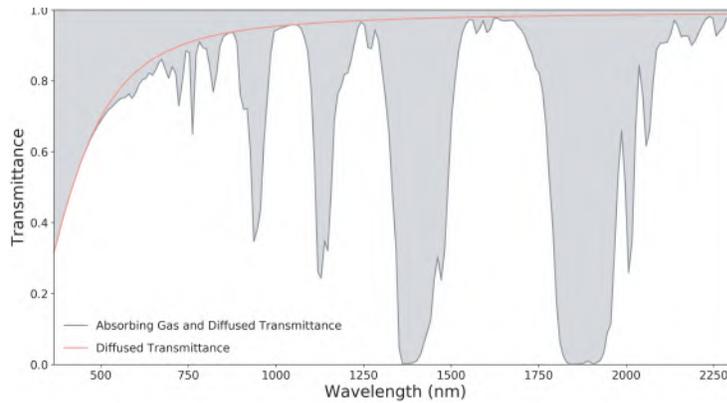
Figure 2.2 Atmospheric transmittance and spectral absorption and scattering of atmospheric constituents. The water vapor concentration is 0.95 cm and ozone is 355 DU and average climatology CO_2, CH_4, N_2O, and O_2. The aerosol optical depth is 0.1 at 869 nm with a 0.8 Ångström coefficient.
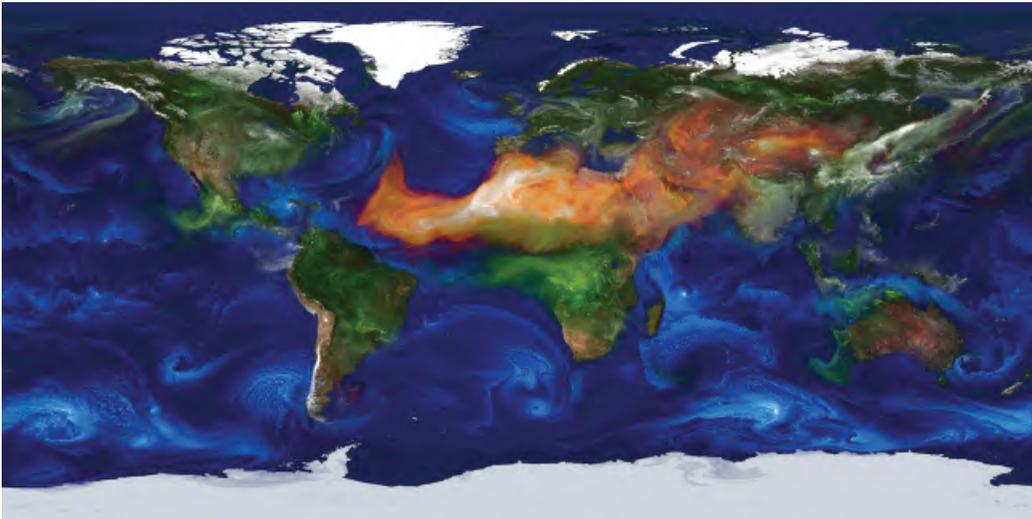
Figure 2.3 Global aerosol model distribution, where the colour-coded aerosol plumes indicate the aerosol type. Red shaded regions indicate presence of dust, blue is sea salt, green is carbonaceous aerosols (organic or black carbon), and gray is sulfate aerosols. Image credit: NASA Global Modeling and Assimilation Office GEOS-5 Nature Run model.

The radiance measured at TOA by a satellite sensor, $L_t(\lambda)$, is the summation of the contribution from every component of the atmosphere-ocean system. All radiances discussed hereafter carry units of $\mu W\,cm^{-2}\,nm^{-1}\,sr^{-1}$. A standard equation of the atmospheric correction is Gordon (1997):

$$L_t(\lambda) = L_r(\lambda) + [L_a(\lambda) + L_{ra}(\lambda)] + T(\lambda)L_g(\lambda) + t(\lambda)L_{wc}(\lambda) + t(\lambda)L_w(\lambda), \qquad (2.1)$$

where $L_t(\lambda)$ is the top of atmosphere radiance measured by the satellite sensor, $L_r(\lambda)$ is radiance due to (Rayleigh) scattering by air molecules, $L_a(\lambda)$ is radiance due to scattering by aerosols, $L_{ra}(\lambda)$ represents the multiple scattering interactions between molecules and aerosols, $L_{wc}(\lambda)$ is radiance resulting from white caps, $L_g(\lambda)$ is radiance resulting from specular reflection of sunlight off the sea surface (sun glint), $L_w(\lambda)$ is water-leaving radiance, $t(\lambda)$ is diffuse transmittance in the viewing direction, and $T(\lambda)$ is the direct transmittance from the surface to the sensor in the viewing direction. $L_r(\lambda), L_a(\lambda),$ and $L_{ra}(\lambda)$ represent radiances generated by the atmosphere, $L_{wc}(\lambda)$ and $L_g(\lambda)$ are radiances generated at, or immediately below, the surface of the ocean, and $L_w(\lambda)$ is water-leaving radiance resulting from light backscattered from below the water surface. $L_w(\lambda)$ is the parameter that the atmospheric correction aims to retrieve.

As L_w depends on conditions of the observations, it is usually expressed after normalization by writing:

$$L_{WN}(\lambda) = L_w(\lambda)\frac{F_0(\lambda)}{E_d^+(\lambda, \theta_0)} \approx \left(\frac{d}{d_0}\right)^2 \frac{L_w(\lambda)}{t_0(\lambda)\cos\theta_0}, \qquad (2.2)$$

where $F_0(\lambda)$ is the mean extraterrestrial irradiance (Thuillier et al. 2003) and $E_d^+(\lambda)$ is the downwelling irradiance just above the sea surface associated with the solar zenith angle θ_0, $t_0(\lambda)$ is the diffuse transmittance of the atmosphere from the Sun to the ocean surface, d is the Sun-Earth distance and d_0 its mean. In effect, this operation corrects for the variations of amplitude of solar irradiance at the water surface. L_{WN} can be further corrected for bidirectional effects (the emerging radiance field generally not being isotropic) using various approaches that express L_{WN} as if the water is observed at nadir with overhead Sun (e.g., Morel et al. 2002).

Additional information beyond the satellite radiances are usually required for the atmospheric correction. $L_r(\lambda)$ can be computed from Rayleigh look-up tables that typically utilize solar-sensor geometry, atmospheric pressure and wind speed (Wang 2005). White cap radiance can be modeled using sea surface wind speed (Moore et al. 2000). In widely-used standard atmospheric correction schemes, sun glint is primarily masked but residual contamination is corrected (Wang and Bailey 2001). Specific atmospheric corrections have been developed to operate in sun glint conditions (e.g., Steinmetz et al. 2011). Other quantities are required to proceed with atmospheric correction, such as relative humidity, to select aerosol optical properties (e.g., Ahmad et al. 2010) and ozone concentrations. These atmospheric quantities are usually obtained at a time resolution of a day or hours from national weather prediction centers or from satellite data.

For the rest of this chapter, we will be using remotely sensed reflectance ($R_{rs}(\lambda); sr^{-1}$) as the primary parameter derived from a satellite radiometer, rather than water-leaving radiance noted in Equations 2.1 and 2.2. The two terms are interchangeable according to the following expressions:

$$R_{rs}(\lambda) = \frac{L_w(\lambda)}{t_0(\lambda)F_0(\lambda)\cos(\theta_0)} = \frac{L_{WN}(\lambda)}{F_0(\lambda)} \qquad (2.3)$$

R_{rs} has become the standard product distributed by space agencies (in some cases multiplied by a factor π, then becoming dimensionless).

Given the large contribution of the atmosphere to the total light reaching a satellite radiometer over the ocean, the largest potential source of error in measuring $R_{rs}(\lambda)$ from space is the error associated with the atmospheric correction. There are differing standard approaches by various space agencies. Historically, standard atmospheric corrections (see Mobley et al. 2016 for a comprehensive summary) decouples the radiance contributions from the ocean and the atmosphere through the "black-pixel" assumption (Siegel et al. 2000). The absorption coefficient for water increases dramatically in the near-infrared (NIR, >780 nm – 2500 nm, Gordon and Wang 1994), such that an assumption can be made that R_{rs}(NIR) is negligible (i.e., black) in open ocean waters — that is, the sensor-measured NIR reflectance results only from Rayleigh scattering and atmospheric aerosols. After subtracting the calculable molecular (Rayleigh) scattering, the remaining NIR reflectance can be attributed to the aerosols, which are extrapolated to, and subtracted from, the visible bands. However, in waters with abundant scattering materials (e.g., suspended sediments or intense algal blooms), R_{rs}(NIR) is no longer negligible, which violates the black-pixel assumption and will lead to atmospheric correction failures. To accommodate these cases, bio-optical modelling is used to estimate NIR optical properties (e.g., Bailey et al. 2010) and an iterative scheme is operated to compute a best-estimate of R_{rs}(NIR). In the case of the Ocean Land Colour Instrument (OLCI), following the heritage from the Medium Resolution Imaging Spectrometer (MERIS) processing, a bright pixel atmospheric correction method (Moore et al. 1999) addresses the NIR contribution by using a coupled atmosphere-ocean model based on inherent optical properties (IOPs) and optimized through inversion over five bands in the NIR.

Beyond this standard approach, a variety of atmospheric correction schemes have been developed, with different applications and principles. IOCCG (2010) provides a review of the published atmospheric correction approaches. Some techniques have used neural networks, non-linear minimization (e.g., Chomko and Gordon 2001) or a Bayesian approach (Frouin and Pelletier 2015). Few of these alternative schemes have been applied in an operational context. Artificial neural networks are used in an operational manner to process MERIS and OLCI data by the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), respectively (Doerffer and Schiller 2007). Another counter-example is the application of the POLYMER (POLYnomial based algorithm applied to MERIS) algorithm (using spectral optimization based on polynomials, Steinmetz et al. 2011) in ESA's Climate Change Initiative.

2.2.2 Ocean

The light emerging from below the water surface is the signal useful to determine in-water inherent optical properties (IOPs) or concentrations of optically significant constituents. The relation between R_{rs} and IOPs can be approximated by the following equation:

$$R_{rs}(\lambda) = G(\lambda) \left(\frac{(b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \right), \tag{2.4}$$

where $G(\lambda)$ (sr^{-1}) is a coefficient that accounts for the air-water interface effect, the bidirectional character of the in-water light field, and the effect of multiple scattering effects, and $a(\lambda)$ and $b_b(\lambda)$ are the bulk absorption and backscattering coefficients, respectively (m^{-1}). The absorption and backscattering coefficients are IOPs, which means that they are independent of the ambient light field. Equation 2.4 puts particular emphasis on the backward part of scattering as it is a major contributor to $R_{rs}(\lambda)$.

The optically active constituents in the ocean show daunting complexity with size, spanning orders of magnitude (e.g., Stramski et al. 2004). For simplicity, their optical contributions are usually partitioned into IOPs associated with pure water, chromophoric dissolved organic matter (CDOM), non-algal particles (NAP), and phytoplankton. The total absorption $(a_t(\lambda))$ is the sum of the absorption of all of these constituents $(a_w(\lambda), a_{CDOM}(\lambda), a_{NAP}(\lambda)$, and $a_{ph}(\lambda)$, respectively, Equation 2.5). Considering their similar spectral shapes, $a_{CDOM}(\lambda)$ (also
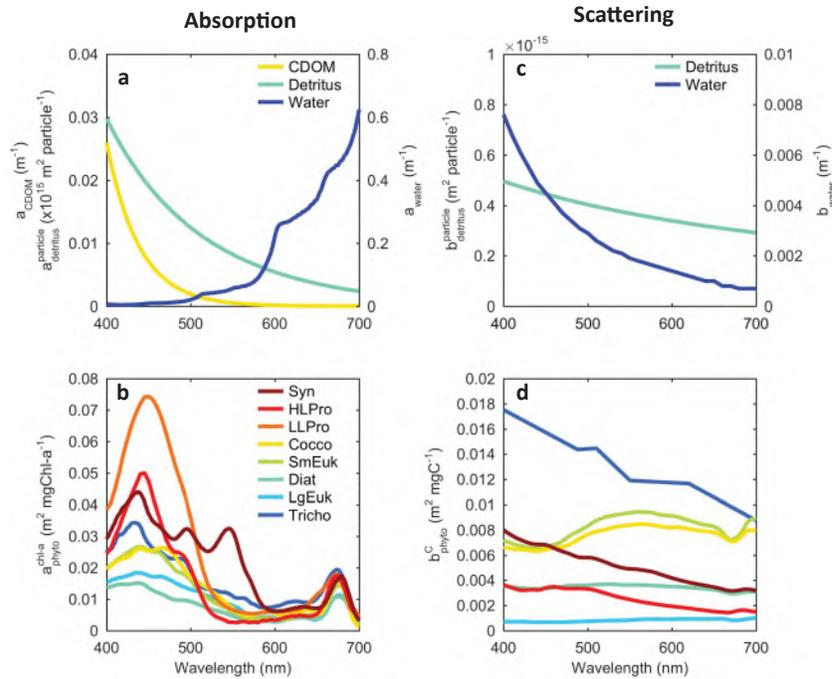
Figure 2.4 Spectral absorption and scattering of ocean optical constituents. Spectral absorption for a) water, chromophoric dissolved organic matter (CDOM), non-algal particles (NAP) or detritus, and b) various phytoplankton groups. Spectral scattering c) NAP/detritus, water, and d) various phytoplankton groups. The phytoplankton types were taken from Dutkiewicz et al. (2015a) and include: Syn, *Synechococcus*; HLPro, *Prochlorococcus*; LL-Pro, *Prochlorococcus*; Cocco, *Emiliania huxleyi*; SmEuk, *Isochrysis galbana*; Diat, *Thalassiosira weissflogii*; LgEuk, *Prorocentrum micans*; Tricho, *Trichodesmium* sp. The phytoplankton optical characteristics were obtained for representative types in culture. Spectral a_{CDOM} are the observed average from several AMT transects (Kitidis et al. 2006). Adapted from Dutkiewicz et al. (2015a), Creative Commons Attribution 3.0 License (CC BY 3.0).

noted $a_g(\lambda)$, for Gelbstoff or yellow substance) and $a_{NAP}(\lambda)$ (also noted $a_d(\lambda)$ for detritus) are often treated together as $a_{dg}(\lambda)$. Their spectral absorption and scattering characteristics are shown in Figure 2.4. CDOM is dissolved (i.e., material that passes through a 0.2μm filter) and thus, does not scatter light. Total backscattering is due to pure water ($b_{bw}(\lambda)$), non-algal particles ($b_{b_{NAP}}(\lambda)$), and phytoplankton ($b_{b_{ph}}(\lambda)$), where non-algal particles and phytoplankton are often grouped into particulate backscattering ($b_{bp}(\lambda) = b_{b_{NAP}}(\lambda) + b_{b_{ph}}(\lambda)$, Equation 2.6). Their spectral absorption and scattering characteristics are shown in Figure 2.4. A comprehensive overview of these optical properties and atmospheric correction can be found: http://www.oceanopticsbook.info/.

$$a(\lambda) = a_w(\lambda) + a_{CDOM}(\lambda) + a_{NAP}(\lambda) + a_{ph}(\lambda) \tag{2.5}$$

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda) \tag{2.6}$$
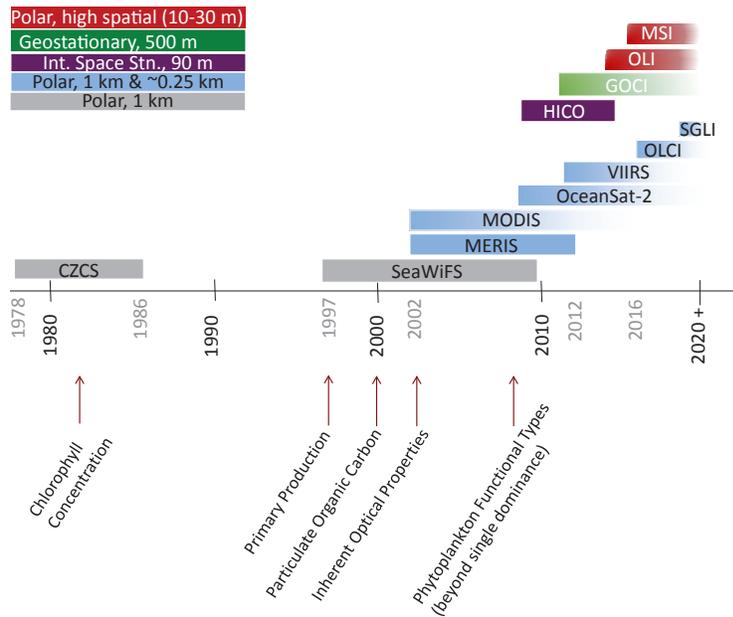
Figure 2.5 Current ocean colour mission lifetime with concurrent timeframe of major product developments. Faded gradient indicates missions in current operation. Note: OLI and MSI are instruments developed for terrestrial observations that have demonstrated success in some coastal applications.

2.3   Historical and Current Instruments

The colour of the ocean has been imaged by satellite sensors for nearly 40 years. The advent was the proof-of-concept CZCS (1978 – 1986), followed among others by SeaWiFS (1997 – 2010), Moderate Resolution Imaging Spectrometer (MODIS, 2002 – present, 250 and 500 m for a subset of bands), Medium Resolution Imaging Spectroradiometer (MERIS, 2002 – 2012, 1000 m and 300 m), and Visible Infrared Imaging Radiometer Suite (VIIRS, 2012 – present, 750 m

and 375 m for a subset of bands), to name but a few (Figure 2.5). This list includes the polar orbiting sensors that provide global coverage at roughly 1000 m spatial resolution, with some of these also observing at medium spatial resolution. Satellite radiometers have four types of requirements that ultimately drive the final design:

1. spatial coverage and resolution,
2. temporal coverage and revisit frequency,
3. spectral coverage as well as number and position of spectral bands, and
4. radiometric quality (IOCCG 2012) (Figure 2.6).
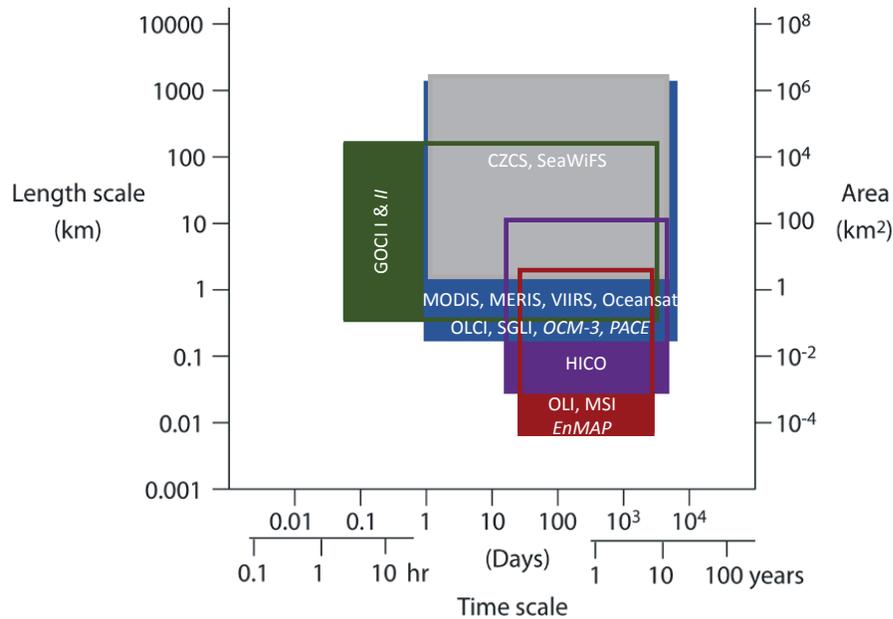
Figure 2.6 Spatial and temporal resolution of heritage, current and planned ocean colour satellite sensors. Planned sensors and missions are italicized. Note: OLI and MSI are instruments developed for terrestrial observations that have demonstrated success in some coastal applications. Adapted from Mouw et al. (2015).

The Korean Geostationary Ocean Colour Imager (GOCI, 2011 – present, 500 m) is the only geostationary ocean colour sensor. It is centered on longitude 130°E, allowing for high temporal resolution (Ryu et al. 2012). More recently, as part of the European Commission Copernicus programme, ESA and EUMETSAT launched the Sentinel-3 platforms carrying the Ocean and Land Colour Instrument (OLCI) in February 2016 (Sentinel-3A) and April 2018 (Sentinel-3B). These provide continuity of MERIS-class polar orbiting observations, with global 300 m spatial resolution. Sensors that have been utilized for coastal and inland waters due to their high spatial resolution include the Hyperspectral Imager for the Coastal Ocean (HICO, 2009 – 2014, 90 m), Landsat-8 Operational Land Imager (OLI, 30 m) and the Sentinel-2 Multispectral Instrument (MSI, 10–60 m). With the exception of HICO, all sensors launched to date have had multispectral imaging capability (Figure 2.7). HICO imaged hyperspectrally with 124 bands at 5.73 nm spectral resolution between 400 to 900 nm. HICO was mounted on the International Space Station, thus its imaging coverage of the Earth was opportunistic. OLI and MSI were developed

for terrestrial remote sensing, thus their band placement, band width and signal-to-noise ratios are not optimized for ocean targets, yet several studies have successfully demonstrated their importance for coastal remote sensing (Pahlevan et al. 2014). A full list of current missions can be found at http://ioccg.org/resources/missions-instruments/current-ocean-colour-sensors/. These instruments are equipped with sensors optimized for measuring remote sensing reflectance over most of the world's oceans, but most are limited in their ability to observe inland or coastal waters (Mouw et al. 2015). In this chapter, we focus on standard products that were designed for application across basin to global scales. Readers interested in coastal and non-standard cases are encouraged to consult Mouw et al. (2015), Zheng and DiGiacomo (2017), and IOCCG (2000, 2012).
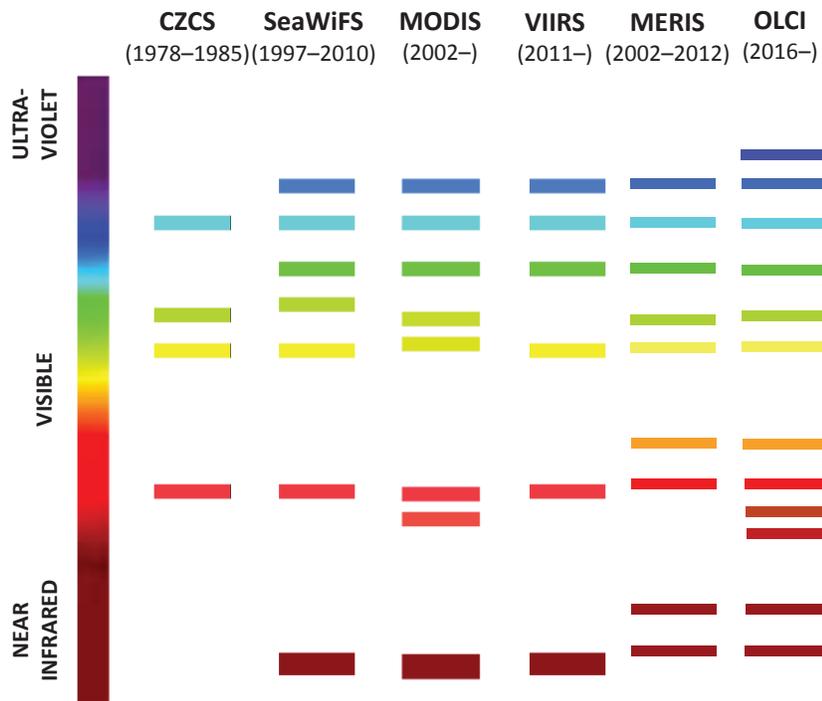
Figure 2.7 Spectral resolution of current polar orbiting ocean colour satellite sensors. Adapted from PACE Science Definition Team Report (PACE 2012). A full list of past, current and future missions can be found at http://ioccg.org/resources/missions-instruments/.

2.3.1 Imagery products

Most ocean colour variables are estimated from $R_{rs}(\lambda)$ (noting that some are derived from $L_t(\lambda)$). The optically significant constituents within the ocean dictate how sunlight is absorbed and scattered in the water column, ultimately shaping $R_{rs}(\lambda)$. Over time, satellite imagery products have evolved from chlorophyll-a concentrations to primary production, specific carbon pools, such as particulate organic carbon (POC), inherent optical properties (IOPs), and single and multiple phytoplankton groups (Figure 2.5). This evolution has been driven

by the ever-increasing number of *in situ* optical observations across the globe driving a continually growing understanding of the distribution and variability of the optical constituents (described in Section 3.1.2) and their relationship with geophysical parameters. These products typically begin with atmospherically-corrected $R_{rs}(\lambda)$, but the methods of derivation and their resulting uncertainties differ, in particular with regards to how far the derivation of geophysical parameters falls from the original $R_{rs}(\lambda)$ and the use of underlying, intermediate optical parameters. Zheng and DiGiacomo (2017) have laid out the idea of level of derivation and associated uncertainty by assigning various products to different "tiers" (Figure 2.8). Their Tier 1 is the top-of-atmosphere radiance observations made directly by the satellite radiometer, and has the least uncertainty. Conversely, Tier 5 variables have the highest level of derivation and the most uncertainty. We have added a Tier 6, to also include primary production and phytoplankton functional types. The uncertainty of the tiers above accumulates to the variables in a given tier.

Tiers of Variables **Added Uncertainties**

Tier-1: Sensor level signal
$L_{tot}(\lambda)$
- Pre-launch calibration;
- On-board calibration;
- Vicarious calibration.

Tier-2: Radiance and reflectance
$L_w(\lambda)$ $nL_w(\lambda)$ $R_{rs}(\lambda)$ Glint reflectance
- "Black-pixel" assumption;
- Aerosol optical property;
- Glint;
- Adjacency effect;

Tier-3: Total IOPs and AOPs
$a(\lambda)$ $b_b(\lambda)$ Turbidity $K_d(\lambda)$ Visual clarity
- Absorption and angular scattering coefficients;
- Sun-sensor geometry;
- Sea surface state;
- Bubble clouds;
- Bottom conditions;
- Multiple scattering;
- Inelastic processes.

Tier-4: Component IOPs
$a_{ph}(\lambda)$ $a_d(\lambda)$ $a_g(\lambda)$
- Spectral shapes of various components;
- Different combinations of component IOPs summing to nearly identical total spectrum (multiple solution problem).

Tier-5: Mass concentrations
[SPM] [Chl-a] [DOC]
- Particle size distribution;
- Mineral and organic contributions;
- Density and iron content of minerals;
- Density, water content, proteins, and humic matter of dry organic materials;
- Phytoplankton intracellular pigments and internal structure;
- Proportion of colored to uncolored DOC;

Tier-6: Primary Production and Phytoplankton Functional Types
NPP
PFT PSC PSD PTC
- Physiological variability
- Underlying empirical assumptions requiring on-going recalibration
- Small deviations in spectral shape/magnitude may be difficult to retrieve

Accumulation of Uncertainty

Figure 2.8 Tiers of satellite-derived products and associated uncertainties introduced at each tier. The list of products is representative, but not exhaustive. Adapted from Zheng and DiGiacomo (2017), Creative Commons Attribution License (CC BY).

The critical step of deriving quantitative in-water, optical, biogeochemical and water quality information from satellite-derived $R_{rs}(\lambda)$ requires the use of bio-optical algorithms. A wide suite of algorithms have been developed, tested, and implemented (Gordon and Morel 1983; IOCCG 2000, 2006), that can be broadly categorized into two groups: empirical and semi-analytical (Figure 2.9). Empirical methods are solely data driven and based on observed statistical relationships, whereas semi-analytical approaches combine data driven relationships with methods based on simplifications to the radiative transfer equation. However, the distinction between the two approaches can sometimes be unclear, as examples exist that blur this differentiation; some empirical algorithms have been developed from methods based on the radiative transfer equation (e.g., Doxaran et al. 2002), and most semi-analytical algorithms contain empirical relationships (e.g., Garver and Siegel 1997; Lee et al. 2002). Both empirical and semi-analytical algorithms can be used for effective generation of biogeochemical products from $R_{rs}(\lambda)$. The earliest ocean colour satellite products focused on the retrieval of chlorophyll concentration, [Chl], in waters where phytoplankton dominate the optical properties or covary with other optically active constituents (Table 2.1).
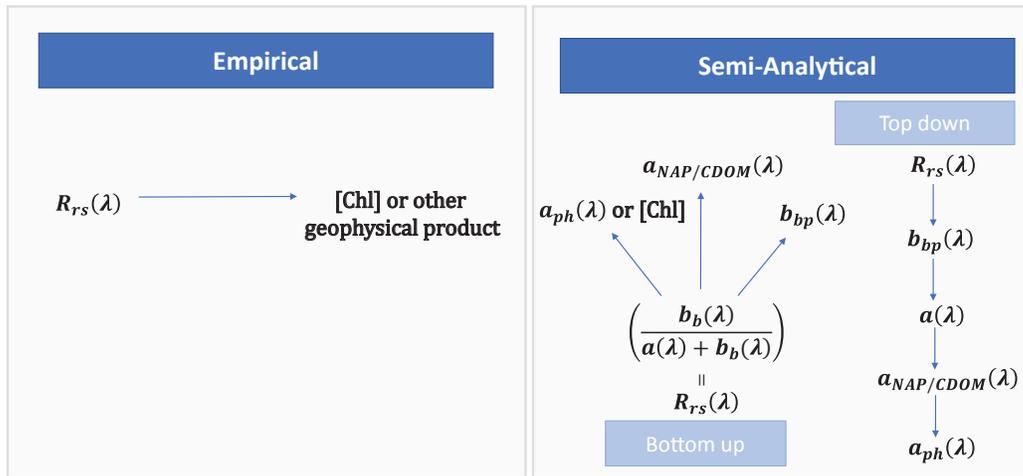
Figure 2.9 Schematic of fundamental connections of empirical vs. semi-analytical algorithm approaches.

2.3.2 Chlorophyll concentration

Empirical algorithms contain explicit or implicit empirical expressions. The most widely used empirical algorithms for the retrieval of [Chl] are based on band ratios (Gordon et al. 1983; O'Reilly et al. 1998). For a given pixel, the ratio of the greatest $R_{rs}(\lambda_{blue})$ (instrument specific bands between 443 and 520 nm) is normalized to a green band (i.e., the instrument specific band closest to 555 nm). This band ratio is then related to [Chl] through a 4th order polynomial.

$$[\log]_{10}(\text{Chl}) = a_0 + \sum_{i=1}^{N} a_i \left(\log_{10} \left(\frac{R_{rs}(\lambda_{blue})}{R_{rs}(\lambda_{green})} \right) \right)^i \tag{2.7}$$

Table 2.1   Assumptions, strengths, and limitations of empirical vs. semi-analytical and PFT algorithms. PFT algorithm summary was taken from Mouw et al. (2017).

Algorithms	Assumptions	Strengths	Limitations
Empirical	• All optical constituents covary with the target parameter	• Easy to implement	• Unable to distinguish the influence of optical constituents that are not covarying with the target parameter • Requires on-going recalibration as environmental change alters the relationship between the optical and biogeochemical parameters
Semi-Analytical Bottom Up	• Requires bio-optical models for each component during the retrieval process and derives each individual component and the bulk property simultaneously	• Ability to retrieve multiple optical components	• Dependent on empirical coefficients in the optical relationships inherent optical properties and the optical constituents retrieved • Does not independently retrieve the spectrum of any component
Semi-Analytical Top Down	• Retrieves the bulk property (total absorption or total back-scattering) first before decomposing into separate individual components, thus applying bio-optical models for each component separately during the inversion process	• Ability to retrieve multiple optical components • Ability to independently retrieve optical components	• Dependent on empirical coefficients in the optical relationships between the inherent optical properties and the optical constituents retrieved
PFT Abundance	• Change in size structure with change in [Chl] based on generalized relationships	• Easy to implement • Strong ecological basis	• Primary empirical relationships with [Chl] that cannot detect regional deviations • Unable to distinguish mixed populations of similar abundance • Requires on-going recalibration as environmental change alters phytoplankton assemblages • Susceptible to physiological variability
PFT Radiance	• After normalization to [Chl], changes in radiance are due primarily to variability in phytoplankton type	• Does not require, or limited dependence on, derived products • Input data (R_{rs}) has lower error than derived products	• Dependent on empirical relationships between radiance and pigments • Difficult to discriminate PFTs with similar normalized radiance signatures • Susceptible to physiological variability particularly normalized spectra
PFT Absorption	• Variability largely the result of composition and pigment packaging • Primary variability in absorption is related to different PFTs	• Not directly dependent on concentration	• Susceptible to physiological variability • Small deviations in spectral shape/magnitude can be difficult to retrieve • Difficult to discriminate PFTs with similar absorption signatures
PFT Scattering	• PSD and b_{bp} have a power-law shape • Relative proportions of bio-volume to total particulate	• Less sensitive to physiological variability	• Includes all particles, not just phytoplankton • Difficult to discriminate PFTs with similar scattering signatures

A band difference approach has been introduced to better characterize clear, low [Chl] water by reducing artifacts associated with residual solar glint, stray light, and atmospheric correction errors (Hu et al. 2012). The $R_{rs}(\lambda)$ difference algorithm, known as the colour index (CI) employs the difference between three-bands: $R_{rs}(\lambda_{green})$ and the reference linearly formed between $R_{rs}(\lambda_{blue})$ and $R_{rs}(\lambda_{red})$.

$$CI \ = R_{rs}(\lambda_{green}) - [R_{rs}(\lambda_{blue}) + \frac{(\lambda_{green} - \lambda_{blue})}{(\lambda_{red} - \lambda_{blue})} \times ((R_{rs}(\lambda_{red}) - R_{rs}(\lambda_{blue}))] \qquad (2.8)$$

The band ratio and band difference approaches (Equations 2.7 and 2.8) have been merged and constitute NASA's standard [Chl] product. Other empirical approaches include principal component analysis (Sathyendranath et al. 1994; Craig et al. 2012) of $R_{rs}(\lambda)$ that contain explicit empirical expressions, and artificial neural networks (Schiller and Doerffer 1999; Doerffer and Schiller 2007) that embed the empirical expressions (and associated coefficients). [Chl] can also be derived through the use of semi-analytical algorithms with appropriate assumptions on the phytoplankton [Chl]-specific absorption coefficient (a_{ph}/[Chl]) (e.g., Maritorena et al. 2002).

For optically-complex waters, and from sensors with appropriately placed bands (MERIS, OLCI, MSI), red edge features can be used for derivation of [Chl] (Gower et al. 1999; Gons 2002; Mishra and Mishra 2012; Moses et al. 2012; Matthews and Odermatt 2015). Though not included commonly in a standard suite of products, open source processors have been made available for the community to use these methods, including through the Sentinel Application Platform (SNAP) and Acolite processor for Landsat and Sentinel-2 (Vanhellemont and Ruddick 2018).

2.3.3   Carbon pools

There are a number of pools of carbon in the ocean that can be quantified to varying degrees of success using satellite ocean colour data. These major pools include particulate organic carbon (POC), particulate inorganic carbon (PIC), dissolved organic carbon (DOC), and dissolved inorganic carbon (DIC).

One of the most mature and readily available remotely-sensed parameters in terms of carbon pools is the concentration of POC. The algorithm of Stramski et al. (2008) is available as a standard product from NASA, and this algorithm has shown good global performance in recent algorithm intercomparisons (see Evers-King et al. 2017). This algorithm is based on a blue/green reflectance ratio, similar to many chlorophyll algorithms, and empirical approaches typically show similar performance to chlorophyll-a algorithms in global ocean contexts. As with [Chl] algorithms, estimation of POC in coastal waters can involve more uncertainty due to the optical complexity of these regions. There are many other experimental algorithms to derive POC, some of which can be applied as simple empirical relationships using derived products (e.g., IOPs — particularly b_{bp}, attenuation coefficients, or [Chl]). More complex methods are in development to incorporate the effects of particle size and type. Phytoplankton carbon (the portion of POC that is contained within phytoplankton cells), is less readily quantifiable from satellite. A number of experimental methods have been developed

but none are routinely included in satellite products (see algorithm intercomparison from Martinez-Vicente et al. 2017).

Products relating to PIC are mostly centered around the detection of calcite associated with coccolithophore blooms — one of the primary sources of PIC in the ocean. Due to their reflective properties, these blooms are relatively easy to distinguish compared to other phytoplankton types. An algorithm from Balch et al. (2005) is included as a standard product from NASA.

Dissolved inorganic carbon is the largest active pool of carbon in the ocean and is very important due to its role in the ocean carbonate buffer system. However, it is also one of the most difficult to quantify from remote sensing: it has no detectable optical signature so that ocean colour is used only indirectly through a variable like [Chl]. A variety of experimental methods have been developed using different combinations of chlorophyll, sea surface temperature (SST), and salinity derived from remote sensing to estimate DIC, pCO_2, and total alkalinity (Stephens et al. 1995; Sarma 2003; Ono et al. 2004; IOCCG 2006; Shutler et al. 2016).

The DOC pool as a whole does not have a single signal that can be captured by optical remote sensing. However, the coloured dissolved organic matter (CDOM) component can be retrieved based on its absorption spectrum, though this is often combined with detritus (which has a similar spectral signature). There are many algorithms for deriving CDOM, however currently it is most frequently provided as a product in the form of the absorption coefficient of CDOM (and sometimes) detritus at a given wavelength. While CDOM and DOC are largely unrelated at basin/global scales (Siegel et al. 2002), significant relationships are found at regional scales in coastal/shelf areas, in particular close to estuaries (Vantrepotte et al. 2015), leading to regional satellite-derived DOC distributions.

2.3.4   Inherent optical properties

Inherent optical properties (IOPs) only depend on the medium, thus are independent of the ambient light field. Semi-analytical algorithms are developed based on relationships derived from simplifications to the basic radiative transfer equation (i.e., Equation 2.4) (Gordon et al. 1975; Morel 1980; Gordon et al. 1988; Morel and Gentili 1993). Equation 2.4 indicates at each wavelength, $R_{rs}(\lambda)$ is a function of at least three different variables (a_{ph}, a_{dg}, and b_{bp}; or four variables when a_{dg} is split into a_{NAP} and a_{CDOM}). These variables are linked to biogeochemical constituents through their mass-specific IOPs, such as the chlorophyll-a specific absorption and mineral or detrital-specific (back)scattering coefficients. Thus, an inverse solution of Equation 2.4 requires multiple spectral bands, assumptions on component spectral shapes, and accurate models of the primary optical relationships.

Various semi-analytical algorithms have been developed, and a comprehensive review of these can be found in Werdell et al. (2018). Werdell et al. (2018) outline two approaches for retrieving IOPs: 1) using $R_{rs}(\lambda)$ after atmospheric correction and, 2) using $L_t(\lambda)$, thus circumventing the need for atmospheric correction. The former is the approach used in standard processing and will be treated here. Many algorithms have been developed that invert Equation 2.4 to derive IOPs and/or concentrations of constituents, such as [Chl] or [TSM]

(total suspended material concentration). The approaches that these algorithms employ can be divided into two categories: bottom-up strategy (BUS) and top-down strategy (TDS) (Figure 2.9). Both BUS and TDS utilize dependence of bulk IOPs on the spectral shape and magnitude of the three primary components, either in the visible domain (O'Reilly et al. 1998; IOCCG 2006) or extending into the red-infrared region (e.g., Gitelson 1992; Dall'Olmo et al. 2003; Binding et al. 2012; Moses et al. 2012). A BUS algorithm requires bio-optical models for each component during the retrieval process and derives each individual component and the bulk property simultaneously. Thus, a BUS algorithm does not independently retrieve the spectrum of any component (Table 2.1). Conversely, a TDS algorithm retrieves the bulk property first before decomposing it into separate individual components, thus not requiring bio-optical models for each component during the inversion process. This allows a TDS algorithm to retrieve the spectrum of some components independently (Table 2.1), which can be used later to determine various water parameters (e.g., Craig et al., 2006) using additional bio-optical models. Examples of BUS include linear matrix inversion (Hoge and Lyon 1996; Wang 2005; Binding et al. 2012), spectral optimization (e.g., Doerffer and Fisher 1994; Bukata et al. 1995; Roesler and Perry 1995; Lee et al. 1999; Maritorena et al. 2002; Evers-King et al. 2014) and look-up-tables (LUT) (Carder et al. 1991; Mobley et al. 2005). Examples of TDS include the Quasi-Analytical Algorithm (QAA) (Lee et al. 2002), the Plymouth Marine Laboratory (PML) algorithm (Smyth et al. 2006), and the Loisel and Stramski (2000) algorithm based on the diffuse attenuation coefficient.

The mass-specific absorption and scattering of each constituent vary spatially and temporally. Thus, individual IOPs cannot be precisely retrieved from $R_{rs}(\lambda)$ and can result in regional and temporal varying uncertainty. A report summarizing the retrieval of IOPs from $R_{rs}(\lambda)$, and the associated difficulties (IOCCG 2006) concluded that the total absorption and backscattering coefficients are the most reliable parameters that can be retrieved. The absorption spectra of the individual components are not constant and often overlap each other, thereby reducing the accuracy of the retrieved individual absorption coefficients (Smyth et al. 2006; Lee et al. 2010), which ultimately affects the derivation of the in-water constituents.

2.3.5 Attenuation coefficient and euphotic depth

The spectral diffuse attention coefficient, $K_d(\lambda)$, is an apparent optical property (AOP), meaning that it depends on the ambient light field. $K_d(\lambda)$ is often used to estimate the depth of the euphotic zone depth (Z_{eu}) in models of primary production. The first optical depth is regarded as the depth for which light exiting the ocean is able to be measured remotely ($1/K_d(\lambda)$). There are three approaches for estimating $K_d(490)$. The first is based on empirical relationships derived from *in situ* measurements of $K_d(490)$ and blue/green band ratios of $R_{rs}(\lambda)$ (Austin and Petzold 1981; Mueller 2000), or more generally on relationships between $K_d(\lambda)$ and $R_{rs}(\lambda)$ revealed, for instance, by neural networks (Jamet et al. 2012). The second approach uses empirical relationships between K_d and [Chl] (Morel 1988; Morel and Maritorena 2001). The third approach uses a quasi-analytic method that first derives absorption and backscattering from $R_{rs}(\lambda)$ and then uses these coefficients to semi-analytically estimate $K_d(\lambda)$ (Lee et al.

2005b). Lee et al. (2005a) compared algorithms representative of these approaches and found the semi-analytically derived $K_d(\lambda)$ to perform the best for the broadest range of water types.

2.3.6 Primary production

Net primary production (NPP), or the rate of production of organic carbon (referred to here as primary production), has been estimated utilizing a variety of satellite inputs with algorithms that vary considerably in complexity. Two important characteristics define these models: the way they treat light in the water column, and their representation of phytoplankton physiology. As far as light is concerned, the most complete models are wavelength-, depth- and time-resolved (they consider propagation of spectral light through the water column at various times during the day, Behrenfeld and Falkowski 1997a). Satellite-derived photosynthetically available radiation (PAR) is now a standard input in that context. In order to be driven by remote sensing data, algal physiology has been kept relatively simple with a varying degree of empiricism. The first developed and most widely used satellite primary production algorithms are based on satellite derived [Chl] as a measure of phytoplankton biomass (and often accessorily to define light attenuation). In that framework, models have often used [Chl]-normalized photosynthetic parameters for photosynthesis-irradiance relationships. For instance, photosynthetic parameters have been defined as a function of temperature (Antoine and Morel 1996; for the Vertically Generalized Production Model, VGPM Behrenfeld and Falkowski 1997b) or varying with season and biogeochemical provinces (e.g., Sathyendranath et al. 1989, 1991). A variety of models and their performance with respect to field data have been described in Carr et al. (2006) and Friedrichs et al. (2009) and Saba et al. (2011).

Phytoplankton change their cellular [Chl] content to acclimate to their nutrient and light environment, which is challenging to represent in satellite-based algorithms. To reduce the uncertainties in primary production estimates inherent in using [Chl], a carbon-based approach was developed. The carbon-based primary production models utilize satellite-estimated b_{bp}, in addition to [Chl], PAR, K_d and SST, to estimate phytoplankton carbon biomass which is related to growth rates from chlorophyll-to-carbon ratios (Behrenfeld et al. 2005; Westberry et al. 2008). Further, the 'Carbon, Absorption, and Fluorescence Euphotic resolving' (CAFÉ) model has been developed that takes advantage of several recent remote sensing advances allowing for the use of phytoplankton absorption, particulate backscattering and phytoplankton carbon to estimate NPP. Net primary production is calculated by the CAFÉ model as a product of energy absorption and the efficiency by which absorbed energy is converted into carbon biomass, while growth rate is calculated as NPP normalized to carbon biomass (Silsbe et al. 2016).

2.3.7 Phytoplankton functional types

A variety of approaches have emerged that attempt to discriminate 'phytoplankton functional types' (PFTs), which include algorithms retrieving phytoplankton size classes (PSC), phytoplankton taxonomic composition (PTC) or particle size distribution (PSD). In this way, a PFT is an aggregation of phytoplankton, where irrespective of their phylogeny, they share similar biogeochemical or ecological roles. PSC, PTC and PSD serve as a further refinement of PFTs, where the

choice of the considered functional type depends on the question at hand (Mouw et al. 2017). Here we highlight the primary PFT algorithm types and output products (for a comprehensive review, see IOCCG 2014 and Mouw et al. 2017). Determination of phytoplankton functional types has taken advantage of both empirical and semi-analytical approaches discussed above.

PFT algorithms are classified according to their theoretical basis, and include abundance-, radiance-, absorption-, and scattering-based approaches. Abundance-based algorithms use [Chl] as a satellite input and are based on the general observation that in the global open ocean, a change in [Chl] is associated with a change in phytoplankton composition or size structure. Radiance-based algorithms classify PFTs based on the shape and/or magnitude or the satellite-observed $R_{rs}(\lambda)$. Radiance-based approaches assume that, after normalization, changes in radiance coincide with changes in PFT composition, as opposed to other in-water constituents such as CDOM or NAP that may or may not co-vary with phytoplankton. The majority of the existing PFT approaches are absorption-based, where there is some level of dependence on the spectral magnitude or shape of phytoplankton absorption. The magnitude of $a_{ph}(\lambda)$ is related to pigment composition and total pigment concentration, dominated by [Chl] at the peak wavelength (for oceanic waters) of 443 nm. Size information is contained in the absorption spectrum due to pigment packaging (e.g., Bricaud and Morel 1986). Some of the approaches utilize chlorophyll-specific phytoplankton absorption in which phytoplankton absorption is normalized to [Chl] (Bracher et al. 2009; Mouw and Yoder 2010; Sadeghi et al. 2012; Roy et al. 2013), either for a specific wavelength or to derive a spectral shape or slope that is related to second order signals including pigment composition and packaging. Backscattering approaches retrieve information on all particles rather than just phytoplankton. Generally, the backscattering coefficient decreases according to a power law function with increasing wavelength (Montes-Hugo et al. 2008; Kostadinov et al. 2009). Smaller particles have a greater backscattering slope (η) than larger particles (Table 2.1). Recent approaches have sought to include the effects of phytoplankton size and structural properties on both absorption and backscattering properties within semi-analytical, spectral optimization algorithms (Evers-King et al. 2014; Lain et al. 2017).

The satellite phytoplankton functional type algorithm products or outputs vary by the algorithm type. The PSC output is most commonly grouped as pico- (0.2-2 μm), nano- (2-20 μm), and/or microplankton (>20 μm) following the size classification scheme proposed by Sieburth et al. (1978). However, a few models allow for multi-component size classes not constrained by the traditional size groupings (Roy et al. 2013; Brewin et al. 2014). The PSD satellite output (Kostadinov et al. 2009; Kostadinov et al. 2010; Roy et al. 2013) can conform to the Sieburth et al. (1978) size classification. The PTC algorithms (Alvain et al. 2005, 2008; Bracher et al. 2009; Hirata et al. 2011; Sadeghi et al. 2012; Ben Mustapha et al. 2014) have a variety of outputs, dictated largely by the information available from *in situ* calibration and/or validation datasets.

Products that require additional satellite data sets include primary production (see Section 3.3.5) and PFTs that are ecologically based (e.g., Raitsos et al. 2008; Palacz et al. 2013). The ecologically-based PFT estimates require additional physical and spatio-temporal information. Raitsos et al. (2008) incorporated geographical, temporal, biological, physical and bio-optical

information into neutral networks, to discriminate four major functional types (diatoms, dinoflagellates, coccolithophores and silicoflagellates). Likewise, Palacz et al. (2013) used artificial neural networks to estimate four PFTs (diatoms, coccolithophores, cyanobateria, and chlorophytes) from satellite estimated SST, PAR, wind speed, and chlorophyll, and model assimilated mixed layer depth, neglecting geographic and temporal information.

2.3.8   Product availability

Satellite imagery is classified by "levels" which describe the extent of processing that has taken place (Table 2.2). Level-0 data are top-of-atmosphere radiance, raw from the sensor and unprocessed (instrument-specific digital counts and file formats). Level-1 data consist of top-of-atmosphere geophysical units ($L_t(\lambda)$) that are time referenced and radiometrically and geometrically calibrated. Level-2 products consist of atmospherically corrected $R_{rs}(\lambda)$, as well as derived geophysical parameters, such as IOPs, and [Chl], among others. At Level-3, the derived products are aggregated, composited, and/or projected into a spatial grid over a defined time period. Thus, Level-3 products consist of daily, 8-day, monthly, annual, etc. time composites that are projected onto a grid. Level-4 data are model output or results from the analysis from multiple sources of data. Primary productivity is an example of a Level-4 product. In relating these levels to the tiers presented in Figure 2.8, Level-1 and tier-1 are equivalent. However, tiers two through five in Figure 2.8 would fall into Level-2 here, until they are aggregated over space and time, at which time, they would be considered Level-3. Tier six from Figure 2.8 would fall into Level-4 here as they require additional satellite products to be able to be derived.

Table 2.2   Summary of satellite ocean colour imagery processing levels.

Level	Processing	Example Products
0	Raw	Spectral top-of-atmosphere radiance
1	Time referenced, radiometrically and geometrically calibrated TOA radiance	$L_t(\lambda)$
2	Atmospheric correction and algorithms are applied for the derivation of geophysical parameters	$R_{rs}(\lambda)$, [Chl], $K_d(\lambda)$, PIC, POC, PAR, $a(\lambda)$, $b_b(\lambda)$, $a_{ph}(\lambda)$, $a_{dg}(\lambda)$, $b_{bp}(\lambda)$
3	Aggregated and/or projected into a spatial grid over a defined time period	Same as for Level-2, but projected and composited to various time periods: daily, 8-day, monthly, annual, etc.
4	Model outputs or products derived from multiple measurements	Primary Production

The level of maturity of products varies. Considering their historical use and perceived maturity, $R_{rs}(\lambda)$, [Chl], $K_d(490)$, and PAR are usually the primary standard products for all space agencies, while at least some IOPs have been increasingly considered. 'netCDF' is now considered the standard file format for distribution. These products are hosted by the agency/institution responsible for the respective missions/products (Table 2.3). While numerous products can be found at the respective websites in Table 2.3, the mo-

delling community could benefit from improved ease of access of the imagery through the use of services that host a large collection of datasets, such as Environmental Research Divisions Data Access Program (ERDDAP, `https://upwell.pfeg.noaa.gov/erddap/index.html`) and Unidata's Thematic Real-time Environmental Distributed Data Services (THREDDS, `https://www.unidata.ucar.edu/software/thredds/current/tds`).

Table 2.3 Satellite ocean colour imagery online access (note: IOCCG has mission specific data access links at: `http://ioccg.org/resources/data/`).

Agency	Data Distribution	URL

NASA	Ocean Color Web
	
	 Ocean Productivity
		`https://oceancolor.gsfc.nasa.gov/`
		
		 `https://www.science.oregonstate.edu/ocean.productivity/`

NOAA	CoastWatch	`https://coastwatch.noaa.gov/`
EUMETSAT	EUMETSAT broadcast delivery system (EUMETCast)	`https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETCast/`
EUMETSAT	EUMETSAT Data Centre	`https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETSATDataCentre/index.html`
EUMETSAT and ESA	Copernicus Online Data Access (CODA) and SciHub	`https://coda.eumetsat.int` `https://scihub.copernicus.eu/`
European Commission	Copernicus Marine Environmental Monitoring Service	`http://marine.copernicus.eu/`
ESA and PML	Ocean Colour – Climate Change Initiative (OC-CCI)	`https://www.oceancolour.org/`
ACRI-ST	GlobColour	`http://globcolour.info/`

There are also many products that have not yet been incorporated into the standard data distribution streams, such as phytoplankton functional types (see Mouw et al. 2017 for a list of PFT imagery product availability), phytoplankton pigments, and various carbon pools. Many of these are considered developmental and under continued investigation. Some of these imagery products have been deposited into public archives by their developers.

The various space agencies that have launched these missions also maintain software for processing, viewing, analysis and quality control of their missions. These include NASA's SeaWiFS Data Analysis System (SeaDAS, `https://seadas.gsfc.nasa.gov/`) and ESA's Sentinel Application Platform (SNAP, `http://step.esa.int/main/download/`). There is also image processing and visualization software available through the United Nations Educational, Scientific and Cultural Organization's (UNESCO) Bilko (`http://www.bilko.org/`, as well as several online imagery visualization websites, including NOAA's Ocean Colour Viewer (`https://www.star.nesdis.noaa.gov/sod/mecb/color/`), NASA's Goddard Earth Sciences Data and Information Services Interactive Online Visualization and Analysis Infrastructure (GIOVANNI, `https://giovanni.gsfc.nasa.gov/giovanni/`), NASA's State of the Ocean (`https://podaac-tools.jpl.nasa.gov/soto`), NASA's Worldview (`https://worldview.earthdata.nasa.gov/`), and the on-line tools from the European Space Agency's Ocean Colour-Climate Change Initiative (`https://www.oceancolour.org/`), among others.

2.3.9   Product Selection

How does a user choose the best satellite imagery product for a given region and/or application? Standard products i.e., those that space agencies routinely provide at Level-2 and -3, are always a safe selection for broad scale applications as they have undergone the most extensive validation. However, this does not mean they are without uncertainties (see Section 2.4). When working regionally, or in coastal and inland waters, regionally tuned and validated products will often have less uncertainties and are recommended if a study is specific only to that region. However, access to these products are not uniform across various regions, thus requiring specialized data processing by the end user.

When confronted with the choice of different products, a user will likely consider the characteristics that are best for the specific application, for instance, spectral properties or time/space resolution and coverage. For example, for event-scale processes such as coastal algal blooms, a daily Level-1 or -2 product may be most appropriate, while for time series analysis of a large-scale climate phenomenon, a merged Level-3 data set may be a better choice. It is also worth considering the assumptions that are underlying all ocean colour products, since these might be a limiting factor in downstream applications. For instance, spectral relationships embedded in algorithms might restrict the possibility of interpreting products for the detection of water types or PFTs. Some modellers might favor suites of products that are fully consistent (for instance, respecting the conservation of energy). Researchers involved in climate research might refrain from using merged products when the merging procedure does not account for inter-mission biases. Algorithms that have gone through a thorough validation process and products going through quality checks should certainly be preferred. For the same reasons, the availability of uncertainty estimates associated with products will likely become a factor of choice, and a requirement for some applications. In that context, high uncertainty estimates do not conflict with a high quality for the product itself if a complete and realistic uncertainty budget has been constructed. Finally, products with a full and clear documentation that are readily accessible in known formats should have a competitive advantage, while dialogue between data providers and users might also direct a proper selection of products.

Given each ocean colour mission sensor has a finite operational lifetime, combining different ocean colour missions to produce a consistent time series is crucial for any long-term study. Each mission differs in their design, capability, and sampling. Different satellite ocean colour sensors have, for example, different band locations, band widths, radiometric sensitivities and orbit. For standard products, there have been efforts to merge imagery from multiple platforms. This allows for reduced impact of cloud cover and better statistical representation when multiple days or longer are composited together. The Ocean Colour – Climate Change Initiative (OC–CCI, Table 2.3) provides multi-sensor (MERIS, Aqua-MODIS, SeaWiFS and VIIRS) merged, consistent, uncertainty-characterized, global satellite essential climate data products. Other predecessor efforts to merge multiple missions included GlobColour (Table 2.3) and MEaSURES (Making Earth Science Data Records for Use in Research Environments, http://wiki.icess.ucsb.edu/measures). MEaSURES merged SeaWiFS, MERIS and Aqua-MODIS and is

periodically updated. GlobColour is active and merges SeaWiFS, MERIS, Aqua-MODIS, VIIRS, and OLCI over a variety of temporal and spatial resolutions. However, there are challenges when merging several missions together. For example, SeaWiFS, MODIS, MERIS, and VIIRS can provide ~20 years of chlorophyll data. However, each instrument has different mission and sensor designs. These include different band locations and widths, radiometric sensitivities and capabilities, as well as orbital crossing times. These differences make the research that relies on several sensors, including the detection of trends, very challenging.

The effects that discrepancies between missions have on the detection of trends in chlorophyll have been noted in several studies (e.g., Gregg and Casey 2010; Mélin 2016). Gregg and Casey (2010) showed that global median chlorophyll from SeaWiFS and MODIS differed by 12.2% and that these discrepancies exceeded the maximum observed interannual variability globally, and in major oceanographic regions. Mélin (2016) found that the assessment of trends can be impacted by drift in the chlorophyll time series resulting from sensor functions. By comparing the slopes of the linear regressions from varying levels of inter-mission bias and drift between SeaWiFS and MODIS for a 15-year period, they found that a threshold on bias exceeding 2% can lead to error in trend detection. They also noted that these findings are regionally specific with low chlorophyll regions having particularly high sensitivity to inconsistency between sensors.
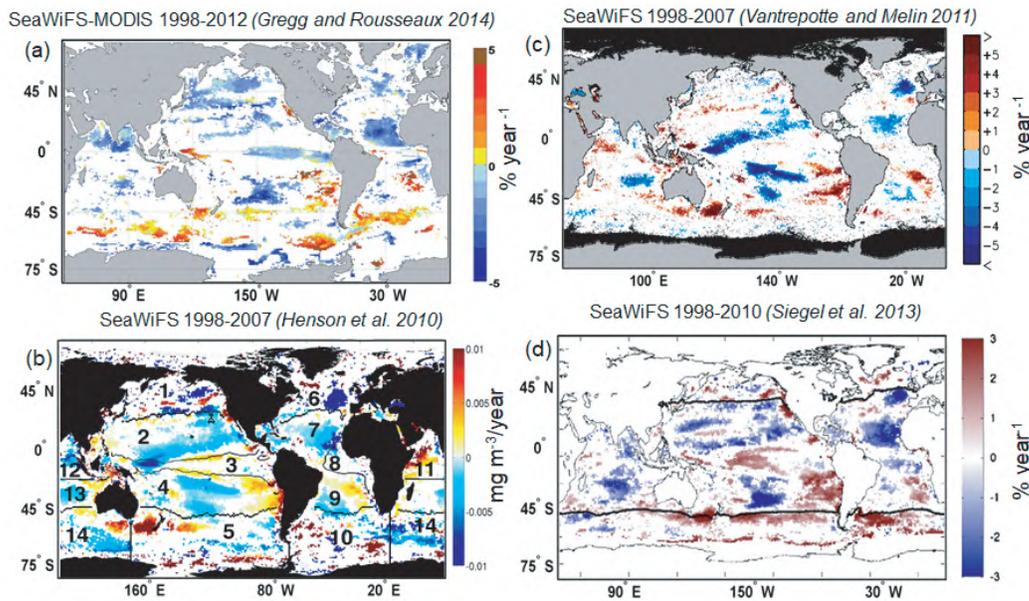
Figure 2.10 Comparison of four representations of chlorophyll trend maps in units of percent per year (except lower left which is mg m^{-3} yr^{-1}). Redrawn using data from Gregg and Rousseaux (2014).

Previous studies have attempted to create consistent multi-mission chlorophyll time series (Gregg and Conkright 2002; Antoine et al. 2005; Martinez et al. 2009; Mélin 2016). For example, Gregg and Rousseaux (2014) integrated three diverse methodologies to provide a consistent and complete global representation, free of sampling biases. They combined ocean colour

data from multiple satellites (SeaWiFS, MODIS-Aqua), bias correction methods based on *in situ* data, and data assimilation. The bias correction approach, the Empirical Satellite Radiance-In Situ Data (ESRID) approach, uses relationships between satellite water leaving radiances and *in situ* chlorophyll data. ESRID applies the standard processing bio-optical algorithm after processing completion, using satellite water leaving radiances, and *in situ* chlorophyll. Using this method, Gregg and Rousseaux (2014) demonstrated that this approach was one way to provide a consistent, multi-mission, satellite ocean colour record, and that the trends they observed agreed with previous studies (Figure 2.10).

2.4   Uncertainty

As described in the previous sections, ocean colour remote sensing functions as a multi-tiered system where uncertainties tend to accumulate at each processing step (e.g., Matthews et al. 2013, their Figure 2.7), from the calibration of the radiometric signal into the TOA radiance, $L_t(\lambda)$, the calculation of the remote sensing reflectance, $R_{rs}(\lambda)$, after atmospheric correction, to the application of algorithms computing optical properties and concentrations of optically significant constituents or more advanced products such as primary production or PFTs (Sections 3.1 and 3.3). IOCCG (2019) provides a complete review of the many sources of uncertainties affecting ocean colour products, a summary of which is given here. The main contributors to uncertainty are from an imperfect atmospheric correction and the relationships between $R_{rs}(\lambda)$ and IOPs, both of which are further exasperated in optically-complex coastal waters.

Unaccounted atmosphere effects are a major source of uncertainty in TOA radiance. The process of atmospheric correction relies on a faithful representation (optical properties, vertical structure) of atmospheric constituents (gas, aerosols) and identification of challenging conditions such as the presence of clouds, either obstructing the field-of-view or affecting neighboring pixels (stray light, shadows). The atmospheric correction must also deal with specific characteristics of the water surface, such as specular reflection (sun glint) or the presence of white caps and bubbles. The presence of sea ice, floating macroalgae, post-bloom surface scums or coccoliths are other phenomena that might affect surface conditions while being challenging to interpret with standard algorithms. The atmospheric correction procedure in the nearshore is further complicated by the potential presence of absorbing aerosols such as smoke, dust, and anthropogenic emissions (e.g., NO_2 and CO_2, Tzortziou et al. 2013) and by possible adjacency effects due to land (Bulgarelli et al. 2014), and the impact of the bottom in optically shallow waters.

The relation between $R_{rs}(\lambda)$ and IOPs through analytical expressions (such as Equation 2.4) are simplifications relying on various assumptions. Bulk IOPs result from the optical properties associated with a constantly evolving myriad of particles of complex shapes and of dissolved samples, and are therefore a very simplified representation of nature. Even though large progress has been achieved, a more detailed budget of backscattering by water constituents covering orders of magnitude in size, is still needed (Stramski et al. 2004). Bio-optical algorithms

often rely on assumptions about the relations between IOPs and concentrations of optically significant constituents (e.g., the link between $a_{ph}(\lambda)$ and [Chl], or between $b_{bp}(\lambda)$ and [TSM]) and about the spectral shape of IOPs. There are large variations in these model parameters in natural waters, as a function of particle characteristics, chemical composition, algal species and status. An abundant literature with descriptions of specific optical parameterizations for the various constituents exists (see reviews in Blough and Del Vecchio 2002; Twardowski et al. 2004 or Berthon et al. 2008). For the same reasons, purely empirical algorithms relating $R_{rs}(\lambda)$ to concentrations of optically significant constituents (like [Chl] or IOPs) are not able to capture the full extent of natural variability.

Coastal and inland waters present specific challenges (Mouw et al. 2015): in these regions, optically significant constituents (OSC) often vary independently, requiring improved spectral and radiometric resolutions, while physical drivers such as tides and geographic boundaries set up different spatial and temporal scales compared to the open ocean, requiring improved spatial resolution than what is currently provided by existing space-based assets. The independently varying optical constituents also present challenges for algorithms. Empirical algorithms often assume the targeted parameters being retrieved co-vary with the other OSCs present. Semi-analytical algorithms are more flexible to consider independent changes between parameters, but there are still empirical coefficients associated with the underlying optical relationships that may not represent well the variability of the system. Thus, algorithms developed for coastal and inland waters need to consider the regional and temporal differences in mass-specific optical properties. With this in mind, blended approaches that utilize a variety of algorithms and choose those that produce the least uncertainty in various water types are advantageous in optically diverse waters (Moore et al. 2001, 2014; Palmer et al. 2015).

Additionally, there are uncertainties associated with the editing process happening between Level-2 and Level-3. When binned into a grid-sized point, pixels might cover only a part of the related surface, limiting the information content. The same phenomenon is at play when combining relatively high-resolution data onto coarser grids. Locally, daily Level-3 data are often based on one satellite pass (for polar-orbiting platforms), which in reality is a snapshot (around local noon) of the period. Finally, the creation of a time composite (typically weekly or monthly) often relies on a limited number of so-called daily data, leaving open the question of what the time composite would be if it were constructed with a full temporal sampling. So, while the process of averaging data tends to reduce the uncertainties at Level-3 (averaging out non-systematic errors), the way incomplete information is combined in space and time might increase uncertainties. Finally, when combining data from different missions, systematic effects at the level of a whole mission may introduce artefacts in the merged series and impact trend analyses (Mélin 2016). Overall, uncertainty propagation in the creation of Level-3 data is a complex and insufficiently characterized process that requires an extensive knowledge of covariance fields.

So far, estimates of ocean colour data uncertainties have mostly depended on validation studies where satellite products are compared with field observations through a well-defined protocol (see IOCCG 2019). A fairly robust body of knowledge has been accumulated on validation results even though the global ocean is still unevenly sampled, particularly for

AOPs and IOPs. As for $R_{rs}(\lambda)$, root-mean-square (RMS) differences between field and satellite data usually decrease with wavelength, with values in the blue mostly in the range 0.0008 – 0.0025 sr^{-1} (Mélin and Franz 2014). Mean absolute relative differences show large variations, only rarely close to 10%, but sometimes exceeding 10's of %. Spectrally, they usually show a horse-shoe shape, with values in the blue and red bands inversely related to $R_{rs}(\lambda)$ (e.g., being high in the blue for low-R_{rs} absorbing waters). Validation studies at the scale of the global ocean have converged on RMS differences of the order of 0.3 (in log$_{10}$ units) for [Chl] from standard algorithms (Gregg and Casey 2004) but these results exhibit spatial variations as Chl-specific IOPs vary across regions (Szeto et al. 2011). An abundance of literature shows that differences can be much higher in coastal, optically-complex, regions. Results obtained for IOPs depend on the IOP and wavelength considered together with the water type. In a large validation study, Brewin et al. (2015) documented RMS differences between log-transformed *in situ* data and algorithm outputs of approximately 0.2 at 412 nm and 0.15 at 490 nm for the absorption coefficients, with higher differences when total absorption is partitioned in component absorption. It is important to note that validation statistics provide imperfect estimates of uncertainties, considering that they are influenced by the issue of representativeness (satellite and field data may be registered at different times and differ largely in their spatial scales) and by the uncertainties of the field data.

It is now well accepted that ocean colour products should be accompanied by uncertainty estimates (and several products are now hosted with associated estimates) and there is currently a strong emphasis on development of relevant techniques and the integration of metrological principles in this field. At present, the available approaches rely on limiting assumptions and consider only some of the sources of uncertainties. To reach a full uncertainty budget with traceability to standards, all the sources should be included and the various uncertainty contributions should be properly distinguished to allow the propagation of the uncertainty fields to higher level products (Level-3 or -4). The amount of information needed to describe the uncertainty fields to users varies across applications. An uncertainty estimate informs on "the dispersion of the values that could reasonably be attributed to a measurand" (GUM 2008) so that its description can go from a single value to a complete probability distribution function. Similarly, the treatment of uncertainty propagation in space and time may require covariance information, which may induce a large increase in data-file size. Biogeochemical modelling is likely to be among the user communities most interested in a fairly complete description of uncertainty fields, for instance for the purpose of data assimilation (e.g., Ford and Barciela 2017).

2.5 Future Capability

2.5.1 Mission capability

As we look to the future, the most recently launched missions, the Sentinel-3 platforms with the Ocean and Land Colour Instrument (OLCI), will be depended upon as the workhorse ocean colour missions as other assets on orbit age out. It is anticipated that our ocean colour sensing

ability into the 2030's, and potentially beyond, will be reliant on OLCI from Sentinels-3A to -3D (Sentinel-3A and -3B were launched February 2016 and April 2018, respectively with two more flight units, Sentinel-3C and -3D, scheduled). Beyond these current and planned Sentinel-3 platforms, there are several future missions planned (see Table 6 in Werdell et al. (2018) and `http://ioccg.org/resources/missions-instruments/scheduled-ocean-colour-sensors/`) that expand spectral, spatial, and temporal shortcomings of the current instrument suite.

Ocean colour missions fall under three categories: 1) high spatial, 2) moderate spatial, and 3) geostationary. High spatial resolution sensors are polar orbiting and provide global coverage, but at infrequent repeat times (10 days or more). Moderate resolution sensors are also polar orbiting and have comprised the majority of past ocean colour missions. Geostationary missions only observe a particular region of the Earth, but at multiple times per day, allowing investigations into diurnal processes. The planned Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission will extend the current moderate spatial resolution, polar-orbiting capability by substantially increasing spectral resolution. The timeframe for the start of the PACE mission is anticipated to be in the 2022/2023 timeframe.

There were missions in planning that greatly expanded temporal and spatial resolution, including the GEOstationary Coastal and Air Pollution Events (GEO-CAPE) mission (95°W, 375 m resolution, possibly hyperspectral, Fishman et al. 2012; Salisbury et al. 2017) that was envisioned to greatly expand temporal resolution, in addition to expanding the number of spectral bands, allowing for multiple images of the same location several times per day. This is highly important for capturing the short-term dynamics found in coastal and inland environments. The Korean COMS mission carrying the Geostationary Ocean Colour Imager (GOCI), centered over 130°E, has demonstrated significant improvement in observing diurnal variability with a similar band suite as many of the heritage sensors (Wang et al. 2013). A second GOCI instrument is planned for launch in early 2020. The Hyperspectral Infrared Imager (HyspIRI) was also in the planning phase and was envisioned to greatly expand both spatial (60 m) and spectral resolution (hyperspectral), but at a 16-day revisit (Devred et al. 2013), allowing it to be most valuable for observing slow-changing properties such as benthic type and bottom depth (Figure 2.4). The planned implementation of the GEO-CAPE and HyspIRI missions has halted and the missions are being reimagined under the broad umbrella of surface biology and geology guided by the recommendations of the most recent Decadal Survey for Earth Observations from Space (National Academies of Sciences and Medicine 2018). Other high spatial resolution missions that are proceeding with planning include a second OLI sensor on Landsat-9, and the German mission 'Environmental Monitoring and Analysis Program' (EnMAP), with 30 m spatial resolution.

2.5.2   Future products

With the anticipated expansion of spectral, spatial, and possibly temporal resolution, an expansion and refinement of products is anticipated. Many algorithm approaches are at an advanced stage where they could be more readily included into standard processing, such as POC concentrations, and merged [Chl] based on optical water types. It is anticipated that

future mature products will allow for improved/expanded phytoplankton functional types and pigment discrimination with improved spatial resolution. Additionally, parameters related to phytoplankton physiology are anticipated to expand and improve as a result of expanded spectral resolution near the region of phytoplankton fluorescence. Improved uncertainty in a_{CDOM} products, and the ability to discriminate the spectral slope of a_{CDOM} (Grunert et al. 2019), is anticipated by expanding spectral observations to shorter wavelengths. This will also result in improved discrimination of all other IOPs and products further downstream, such as phytoplankton functional types, and pigments and carbon pool products (such as PIC and phytoplankton carbon concentrations). The uncertainty of all products is anticipated to improve.

2.6   Recommendations

The following recommendations are anticipated to better facilitate the ease of product access, product selection, and new product development that can be of use to the modelling community:

❖ Continued development and support of future satellite missions that expand spectral, spatial and temporal resolution, allowing for product improvement and new product development.

❖ Expand merged products, not only to merge products in time, but also in space (i.e., blended products where the most suitable regional algorithm is selected and combined to form a single output).

❖ Improve the ease of access of the imagery through the use of Environmental Research Divisions Data Access Program (ERDDAP, `https://upwell.pfeg.noaa.gov/erddap/index.html`) and Unidata's Thematic Real-time Environmental Distributed Data Services (THREDDS, `https://www.unidata.ucar.edu/software/thredds/current/tds/`).

❖ Encourage greater sharing of community-derived custom data products through open data principles and services.

❖ Provide greater transparency on the products through more documentation.

❖ Develop ocean colour methods and algorithms that employ additional hydrographic and biogeophysical information on environmental conditions.

❖ Provide additional information on data product uncertainties.

Biogeochemical And Ecosystem Models: What Are They And How Can They Be Used?

Stephanie Dutkiewicz, Cecile Rousseaux, Stefano Ciavatta, Oliver Jahn, Charles A. Stock, Mark Baird, Fei Chai, Barbara Muhling and Marion Gehlen

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

The purpose of this chapter is to introduce scientists (particularly the ocean colour community) to the theory, practicalities of implementation and uses of biogeochemical/ecosystem models. We attempt to do this with limited amount of jargon. We include a table (Appendix 3) with definitions of some of the field's technical phrases and acronyms (some of which we cannot avoid using in this chapter and elsewhere in the report). Our treatment of the subject is, by design, brief so as not to be overwhelming, and is therefore not intended to be exhaustive.

We have used "biogeochemical" and "ecosystem" modelling in the report title to signify that we are encompassing models with different interests. In the modelling community, "biogeochemical models" are distinguished from "ecological models" by their focus on the processes that are most relevant: biogeochemical cycles as opposed to biology/ecology of organisms. This is, however, a blurry delineation, as ecological models require the underlying biogeochemistry, and biogeochemistry models require at least some parameterization of ecosystem. Biogeochemical models may have more focus on the cycling of elements such as carbon (e.g., air-sea flux of carbon dioxide). Ecosystem models might ask questions such as why different organisms live in specific geographic locations, why certain types can live together and, how biodiversity is controlled, and what controls global fish distributions.

We provide a brief overview of models (Section 3.1), followed by a more in depth section on the concepts, and basic building blocks of biogeochemical/ecosystem models (Section 3.2). Section 3.3 deals with how modellers have treated natural light, and, in particular, some of the new developments to include radiative transfer calculations, absorption and scattering, and spectral-resolved light. Section 3.4 provides a brief description of some of the types and uses of biogeochemical/ecosystem models. The final Section (3.5) provides some insight into choosing models and model output for specific studies.

3.1 Introducing the Theory, Practicalities of Implementation and Uses of Biogeochemical Models

Marine biogeochemical and ecosystem models are composed of various components covering different aspects of the natural systems they capture (Figure 3.1). Most biogeochemical and ecosystem models include the physical environment (temperature, salinity and circulation of the ocean), the cycling of inorganic and detrital matter (biogeochemistry), and the explicit representation of some portion of the living component of the ocean (e.g., phytoplankton, zooplankton). In this report we will refer to models

that include the physics, chemistry and biology of the ocean as biogeochemical and/or ecosystem models. This is by no means a unique definition, but one that we use here to distinguish from other models such as habitat models.
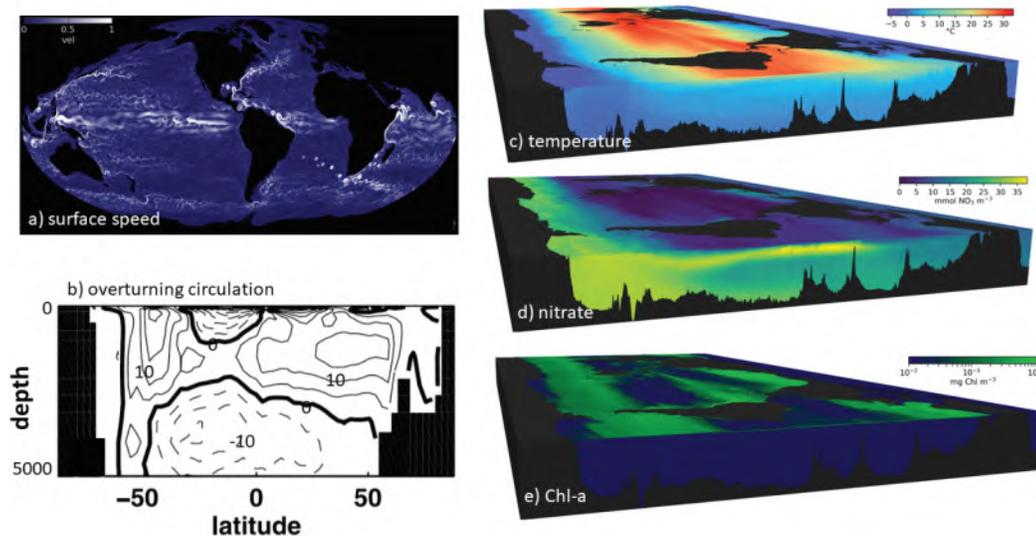
Figure 3.1 Anatomy of a biogeochemical/ecosystem model. Models capture the 3-dimensional physics (including the flow of water), altering distribution of inorganic nutrients (such as nitrate) and components of the marine ecosystem (e.g., phytoplankton biomass and Chl distributions). (a) surface speed (m s^{-1}) showing currents and eddies; (b) global integrated overturning circulation (Sv) showing the importance of the vertical flow in transporting mass with depth and latitude; (c) temperature ($^\circ$C); (d) nitrate (mmol N m^{-3}); (e) Chl-a (mg m^{-3}) with the deep maximum showing as a thin line at about 100–150 m through the tropical regions. (c)–(e) illustrate the 3-dimensional nature of the ocean that cannot as yet be captured from satellites. The model highlighted here has physical fields (e.g., temperature, speeds) from the project for Estimating the Circulation and Climate of the Ocean (ECCO2 physical model at 18 km resolution, Menemenlis et al. 2008). These physical fields (as well as explicit three-dimensional velocities and mixing variables) are used to move and mix the biogeochemical (nutrients, organic matter) and ecosystem (diverse plankton community) variables.

The circulation of the ocean is essential for re-distributing the inorganic and organic pools. As such biogeochemical/ecosystem models must include a representation of currents, mixing, temperature, salinity and density. Patterns of upwelling and vertical mixing are particularly important for controlling the supply of nutrient-rich deep waters to the well-lit euphotic zone to fuel phytoplankton growth (see e.g., Figure 3.1d).

Biogeochemistry is formally defined as the quantification of chemical species exchanged between reservoirs of the Earth system, along with transformations within these reservoirs, both being mediated by biological activity. Cast within the context of the marine environment, the discipline focuses on the cycling of carbon and nutrients (Si, N, P, Fe) between the living and dead organic and inorganic compartments of the ocean.

Ecosystem components of biogeochemical/ecosystem models represent how inorganic matter is converted into organic matter, and most resolve at least part of the marine foodweb. The ocean biology re-positions elements such as C, N, Si, P by consuming them as nutrients in the surface and sinking as part of the detrital matter to depth where they are remineralized (i.e., returned to inorganic form). Some models parameterize this "biological pump" based solely on this re-positioning of inorganic nutrients rather than incorporating explicit food webs (e.g., Matsumoto et al. 2004; McKinley et al.

2004). However, most models specifically consider the flow of elements (e.g., nitrogen) as they go from inorganic nutrients (N), to phytoplankton (P), to zooplankton (Z), how organic matter lands in a non-living organic pool (detritus, D), and how it is remineralized back to the inorganic pool. These are known as NPZD models (Riley 1946; Fasham et al. 1990; Franks 2002) (Figure 3.2).
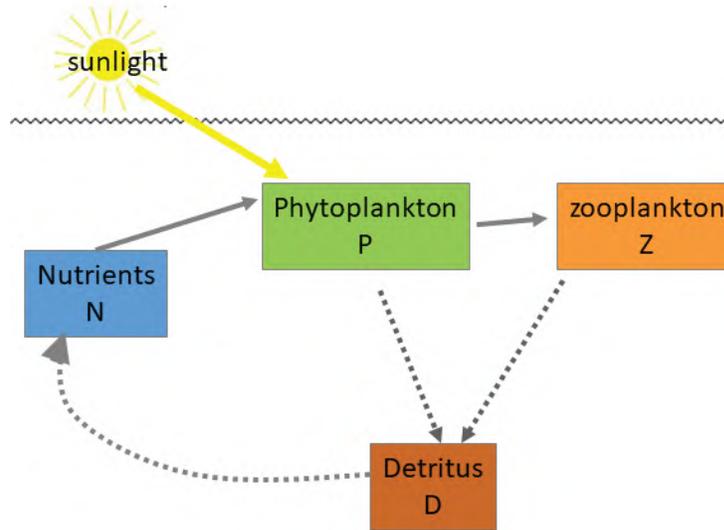
Figure 3.2   Schematic of typical ecosystem components of models. Matter is followed from an inorganic pool (nutrients, N) while it is taken up by phytoplankton (P), which are in turn consumed by zooplankton (Z). Both phytoplankton and zooplankton produce non-living organic matter (detritus, D) through excretion, death and sloppy feeding. Detritus is remineralized back to the inorganic nutrient pool.

With increased computer power and better understanding of the processes, there has been an evolution of models to include multiple nutrients, multiple phytoplankton types (e.g., Chai et al. 2002; Aumont et al. 2003; Moore et al. 2004; Follows et al. 2007), bacterial and virus dynamics to better resolve microbial food webs (e.g., Weitz et al. 2015; Zakem et al. 2018), and multiple zooplankton types to better resolve grazing controls on phytoplankton and energy flow from phytoplankton to fish (e.g., Stock et al. 2014b). More recently, physics, plankton and fish/fisheries models have been combined to form "end-to-end" models (Rose et al. 2010), although often outputs from a biogeochemical/ecosystem model are used to drive fish models (e.g., Cheung et al. 2010; Lefort et al. 2015; Watson et al. 2015). There are, however, a growing number of fully integrated plankton and fish food web models (e.g., Kearney et al. 2012; Aumont et al. 2018). In this report we will not focus on these latter end-to-end fish models, but rather on those that capture only the biogeochemistry and the lower trophic levels (see Section 3.4).

Biogeochemical/ecosystem models often employ a highly simplified grouping of organisms which perform major functions in biogeochemical cycles. These groupings are often referred to as plankton functional types (PFTs) (Baretta et al. 1995; Le Quéré et al. 2005; Hood et al. 2006; IOCCG 2014). Organism size is frequently used to differentiate between different PFTs. Examples of typical PFTs include calcifying phytoplankton (e.g., coccolithophores), silicifying phytoplankton (diatoms), nitrogen fixers (diazotrophs) and small and large zooplankton. Models which require more diversity frequently adopt a trait-based approach (see e.g., Barton et al. 2016). A trait is any quantitative organism characteristic that affects growth, reproduction or survival (e.g., resource acquisition and growth rates) that can be compared across species (McGill et al. 2006). Size is a specific trait that can be used in models (e.g., Ward et al. 2012). The fitness of different organisms along environmental or biological

gradients is tied mechanistically to their functional traits and their trade-offs.

Many biogeochemical and/or ecosystem models use the standard Redfield elemental ratios (C:N:P = 106:16:1) of organisms and fluxes within the marine food-web (Redfield 1934; Redfield et al. 1963). This simplification allows models to use a single "currency" (e.g., carbon, nitrogen or phosphorus). Uptake of other elements are assumed to be in the Redfield ratio with this base currency. This greatly reduces computational costs. However, there are significant deviations around this Redfield stoichiometry (Martiny et al. 2013). Better understanding of the drivers of these deviations has enabled the development of a growing number models with dynamic stoichiometry (Baretta-Bekker et al. 1997; Pahlow and Oschlies 2009; Ward et al. 2012; Daines et al. 2014). This dynamism comes with a computational cost: a tracer is needed for each element within each plankton type, rather than tracking a single nutrient currency for each. Chlorophyll is usually also included within the models, often with dynamic chlorophyll-to-carbon ratios (Geider et al. 1997; Geider et al. 1998), often specifically so this model output can be compared to ocean colour products (see Chapter 4).

The level of complexity of each biogeochemical/ecosystem model primarily reflects 1) the objectives of the model; 2) computational constraints; and 3) constraints imposed by the boundaries of ecosystem science and understanding. Most of the ecological resolution will be allocated to those processes relevant to the primary objective of the model (deYoung et al. 2004). For example, a model designed to understand phytoplankton biodiversity may include dozens of phytoplankton types (e.g., Follows et al. 2007). Those designed primarily for multi-century global biogeochemical simulations will often invest more on resolution of nutrient cycling, with more simplified representations of plankton food webs (e.g., Dunne et al. 2013; Moore et al. 2013). The appropriate level of model complexity is subject to debate (Anderson 2005; Flynn 2005), though rigorous model skill assessment against observation provides one measure of success. Chapter 5 will discuss the crucial role that ocean colour radiometry has played in such skill assessment.

3.2 Concepts, Equations, Code and Computers

Though most oceanographers are aware of, use results from, and often cite work by biogeochemical and ecosystem models, such models often remain a "black box" to much of the community. Here we very briefly unpack this black box.

In general, models are formulated as a set of equations that specify how each component of the model (e.g. biomass of phytoplankton) will change with time as a consequence of the combined effects of physical, chemical and biological factors (see Figure 3.3). These equations are then adapted to computer code such that that variable can be stepped forward in time. The many computations to track how the variable changes over days/year/centuries is done by a computer or cluster of computers. Output from a simulation are concentrations of the modelled variables (e.g., biomass) either instantaneously, or as means over specific time periods (daily/monthly). Rate of change of concentrations can also be provided as model output. Below we provide a very short introduction to some of the specifics of these steps.

3.2.1 Equations, parameters and state variables

The biogeochemical component of a model captures the flow of matter and energy between inorganic nutrients and sunlight into organic matter (phytoplankton biomass) and back to inorganic nutrients. The ecosystem component follows organic matter through (parts of) the food chain (see e.g., Figure 3.2). Models are based on equations that parameterize this flow. For instance, the rate at which phytoplankton biomass changes at any location can be written as (for more explicit representation see Equation 3.2 in Box 3.1):

change in phytoplankton biomass = + growth – losses ± transport/mixing

Phytoplankton growth can be a function of the available nutrients, light and temperature. Losses can include grazing by zooplankton, sinking, viral lysis, cell death and leakage from cells. Transport and mixing refers to the process whereby biomass of the phytoplankton is removed or added at that location by ocean physics (current flow, vertical mixing). Changes in season, interannual variability, and mesoscale processes will lead to variation in the magnitude of the terms on the right of the equation and thus resulting in changes in the phytoplankton biomass. However, in the long term (i.e., years), and with no change in climate, the terms on the right should balance, and a "steady state" should be achieved. Similarly, the rate of change in biomass of a zooplankton can be written as (see Equation 3.3 in Box 3.1):

change in zooplankton biomass = + assimilation of grazed phytoplankton – losses ± transport/mixing

Grazing is a function of the available phytoplankton and how quickly and efficiently the zooplankton can capture and process a phytoplankton. Loss terms can include zooplankton being eaten and death. Additional equations can be written for the consumption and sources of nutrients, as well as the production and remineralization of detrital matter. Some models explicitly include heterotrophic bacteria, instead of parameterizing demineralization, and in these cases there is an equation for them as well.

Figure 3.3 Schematic on developing and using a numerical model. First the set of appropriate equations are devised for all model state variables (e.g., for each box of the NPZD model, see Figure 3.2). Next the equations are written as computer code that is made into an "executable file" that a computer (or cluster of computers) will run. Output, in terms of concentration of variables (such as biomass), are provided for analysis (and maps as displayed here). The model output is 3-dimensional and varies with time.

We refer to the pools of matter that each equation details (e.g., nutrients, phytoplankton biomass, zooplankton biomass) as "state variables". The equations control how matter (and hence also energy) are moved from one pool to another or, in model jargon, how matter moves "from one state variable to another". Box 3.1 below provides sample equations and a further (though still non-exhaustive) discussion of the parameterizations often used in models.

BOX 3.1. NPZD MODEL EQUATIONS. Simple versions of the equations used in a NPZD model

In this box we provide a very generalized example of the types of equations used in NPZD-type models, elaborating with a more formal mathematical approach than the discussion provided in Section 3.2.1. Here N represents a limiting nutrient (e.g. nitrate, phosphate, iron), P phytoplankton biomass, Z zooplankton biomass, and D detrital matter (dead organic matter).

$$\frac{\partial N}{\partial t} = -\mu_{max}\frac{I}{I+I_o}\frac{N}{N+k_N}P - r_D D - \nabla \cdot (\mathbf{u}N) + \nabla \cdot (K\nabla N) \tag{3.1}$$

$$\frac{\partial P}{\partial t} = +\mu_{max}\frac{I}{I+I_o}\frac{I}{N+k_N}P - g_{max}\frac{P^n}{P^n+k_P}Z - m_P P - \nabla \cdot (\mathbf{u}P) + \nabla \cdot (K\nabla P) \tag{3.2}$$

$$\frac{\partial Z}{\partial t} = +\gamma g_{max}\frac{P^n}{P^n+k_P}Z - m_Z Z - \nabla \cdot (\mathbf{u}Z) + \nabla \cdot (K\nabla Z) \tag{3.3}$$

$$\frac{\partial D}{\partial t} = -r_D D + m_P P + m_Z Z + (1-\gamma)g_{max}\frac{P^n}{P^n+k_P}Z - \nabla \cdot (\mathbf{u}D) + \nabla \cdot (K\nabla D) \tag{3.4}$$

The last two terms in these equations represent the advection of the state variables by the three dimensional fluid flow $\mathbf{u}$, and diffusion by sub-gridscale processes, K. The rest of the terms are representations of the different biological processes.

In these equations, growth of phytoplankton is controlled by light I and nutrients N. Frequently these are assumed to be saturating function (i.e., asymptoting to a maximum value at high I or N). I_o and k_N are the values where the respective functions are half their maximum value. k_N is called the nutrient half saturation constant. For the nutrient response, this is called a Michaelis-Menten functional form. Maximum growth rate is set by μ_{max}. In most models the incoming light (I_o) is time varying to include the impact of seasons.

Grazing of phytoplankton by zooplankton, is also often represented as a saturating function with a maximum grazing rate of g_{max} at high phytoplankton biomass. A refuge from grazing at low phytoplankton biomass can be parameterized depending on the value of the exponent n. When $n = 1$ this is called a Holling type II function, and when $n = 2$ it is called a Holling type III function (Holling 1959). Only some of the grazed organic matter, γ, is assumed to be assimilated by the zooplankton (called "sloppy feeding"), the rest enters the detrital pool.

Other loss rates from P and Z (potentially cell death or viral lysis) are often represented as linear functions and they supply dead organic matter to the detrital pool. In models which do not resolve the heterotrophic bacteria, detrital matter is often assumed to remineralize back to inorganic N at a constant rate r_N.

There is considerable debate, based on theory and observations, on the best formulations for the processes discussed above, particularly for the treatment of light limitation and grazing. As mentioned above, many models include more than one nutrient, phytoplankton and zooplankton and might include bacteria rather than using a remineralization timescales. There is a long literature on the different functional forms that the various parameterization should take (e.g., Franks 2002; Vallina et al. 2014) and for additional complexity of the phytoplankton groups (Le Quéré et al. 2005; Follows and Dutkiewicz 2011). Where possible, the parameters mentioned above (e.g., maximum growth, grazing, remineralization rates and half saturation constants) are constrained by laboratory and field experiments.

3.2.2 Grids, resolution, spatial scales

The equations discussed in Section 3.2.1, and more completely in Box 3.1, are solved at only a finite number of locations in the model domain (see Figure 3.4). The model domain is divided into a number of grid cells ("boxes" in Figure 3.4) and the biomass, nutrients, detritus (and any other model variables) are assumed to be homogenous within each grid cell. The size of the grid cells defines the model spatial (vertical and horizontal) "resolution". The smaller the size of the grids, the higher the resolution, and the more physical (and hence biological) processes can be captured. But small grid cells mean that there needs to be many more of them to cover the same region than if using larger grid cells. This translates into higher computational cost: processing of more grid cells requires more Central Processing Units (CPU), and more "wall time" to run the model. Thus spatial (vertical and horizontal) resolution is decided as a compromise between computational cost and processes resolved, and based on the questions to be addressed. A coastal study looking at sharp fronts will need much higher horizontal and vertical resolution than a climate model looking at large scale processes. For open ocean global scale models, "high resolution" is grid cells of the order 10 km in latitude and longitude (~1/12deg) and several meters thick in the vertical. Most climate scale models have coarser resolution with grid cells covering 100 km horizontally and 10+ m in the vertical. Coastal models can have grid cells in the order 1 km or even finer, but only cover a limited region (e.g., 100s kms). The same model can have different resolution in different locations — for instance higher resolution in locations where the processes of interest might be at smaller scale (such as a coastal inlet) and coarser resolution in the open ocean. Such a model is said to have a "nested grid". More details on some typical grid configurations are given in Arakawa and Lamb (1977).
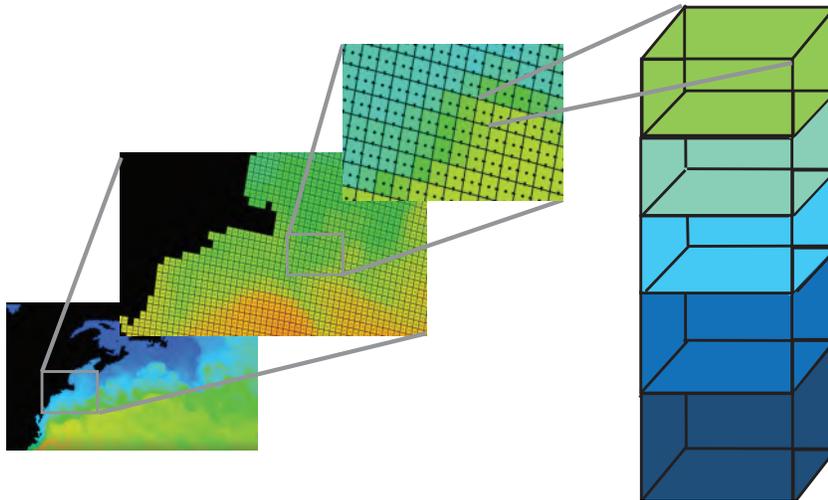
Figure 3.4 Schematic representation of model grid. Model domain is divided in many grid cells within which all model variables (e.g., temperature, Chl-a etc.) are assumed homogeneous. The grid cell is three dimensional. Grid cells are stacked on top of each other to resolve depth (right). Flow comes into and out of the box on each face (both horizontally and vertically).

3.2.3   Code and integration

Models take the equations as delineated in Section 3.2.1 and Box 3.1 and transform them into computer code. The computer code then steps the state variable (i.e., each pool of matter such as nutrients, phytoplankton concentration, zooplankton concentration) through time to see how they change. The process by which they do this is:

state variable at time 2 = state variable at time 1 + rate of state variable change*timestep

where the rate of change of a state variable is provided by the sources and sinks represented in the equations in Box 3.1. Thus, though the equations in Box 3.1 are written as an instantaneous change, models approximate this through a series of finite steps of duration "timestep". The timestep chosen for a model will depend on the types of processes it captures, both the physical and biological. For many biogeochemical and ecosystem models the timesteps are minutes to hours. Smaller time steps are often required to maintain stability in the model with smaller grid cells (Courant et al. 1928), compounding the computational cost of higher resolution. A factor of 2 increase in horizontal resolution generally leads to approximately 8 times increase in computational cost — it leads to 4 times as many grid cells and requires a timestep half as long as the coarser resolution.

The models step the state variables at each grid cell from some initial condition through many timesteps. The number of timesteps, and hence the length of the simulation, will depend on the questions the model is addressing. For instance, a model considering climate change will integrate from pre-industrial conditions until the end of the 21st century and will integrate for several hundreds of (model) years. With a timestep of 1 hour, this can mean several million integrations steps. However, a model looking at a short term process might only integrate for months or a year. A spin-up is needed to bring the models physics, chemistry and ecosystem into equilibrium with one another. Models studying biogeochemistry, in particular carbon uptake, may need hundreds or even several thousands of (model) years of spinup such that the deep ocean biogeochemical drifts are minimal. Otherwise the model air-sea flux results will include significant drifts. Note here that when we talk of simulations of many years, we refer to simulated "model years", not actual time the computers take to do the simulations. This latter is known in model jargon as "wall times".

3.2.4   Model output

Output from the models include the variables that the equations capture, such as nutrient concentration and biomass of phytoplankton and zooplankton (see Section 3.2.1 and Box 3.1). Usually models will also output rates such as primary production and nitrogen fixation. All of these output are provided as numbers either for an instant in time, but more often as averages over a time period (e.g., daily, monthly). Except for 0 and possible 1-D models, the amount of output is too large for something like an Excel spreadsheet. Thus output is usually in a format such as NetCDF or binary that can handle large amounts of data. Maps, transect, and timeseries (e.g., Figure 3.3) can then be produced from these numbers. Outputs are often referred to as "model diagnostics", since they allow the user to diagnose the dynamics underlying simulated

patterns. Storing model diagnostics for large models requires substantial computer disk space and can often pose as large a challenge as running the model itself.

3.3   Treatment of Light

Through photosynthesis, light has a critical influence on ocean biology. When light hits the surface of the ocean, it is absorbed and scattered by various components in the water. The amount of light that enters the ocean itself depends on the season, time of the day, solar zenith angle, presence of clouds, aerosols, etc. All of these conditions, as well as the type of components in the water, influence the amount of light available for photosynthesis. Despite these complex processes and the complex nature of light and its path in the oceans, models that include a high level of complexity in the transfer of light in the oceans are rare.

Typically, light at the surface of the ocean is prescribed (in model jargon, a "forcing field"). The most basic representation is to use a surface bulk irradiance (i.e., shortwave radiation e.g., Maier-Reimer et al. 2005; Doney et al. 2006) or a photosynthetically active radiation (PAR) at the surface (e.g., Palmer and Totterdell 2001; Zielinski et al. 2002). Most biogeochemical/ecosystem models include depth resolution (i.e., many grid cells in the vertical), and often to the bottom of the ocean, and thus it is important that a parameterization of light attenuation is included.

3.3.1   Typical treatment

The most basic models usually distinguish between the attenuation by water and total phytoplankton in the water column (e.g., Jiang et al. 2003; Maier-Reimer et al. 2005; Xiu and Chai 2014). In general, such models use an exponential decay of light, such that light at depth $I(z)$ is a function of the light at the surface I_o, attenuation by water molecules (coefficient c_w, units m^{-1}), and by phytoplankton Chl-a (attenuation coefficient, c_{chl}, units m^{-1} (mg Chl $m^{-3})^{-1}$).

$$E(z) = E_o \exp\Big(-(c_w + c_{chl} * Chl)\Big) \tag{3.5}$$

3.3.2   Including additional optically important constituents

There is growing recognition of the importance of other water constituents, such as coloured dissolved organic matter (CDOM) and non-algal particles (including non-living organic particulate detritus) for ocean biogeochemistry modeling (e.g., Dutkiewicz et al. 2015a; Kim et al. 2015). These components can be included by adding additional terms in the exponent of Equation 3.5

In a series of sensitivity experiments, Dutkiewicz et al. (2015a) showed that including other optically important constituents (such as CDOM and detrital matter) led to significant differences in the depth of the subsurface Chl-a maximum and the 1% light level (Figure 3.5b,c) relative to *in situ* observations (Figure 3.5a). Several models now include CDOM and detritus as important components affecting the light levels (e.g., Gregg and Casey 2007a; Xiu and Chai 2014; Kim et al. 2015; Baird et al. 2016). Additional constituents, such as minerals and

viruses are likely to also be important (Stramski et al. 2001), but have not yet (at least to our knowledge) been addressed by models.
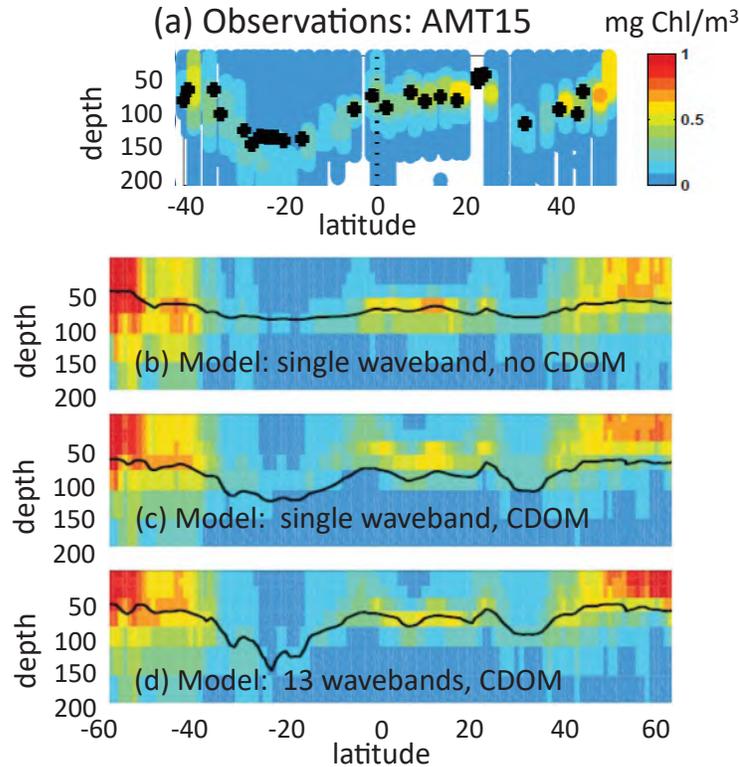
Figure 3.5 Sensitivity experiments showing the impact of resolving optical complexity. (a) Observations of Chl-a (mg m^{-3}) with depth, taken as part of the Atlantic Meridional Transect 15. (b)–(d) Model output of the same transect from three sensitivity experiments: (b) a single waveband of PAR, and attenuation of light only through water molecules and Chl-a; (c) single waveband of PAR, but now including optical impact of CDOM; (d) thirteen 25 nm wavebands of irradiance included, as well as optical impact of CDOM, and detrital material. Note that experiment in (b) includes only a simple attenuation code (such as Eq. 3.5) while (c) and (d) have a 3-stream radiative transfer code (see Section 3.3.3). Black line and symbols indicate the 1% irradiance depth. Figure adapted from Dutkiewicz et al. (2015a), Creative Commons Attribution 3.0 License (CC BY 3.0).

3.3.3   Including radiative transfer model

Light is not simply attenuated by water and other constituents as described in Equation 3.5. In fact, light is scattered and absorbed (Figure 3.6). Downward light (E_d) is absorbed (a) by dissolved substances and particles (e.g., phytoplankton, detritus, CDOM). The light is also scattered in all directions, including a forward component (b) and a back component (b_b). This latter scattering leads to upwelling light. For more details, see Chapter 2.1.2.

Capturing these complex processes requires a radiative transfer model (that is the equations/computer code that captures the radiative transfer interactions of light scattering and absorption through the water column). Recently a few numerical models have included
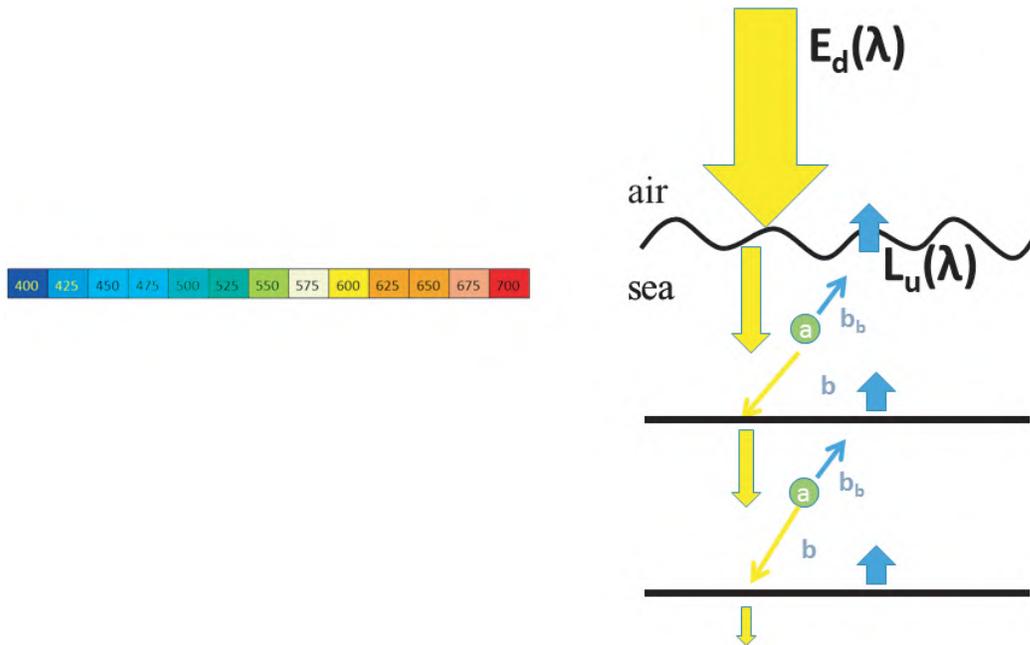
Figure 3.6 Schematic of light passage through the water column. Downward irradiance (E_d) is absorbed (a) by substances and particles (e.g., phytoplankton, detritus, CDOM). The light is also scattered in all directions, but includes a forward component (b) and a back component (b_b). This latter scattering leads to upwelling radiance, L_u (see Chapter 2.1).

such radiative transfer modules. The standard in radiative transfer models is Hydrolight (`www.hydrolight.info`; Mobley et al. 1993; Mobley 1994; Mobley and Sundman 2008a,b). However, Hydrolight is computationally expensive to use with biogeochemical models. The Ecosystem Light Subroutine (EcoLight-S) radiative transfer code (Mobley 2011) was developed to address the need for accurate but computationally fast irradiance calculations in any water body. Ecolight-S has similar computational algorithms to HydroLight, but EcoLight is a modular package designed to be implemented into any ecosystem models to improve the optical calculations. In Mobley et al. (2015), EcoLight-S was compared against a simple exponential light attenuation formula. EcoLight produced 57% more chlorophyll concentrations compared to the simpler formulation, suggesting that using a simpler model is not likely to capture light, and hence chlorophyll, correctly.

Another computationally less expensive option is the 3-stream model (Ocean Atmosphere Spectral Irradiance Model-OASIM, Gregg and Casey 2009) used in Gregg and Casey (2007a). Here, two downward streams of light (direct and diffuse) and one upwelling stream are captured. This type of model captures sufficient complexity to provide better representation of light fields and impacts on Chl-a (Figure 3.5, and see Dutkiewicz et al. 2015a; Gregg and Rousseaux 2017) but with less computational expense than a full radiative transfer model. This inclusion of complexity does improve the realism of the light field (e.g., 1% light level, Figure 3.5d) relative to observations (Figure 3.5a) with consequences for modelled phytoplankton

populations (Dutkiewicz et al. 2015a).

3.3.4 Including directional and spectral light

Most models have traditionally used a single waveband of photosynthetically available radiation (PAR, irradiances in the visible spectrum between 400 and 700 nm). But the optically important water constituents, such as CDOM as well as phytoplankton, absorb and scatter light differently at different wavelength (see Figure 2.4). For instance, phytoplankton absorb more at the blue end of the spectrum than in the green and yellow. Recognition of the importance of including spectral light in models dates back to the late 80's (Sathyendranath and Platt 1988) from *in situ* investigations of phytoplankton light absorption and estimates of primary production. More recently Fujii et al. (2007) and Kettle and Merchant (2008) further reinforced the importance of resolving spectral light fields in radiative transfer models, especially in ability to compare to more observations (e.g., optical). Some examples of studies that incorporate spectral irradiance include a 1-dimensional regional model of the Sargasso Sea (Bissett et al. 1999a), the West Florida Shelf (Bissett et al. 2004), and 3-dimensional models of the eastern U.S. continental shelf (Mobley et al. 2009), the North Pacific (Xiu and Chai 2014), the Great Barrier Reef (Baird et al. 2016) and globally (Dutkiewicz et al. 2015a; Gregg and Rousseaux 2017). The study of Dutkiewicz et al. (2015a) showed that including spectral light further changes the shape of the deep Chl-a maximum and 1% light level (Figure 3.5c) and found that spectral light was also important for the composition of phytoplankton communities.

Incorporating a spectrally-resolved radiative transfer module into physical and biological models can not only improve the visible light calculation (400–700 nm) but can also lead to better short-wave irradiance calculations (400–1000 nm) which are important for the upper ocean heat budget. Mobley et al. (2015) used EcoLight-S with the ROMS-CoSiNE to demonstrate that the changing biological conditions can also affect the total irradiance (400–1000 nm) calculation which impacts upper ocean thermal structure. For example, the surface chlorophyll values computed with EcoLight-S were 40% greater than the values computed with a simple light model, and modelled temperature was as much as 0.3°C warmer at the surface and as much as 0.1°C cooler at depth (Mobley et al. 2015). This thermal change shows the effect of including absorption by phytoplankton in near-surface waters, and consequent shading of deeper waters, which results in altering upper ocean thermal structure. The biological feedback to upper ocean heating can be calculated with a spectrally-resolved radiative transfer model coupled with physical-biological models.

Gregg and Rousseaux (2016) used the three-stream radiative transfer OASIM model (see Section 2.3.3) to assess the importance of including directional and spectral light in simulations of ocean radiative transfer. By sequentially removing directional (i.e., direct versus diffuse) and spectral irradiance they showed that including directional and spectral irradiance when simulating the ocean light field can be important for the biology, but that the magnitude varies with variables and regions of interest. For example, assuming that all irradiance was diffused had little effect on the primary production, but nitrate and chlorophyll concentrations declined by about 20% globally. The changes in nitrate and chlorophyll were much larger in the tropics

and sub-tropics. High latitudes are naturally dominated by diffused irradiance due to the presence of persistent clouds, so treating all the irradiance here as diffuse had less of an effect than in the tropics and sub-tropics, since there is already less direct irradiance to begin with. Disregarding spectral irradiance had effects that depended upon the choice of attenuation wavelength.

3.3.5 Including spectral inherent optical properties

Radiative transfer models can also include the optical properties of the various water constituents (see e.g., Section 3.3.2). The inherent optical properties (IOPs, see Chapter 2, Section 2.3.3), such as the absorption and scattering properties of the different biogeochemical standing stocks, can be included in models in terms of their spectral qualities. Fujii et al. (2007), for example, included the spectral effects of absorption and scattering of various constituents, and highlighted the fact that those optical properties play an important role in identifying and reducing the uncertainties in ecosystem models. CDOM in particular absorbed strongly in the blue, making it an important quantity to include spectrally. Dutkiewicz et al. (2015a) used a radiative transfer model with an explicit treatment of water molecules, detrital matter, CDOM, and several phytoplankton types to calculate surface upwelling irradiance and evaluate the resulting light fields with several data sets. That study found that including these constituents explicitly improved the distribution of light through the water column (Figure 3.5).

3.3.6 Impact of the sea floor

It is common for global biogeochemical/ecosystem models and the ocean colour products used to assess models, to consider only regions of the ocean where the fraction of light returning from the seabed is small. This criterion can result in ocean colour products from waters as deep as 25 m being excluded from consideration. In regional studies using biogeochemical/ecosystem models, such as in estuaries or coral reef systems, models often focus on shallow regions, including the inter-tidal zone. In these shallow regions, it is often the dynamics of bottom habitats such as seagrass meadows and coral reefs that are the motivation for the models (Baird et al. 2016). Ocean colour remote sensing in optically-shallow waters can be used to determine water optical properties, bathymetry and habitat distribution if the benthic reflectance is considered (Dekker et al. 2011; Garcia et al. 2018). Further, bottom reflectance can be used to improve the model simulation of remote-sensing reflectance, thereby allowing a more accurate model assessment against ocean colour (Baird et al. 2016).

Traditionally, shallow water systems have been studied using high resolution sensors with poor temporal coverage, such as Landsat. More recent satellites such as Sentinel-3 now provide good spatial and temporal coverage, opening up new opportunities for integrated ocean colour/biogeochemical model studies in shallow water systems. Thus it should be possible to use remotely-sensed estimates of seagrass biomass to the dynamics of seafloor biota in a biogeochemical/ecosystem model.

3.4   Different Models for Different Applications

Biogeochemical and ecosystem models are used for a variety of applications. As such they focus on different aspects of the system (Section 3.1). This report will largely consider models that focus more on the bottom of the foodweb, investigating phytoplankton/zooplankton ecology and ocean carbon and nutrient cycling. As described in Section 3.1, these models often truncate the foodweb (Figure 3.7a). Such models often parameterize the impact of higher trophic levels (e.g., a loss term that represents grazing by fish). Models might also focus on the upper water column, especially if they are designed for open ocean questions. Models focusing on coastal issues might include details of the benthic organisms and sedimentation (e.g., Baird et al. 2016; Butenschön et al. 2016). Spatial and temporal scales will depend on questions of interest: whether regional or global, days (e.g., for studying a spring bloom events), multi-annual (e.g., for interannual variability), to centennial (e.g., climate change), or millennial (e.g., for paleo-oceanography). Here we provide a non-exhaustive flavour of the different types of applications.
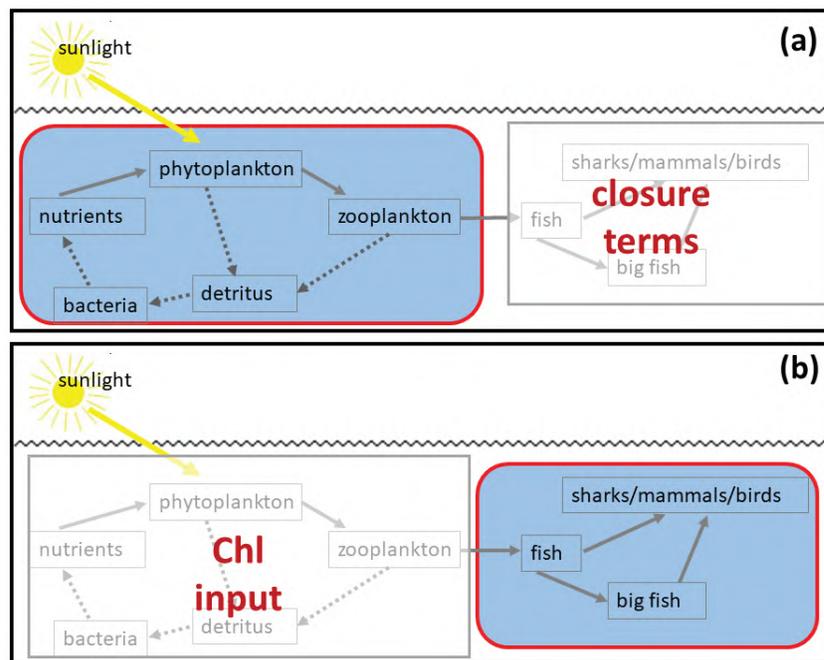
Figure 3.7   Schematic of different types of models. (a) Some models focus on the base of the foodweb, and higher trophic levels are heavily parameterized (in jargon often called "closure terms"). (b) Some models use Chl as an input and focus on the higher trophic levels (often a single fish species). In this report we will focus on the types of models represented in (a).

While the focus of this report is on biogeochemical and/or ecosystem models, it is worth noting that ocean colour products are also critical to other modelling activities. Another area of expanding usage is in species distribution models (SDMs) for marine species (schematically shown in Figure 3.7b). SDMs which include chlorophyll have been used for optimizing survey

efficiency (Zwolinski et al. 2011), helping fishers avoid high bycatch areas (Hazen et al. 2018), and examining risk of ship strikes to whales (Redfern et al. 2013). The use of ocean colour variables in SDMs is usually justified on the basis that it is a proxy for "foraging habitat". Models validated against ocean colour for the current climate (Chapter 5) can be used for chlorophyll-influenced predictions of changing species distributions and fish catch (Cheung et al. 2010) when used in climate change studies. Other uses of chlorophyll in similar statistical modelling frameworks include applications to harmful algal blooms (e.g., Stumpf and Tomlinson 2005) and hypoxia (e.g., Rose et al. 2017). These models usually do not incorporate the specific processes, but rather statistically relate fish catch or other observations to concurrent environmental factors to understand the conditions under which a particular species is most often found.

The skillful simulation of ocean biogeochemistry in a model relies on transport and physical properties provided by an ocean general circulation component of the model. Models of ocean circulation and biogeochemistry can either be run sequentially (off-line mode) or together (on-line mode). In the former, ocean velocities are computed independently from biogeochemistry and used as time-averaged fields (e.g., weekly, monthly, yearly averages) to drive tracer transport. Coupled physical-biogeochemical ocean general circulation models can be run in an "ocean-only" mode (i.e., where the atmospheric forcing is prescribed and the ocean heat/carbon cycles do not feed back to the atmosphere), or as an interactive component of an Earth system model including ocean, atmosphere, land and cryosphere components (i.e., where heat and carbon feedbacks between components are explicitly included).

3.4.1 Hindcast modelling

A hindcast simulation tests whether a model can explain or reproduce historical events. Ocean hindcasts typically use atmospheric forcing (winds, heat fluxes, precipitation) from historical reconstructions of atmospheric conditions derived by combining observations and models to reproduce oceanographic conditions. These products are referred to as atmospheric re-analyses, and there are many options, for example, MERRA (see Gelaro et al. 2017), NCEP (see Kalnay et al. 1996) and ERA-Interim (see Dee et al. 2011). The atmospheric forcing typically covers several decades, based on the observations available. The ocean modelling community uses hindcasts (which are also sometimes referred to as retrospective ocean simulations) to intercompare and assess the quality of ocean simulations as part of the Ocean Model Intercomparison Project (OMIP, Griffies et al. 2016; Orr et al. 2017). Similar intercomparison efforts have been extended to global biogeochemical and ecosystem simulations (Orr et al. 2017, MAREMIP, `https://pft.arc.hokudai.ac.jp/maremip/index.shtml`) and hindcasts are a broadly used configuration for regional applications.

Hindcast models provide a tool to explore the mechanisms controlling seasonal and interannual variability (e.g., Christian et al. 2001; Schlitzer 2004; Pastor et al. 2013). For instance, the hindcast model of Rousseaux and Gregg (2012) investigated the effect of ENSO on phytoplankton community composition in the Equatorial Pacific, finding that the decline in total chlorophyll observed during El Niño events using ocean colour data coincided with a

shift in phytoplankton community composition. Concentration of diatoms decreased during an El Niño event while cyanobacteria concentration increased (Figure 3.8).
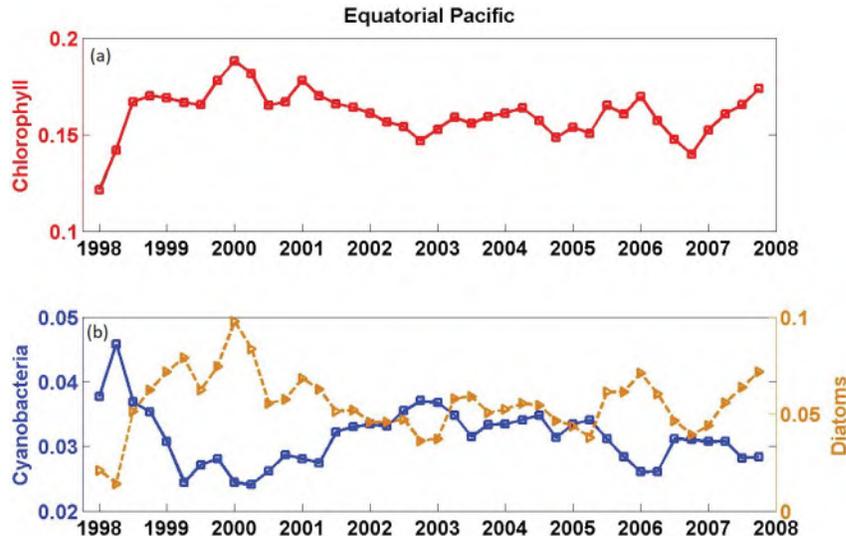
Figure 3.8 Hindcast Model. Temporal variation of Chl and phytoplankton concentration in the Equatorial Pacific. Seasonal average (Jan/Feb/March, detrended, seasonal climatology removed and average added) of (a) chlorophyll-a (mg chl-a m^{-3}) and (b) abundance of modelled cyanobacteria and diatoms (mg chl-a m^{-3}). Figure redrawn from Rousseaux and Gregg (2012).

3.4.2 Climate and Earth system models

Coupled global climate models include representations of the energy and of chemical species exchange across the three Earth system reservoirs — atmosphere, continental surfaces and ocean. The "coupling" in this context refers to the integration of atmosphere, ocean, and land models. These models produce their own climate in response to imposed radiative forcing at the top of the atmosphere, volcanic eruptions, green houses gases (e.g., CO_2, N_2O, CH_4) and other climate active species of anthropogenic origin (e.g., anthropogenic aerosols). The purpose of these projections is to simulate the evolution of Earth's climate over multi-decadal to centennial time-scales. As such, while global climate models simulate the correct frequency of modes of natural variability (e.g., ENSO), climate change projections are not designed to capture the historically observed timing of natural fluctuations. These models are sometimes referred to as Earth system models (ESMs). From a climate impacts' perspective, including a biogeochemical/ecosystem model enables projections of a range of potential ocean ecosystem stressors (Bopp et al. 2013), including changes in ocean colour (Dutkiewicz et al. 2019). For a review of construction characteristics of projection models see Stock et al. (2011) and Heavens et al. (2013).

Output from climate and Earth system projections from around the globe are collected as part of the Coupled Model Intercomparison Project (CMIP). Contributions include projections

under alternative greenhouse gas projection pathways. These outputs provide an essential contribution to the "Assessment Reports" prepared under the auspices of the "Intergovernmental Panel of Climate Change" (IPCC). Such models can suggest how biogeochemically important components (e.g., Chl) may change over the 21st Century (Figure 3.9).
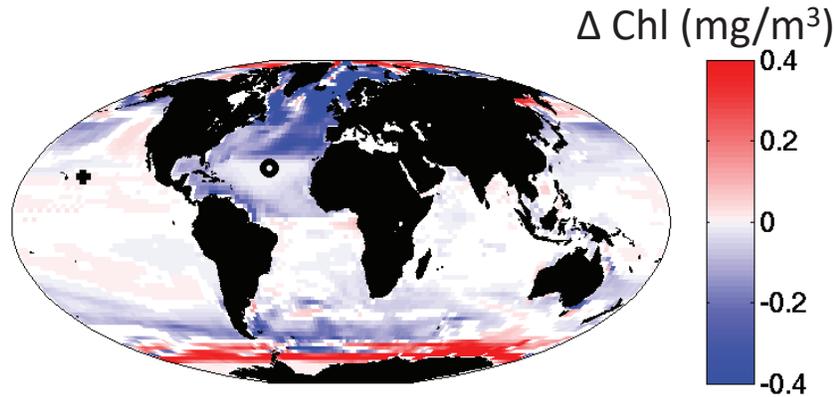
Figure 3.9 Change in Chl-a (mg Chl m^{-3}) between present (1998–2015) and the end of the century (2085–2100). Greenhouse emissions are not curtailed. Results are for a "business as usual" scenario, where blue indicates a decrease in Chl-a by the end of the century. Only areas with a statistically significant differences between the two periods ($p < 0.05$) are shown. Figure redrawn from Dutkiewicz et al. (2019), Creative Commons CC BY license.

Ocean colour products are used by the climate model community for evaluating models within the 20 year record for which there have been such products (Séférian et al. 2014). More recent applications include decadal predictions of marine productivity (Séférian et al. 2014) and the identification of emerging constraints on projected primary production (Kwiatkowski et al. 2017). For more information on the use of such models for seasonal to multi-annual predictions, and their use for marine resource applications, see Tommasi et al. (2017b).

3.4.3 Regional modelling

Regional models simulate limited portions of the global ocean, from the spatial scales of large ocean basins (e.g., the North East Atlantic, Ciavatta et al. 2016, Figure 3.10) down to coastal estuaries (e.g., Chesapeake Bay, Da et al. 2018). They are employed to investigate confined ecosystem dynamics (e.g., shelf-sea carbon fluxes, Wakelin et al. 2012; Mannino et al. 2016, and regional phytoplankton communities, Ciavatta et al. 2019) and often focus on local environmental issues that are relevant to marine policy and management of coastal and shelf areas, such as eutrophication (Laurent et al. 2018), harmful algae blooms (Glibert et al. 2014), hypoxia (Rose et al. 2017), and regional response to a CO_2 enriched atmosphere (Artioli et al. 2014). Nonetheless, they can contribute to global studies, for example, to quantify the contribution of shelf-seas to the global biogeochemical cycles (Fennel et al. 2006) or fisheries (Merino et al. 2012).

The objectives of regional models usually require high spatial resolution (from a few meters to ~10 km). Reducing the domain of regional models, however, makes such spatial
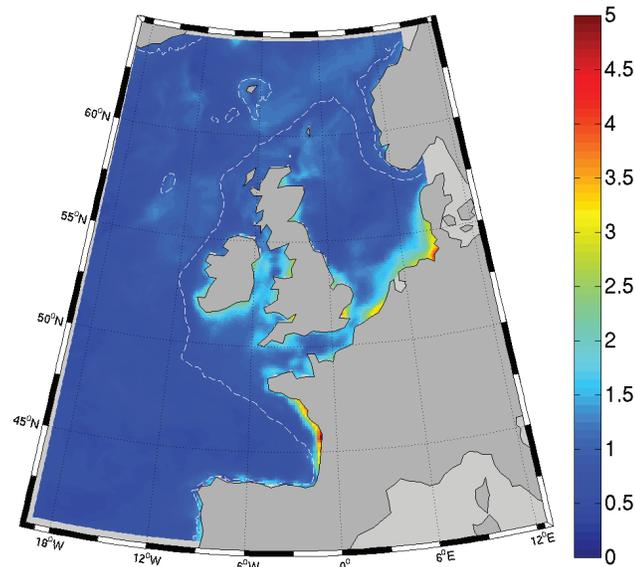
Figure 3.10 Regional Model. Surface chlorophyll concentration (mg m^{-3}) from the simulation of a regional model of the North East Atlantic: average value in the years 1998–2009. The white dashed line represents the shelf border, i.e., the 200 m isobath. Figure adapted from Ciavatta et al. (2016), Creative Commons Attribution 4.0 International (CC BY 4.0).

resolution computationally affordable. Global-scale models are still restricted to resolutions of roughly 1/12° to 1°. A challenging issue of regional models (compared to the global ones) is the need for defining values of the physical and biogeochemical variables at the open boundary of the spatial domain, i.e., defining the open boundary conditions. These conditions are often approximated by downscaling larger-domain models, or using climatological data from global reanalysis or datasets, or a combination of these approaches.

Regional models of shallow coastal basins should take into account energy and matter fluxes at the land and sediment interfaces, and often require the explicit representation of bottom friction and sediment resuspension, realistic forcing data or models for riverine inputs of fresh waters and nutrients, and full coupling of pelagic and benthic ecosystem dynamics (e.g., Butenschön et al. 2016).

Regional models can make use of ocean colour data for calibration, validation and data assimilation in the same way as their global counterparts. However, a distinguishing challenge is that they are often set up for shelf and coastal regions encompassing optically-complex, Case 2 waters. There, land inputs of coloured dissolved organic matter, sediment interferences and resuspension of particles can imply particularly high errors in ocean colour chlorophyll (IOCCG 2014). A careful account of the uncertainty of global chlorophyll products (e.g., per-pixel errors of the ESA-CCI product, Ciavatta et al. 2016), the application of regional chlorophyll products (Baird et al. 2016) or the use of alternative products (e.g., remote sensing reflectance, Jones et al. 2016), is therefore recommended when using ocean colour in combination with regional models of shelf-sea and coastal ecosystems.

3.4.4 Data assimilation

Data assimilation includes a number of formal methods for bringing models into consistency with observations, thus embracing the best of both model and observations. Assimilation of observations in order to improve biogeochemical and ecosystem models has a long history, beginning with parameter optimisation studies (Matear 1995) and extending through to modern day operational systems (Fennel et al. 2019). Ocean colour is often the first choice of data sets for assimilation into biogeochemical models, due wide availability and good spatial and temporal coverage. However, many challenges remain in using ocean colour in data assimilation, which are considered in depth in Chapter 6.

The products of data assimilative ocean-ecosystem runs are often referred to as ocean-ecosystem state estimates, ocean reanalyses, or, debatably, data-assimilative hindcasts. The integration of observed properties generally improves the fit to ocean and biogeochemical properties, though this also results in some uncertainty regarding the model deficiencies and processes underlying the corrections arising from data assimilation.

3.4.5 Operational models

Operational ocean models provide a nowcast and forecast of ocean conditions, such as water levels, temperature, salinity, currents, and biogeochemical properties, for global, regional or coastal areas. Very similar to weather forecasting systems, the ocean weather forecasts are derived from a combination of a suite of models and real-time observations. The operational ocean models include a series of ocean circulation, wave and ice, and biogeochemical models. These integrated models are forced by the outputs from weather forecast systems. In addition, operational ocean forecasting systems also include real-time ocean observations from both *in situ* and remote-sensing products for data assimilation by the operational ocean models. The real-time observations are also used to assess model skill and are incorporated into the long-term database to improve model hindcast results.

Nowcasts and forecasts are simulations of the present and future states of ocean conditions. A nowcast typically covers the period of time from the past few days to the present, and provides model outputs on locations where observational data are not available. A forecast provides information from the present time up to a few days as a short-term prediction, and/or up to a few months depending on the application needs (e.g., El Niño forecast). A forecast is usually initiated by the results of a nowcast.

There are several operational ocean forecasting systems producing physical variable short-term forecasts on global and regional levels. The European Marine Environment Monitoring Service (`marine.copernicus.eu`) has several versions of global and regional models, with higher spatial resolution for regional forecasting systems, combined with real-time data assimilation to produce daily to weekly forecasts of water levels, temperature and salinity, and ocean currents (Le Traon et al. 2017). The HYCOM consortium with multiple institutional partners (`hycom.org`) uses a global high resolution and adaptive vertical coordinate system to produce nowcasts and forecasts of ocean conditions, which include three-dimensional ocean temperature, salinity, and current structure; surface mixed layer depth; and the location of mesoscale features such

as eddies, meandering currents, and fronts. At regional and local levels, NOAA Operational Forecast System (`tidesandcurrents.noaa.gov`) has developed and maintains a network of nowcast and forecast hydrodynamic model systems driven by real-time data and meteorological, oceanographic, and/or river flow rate forecasts to issue short-term predictions of physical parameters (e.g., water levels, currents, salinity, temperature, waves) and disseminates them to users.

Compared to physical forecasting systems, developing operational models for predicting biogeochemical and ecosystem variables has been challenging. There are several operational centers which have started to take up this challenge by integration of biogeochemical models into operational systems (Gehlen et al. 2015). Some demonstrations at global and regional levels have shown that the concepts of biogeochemical data assimilation using advanced biogeochemical models are mature enough to be transitioned to operational centers (Ford et al. 2012; She et al. 2016; Skákala et al. 2018; Teruzzi et al. 2018). The capacity for producing biogeochemical forecasts along with reanalyses is a key requirement for the development of biogeochemical/ecological applications and services (Gehlen et al. 2015; Fennel et al. 2019).

3.5   Model and Model Output Selection

How does a user choose the best model and model output for an application? The choice should be based on the type of question to be addressed, and the spatial and temporal scale required. It is a good idea to contact the model developers to discuss the use of model output and the appropriateness to the question of interest.

If the question addresses a small region of the ocean (i.e., the North Sea), and if there is an appropriate regional model (see Section 3.4.3), then this is likely the best choice. Issues to keep in mind in this case are the appropriateness of the boundary conditions specific for the question being asked. For instance, if interested in how upwelling nutrients impacts coastal Chl-a, then the model needs to adequately capture this input through the open boundary. If there is no regional model, then a larger scale or global model may be helpful, but the non-resolved coastal physics must be kept in mind. If the question is global in nature, then a global model is required. One can also go one step further and use several global models (see e.g., Henson et al. 2010; see Chapter 8, Case Study 8.3.1) so as to cover some of the uncertainties in each model.

The length of the model simulation should also be considered. If the question involves understanding how interannual variability affects ocean colour, then using a model specifically designed for a hindcast (i.e., forced by the appropriate winds/heat for years, see Section 3.4.1) should be considered. Careful evaluation of the model is necessary to ensure that it captures the appropriate processes and at the required temporal scales. If the question is more immediate (e.g., the potential for a harmful algal bloom in the next week), an operational model (see Section 3.4.3) with a now- and forecast would be more appropriate. If the question is on climate change, then an Earth system model (see Section 3.4.2) would be more appropriate, but here it must be remembered that Earth system models have their own modes of variability and

may not capture, for instance, El Niño in the same years as the real ocean.

The level of complexity of the ecosystem model should also be considered when making a selection. For questions concerning changes in phytoplankton communities, one will require a model that captures at least some level of complexity of functional types or size classes. However, a more complex ecosystem model might not be as important to explore biogeochemical questions such as change in carbon cycling over the 21st Century. If the question is more optical in nature, then there are a few models that have incorporated various elements of the optics and radiative transfer that might be more appropriate (see Section 3.3).

Model output and how to compare to ocean colour (or other) products also requires careful consideration. For instance, model Chl-a may not always be exactly comparable to satellite derived Chl-a. This difference could be important when considering issues such as phenology (see e.g., Dutkiewicz et al. 2018). Once again contacting the model developers to discuss the applicability of the specific model to the specific problem to be addressed, would be very beneficial. See Chapter 4 for more discussion on the discrepancies between model output and ocean colour products, and Chapter 5 for model skill assessment. A relative knowledge of the model to be used and the meaning of the output is also recommended.

The (Mis)match between Biogeochemical/Ecosystem Model Variables and Ocean Colour Products

Stephanie Dutkiewicz, Anna Hickman, Colleen Mouw, Cecile Rousseaux, Stefano Ciavatta, Mark Baird, Charles Stock and Fei Chai

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

Satellite products, with their temporal and spatial coverage, are essential in model validation (see Chapter 5) and data-assimilation (see Chapter 6). However, it is important that modellers use these products in an appropriate manner. The discrepancies between what a model outputs and what the ocean colour product produces, as well as biases and missing data must be incorporated into comparisons between the two. This chapter is designed to elucidate these issues to modellers, and to provide the ocean colour community a context for understanding the modeller's dilemmas when using ocean colour products.

The link between ocean colour products and the variables that are used in biogeochemical models are not as clear or easy as might be hoped. Models have developed from interests such as carbon cycling and ecology, so have been structured to follow pathways of matter and energy, and capture basic groups of marine organisms. Though there has been a considerable development of radiative transfer modeling (e.g., Hydrolight, Mobley et al. 1993; Mobley 1994; Mobley and Sundman 2008a,b), radiative transfer and optics have not, in general been integrated in biogeochemical/ecosystem models (though see Section 3.3 below and Section 4.1.1 in Chapter 4). And since biogeochemical/ecosystem models are not (usually) developed for optical interests, the link to optical and ocean colour products are not clear. Moreover, the terminology between modelling and ocean colour communities sometimes varies. For instance, in the modelling communities "POC" often specifically refers to non-living particulate organic carbon. However, many other communities, who measure "POC" *in situ* or estimate it from satellite, include all living and dead organic carbon — the precise definition often depends on measurement technique. Mis-communications such as these can lead to confusion.

Modellers and other non-ocean colour experts often equate satellite-derived products with their "*in situ*" counterparts (e.g., satellite derived Chl-a and *in situ* Chl-a) without fully appreciating that there are distinct differences. The uncertainties are compounded going from first order measurements from satellites (such as water leaving radiances) to derived products such as Chl-a, and even more so for primary production. This degree of "derivedness" is encapsulated in Figure 4.1 (see also Figure 2.8). Deriving information from remote sensing reflectance is done empirically (i.e., using empirical relationships between remote sensing reflectance and the property of interest) or semi-analytically (i.e., using the radiative transfer equation) or a combination of both. Thus, most ocean colour products are themselves a type of model (in this report we refer to these as "algorithms"). The different information used, including the observed remote sensing reflectance, theoretical and empirical relationships, and the necessary assumptions and averaging, can each contribute to mismatches in magnitude and/or

variability of properties when compared to output from numerical models.
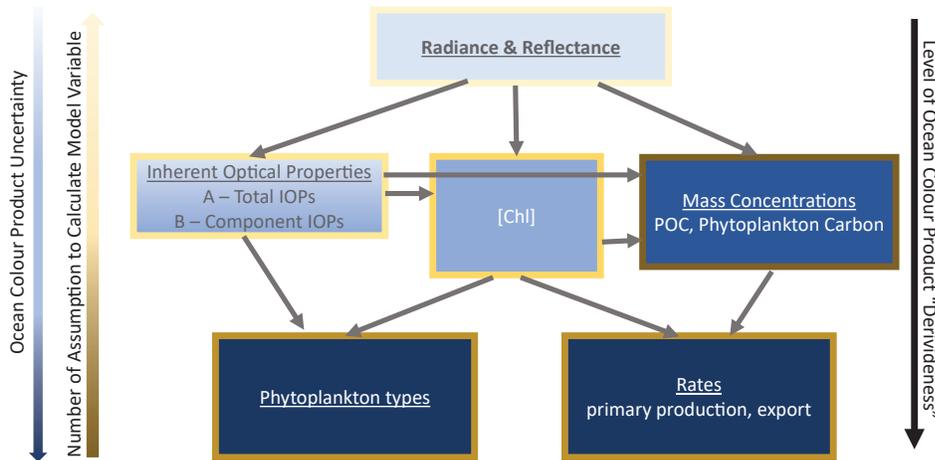
Figure 4.1 Schematic depicting uncertainties and level of "derivedness". For ocean colour products, uncertainty is coloured in blues, with darker colours suggesting more uncertainty. Assuming radiance and reflectance are the least derived ocean colour, the vertical position (tier) of each box depicts how many levels of products are used to derive it. The gray arrows indicate which products are used to derive which other products. For instance, Chl-a can be derived either from radiances and reflectance or from IOPS. In general, the more derived the product is, the more uncertain it is. The brown colours surrounding the boxes indicate the level of assumptions needed to calculate the variable in the model. Most models follow carbon, so phytoplankton carbon is one of the most basic variables. IOPs and reflectance require the most assumptions in the model to calculate.

The central currency of many biogeochemical/ecosystem models is an element (e.g., carbon, nitrogen). Nitrogen is often used as it is the proximal limiting nutrient in many ocean regions, and carbon and nitrogen are linked in a relatively constant proportion compared to more dynamic fluctuating C:P or C:Fe (Redfield 1934; Redfield et al. 1963; Martiny et al. 2013). Quantities for comparison with ocean colour must be derived from these central currencies and additional models of their relationship with chlorophyll or other optical properties (Geider et al. 1998). Thus there is an opposite direction between models and ocean colour products (see coloured arrow in Figure 4.1), with model (usually) going from a base unit of carbon, and ocean colour products inverting carbon from the optics and algorithms.

Models have regular grids and there is no "missing data", though comparisons can be limited by the timescale over which the output is averaged (days/months/years, see Section 3.2.4). Satellite measurements have missing data due to orbital periods, clouds, aerosols etc., so that satellite products are not true means. The ocean colour community call the resulting data "climatologies", whereas the word "climatology" to a modeller usually refers to a long term mean (e.g., over several years), with continuous output (i.e., a true mean). Biogeochemical/ecosystem model grids are also three dimensional, with depth-averaged data within each grid point. Ocean colour products are provided as two-dimensional fields, usually "seeing" one optical depth (1/attenuation coefficient, see Appendix 3). There can thus be confusion between the two communities about the different resolution and averaging of data.

In this chapter we highlight how comparing ocean colour products, *in situ* observations and biogeochemical model output is rarely like-for-like (see for instance Figure 4.24). Even though satellite products and model output can both be compared directly to *in situ* data they will not necessarily compare to each other because they are built on different assumptions/have different biases. In general,

in situ measurements have less uncertainty (i.e., they are closer to the truth) than satellite-derived products. However, different *in situ* measurements have different levels of uncertainty (e.g., HPLC vs. fluorescence measurements of Chl-a). Similarly, different satellite products of the same quantity have different levels of uncertainty (see Chapter 2). What should models compare too? How should the relative uncertainties be weighed against each other? We only begin to address these questions in this report (though see Chapter 5), but pose these as important issues to be considered when considering model variables against both ocean colour products and *in situ* measurements.
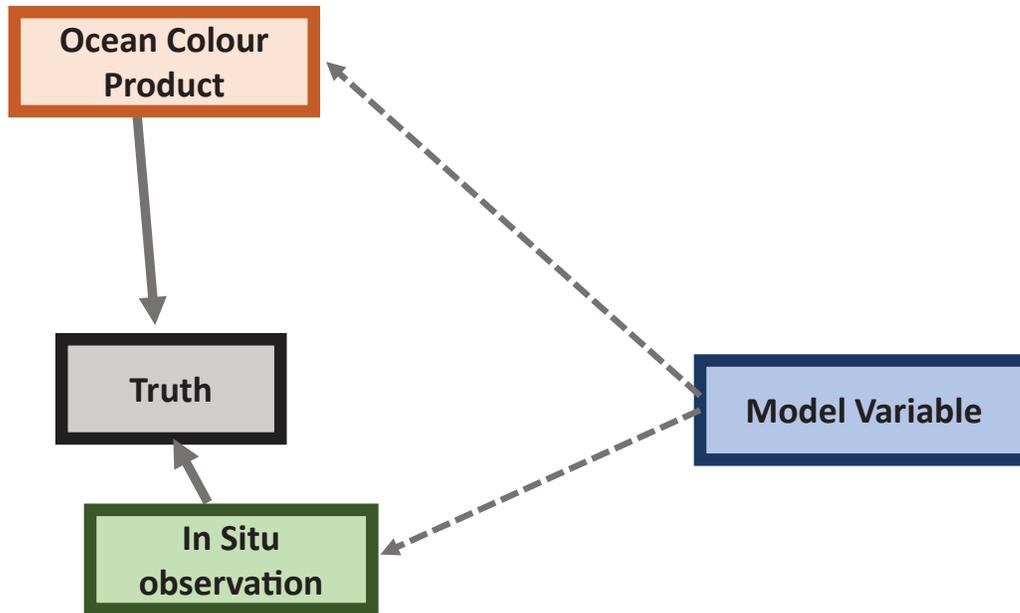
Figure 4.2 Schematic of uncertainties in *in situ* and ocean colour products. Truth (e.g., the actual Chl-a concentration at a location) can be measured in the field by several techniques (e.g., HPLC, fluorometry, see Section 4.1.3) and there are several ocean colour products of Chl-a (e.g., derived from reflectance ratios or semi-empirical methods, see Section 2.3.2). The length of solid arrows suggests the level of uncertainty of the measurements. Generally, *in situ* measurements are less uncertain than the ocean colour products. Model variables can be compared to either *in situ* or ocean colour products (usually both). It is important to remember that the two types of measurements are different, with different uncertainties and biases. Even though the quantities can be given the same "name", they often have considerable mismatches in what they actually capture.

Section 4.1 below discusses the mismatch between model and satellite product for several different products (and where appropriate, their *in situ* counterpart), focusing on the "mismatches in kind". Next, we discuss issues arising from the difference in the spatial and temporal resolution (Section 4.2).

4.1 Same Name, Different "Measurement"

In this section, we consider the differences between ocean colour products and model variables that share the same name. For instance, ocean colour produces an estimate of Chl-a from remotely sensed reflectance, but this is not estimated in the same way as *in situ* Chl-a (which, in turn, can be measured in different ways). The model variable "Chl-a" is defined in a way more similar to *in situ* Chl-a: it is the actual Chl-a content of the model phytoplankton, as opposed to the satellite product which is derived from satellite optical measurements. We will discuss such discrepancies and possible implications

below. We focus on links between ocean colour products that are freely available, usually globally, and not on regional, project-specific products. As such, this section does not provide a good overview of coastal products. We refer the reader to Mouw et al. (2015) and Zheng and DiGiacomo (2017) for a comprehensive review of coastal remote sensing.

There are distinct advantages to models producing output closer to what satellites measure (e.g., reflectance) and thus closer to the satellite derived products (e.g., Chl-a, albeit with care taken to consider discrepancies between the measurements and the model output). However, there is computational costs and additional complexities for models to provide more comparable output, especially reflectance and optical properties. Additionally, end users of models (e.g., ecosystem managers) are often more familiar with quantities such as Chl-a rather than reflectance. We anticipate that different models and different scientific motivations will dictate how closely models will link to ocean colour products in the future.

While this chapter deals with fields that directly relate to model state variables (i.e., Chl-a and phytoplankton carbon), we also consider other direct links between model output and ocean products such as reflectance and inherent optical properties (see Section 2.3.4 in Chapter 2). Fujii et al. (2007) suggested that including explicit optics in an ecosystem model allowed a more accurate subsurface light field as well as additional constraints on model parameters. Several additional studies have demonstrated the value of adding optics to biogeochemical models (e.g., Babin et al. 1993; Sathyendranath and Platt 2007; Kettle and Merchant 2008; Dutkiewicz et al. 2015a). Thus, we also discuss here the links between models that include optical properties and their relevant ocean colour products. We follow from least to highest level "derivedness" and uncertainty from the satellite products' perspective (Figure 4.1).

4.1.1   Water leaving radiance and reflectance

Water leaving radiance and reflectance (see Section 2.2.2 in Chapter 2) are the primary variables observed by ocean colour radiometers, and can be thought of as the primary ocean colour products. They have the least associated uncertainty of any satellite ocean colour product (Figures 4.1 and 4.2). However, very few numerical models simulate these quantities explicitly (see discussion in Section 3.3, Chapter 3). Here we highlight the few models that do capture these quantities, and show that even then, there is a mismatch between the ocean colour and model quantities.

Satellite sensors usually take measurements close to noon, but the solar angle still varies due to latitude. Water leaving radiances provided by space agencies have been converted to approximately zero zenith angle (i.e., as if the sun is directly overhead) by dividing by the cosine of the solar angle, as well as including corrections for other atmospheric, air-sea interface and orbital cycle effects (see Chapter 2.1). While these assumptions correct for the effect of changing solar angle on the downwelling radiance (and hence water leaving radiance), they cannot correct for the ocean colour products seeing deeper in the water column at zero zenith angle. A simple example of this is demonstrated by the approximately 20% decrease in Secchi depth at a sun zenith angle of 50° (Lee et al. 2015b). However, products that use more than one band implicitly assume that the ocean is mixed such that the different penetration depth makes little difference to the output. Satellite downward irradiance is also a product, as while these estimates are based on measurements, there are conversions used resulting in the final estimate. Thus, provided a biogeochemical/ecosystem model uses the same downward irradiance as the satellite product, water-leaving radiance is likely the closest comparison between model and satellite products.

While radiance measurements made at sea do not have atmospheric correction issues and are measured from directly overhead, they may not be observed at noon, and have challenges associated with sensor and water shadowing. As such, there is also a mismatch between *in situ* radiometric measurement and even the most basic of satellite ocean colour measurements.

4.1.1.1   EcoLight-S

EcoLight-S has similar computational algorithms to HydroLight (Mobley 1994), but EcoLight is a modular package designed to be embedded into ecosystem models to improve the optical calculations (Mobley 2011). EcoLight-S solves the radiative transfer equations for radiance. It automatically accounts for transmission of sky radiance (hence of irradiance) through the sea surface as a function of wind speed, solar zenith angle, and atmospheric conditions. The spectral absorption and scattering properties of water are used as inputs to EcoLight-S, which can describe any water body from the clearest Case 1 water to the most turbid Case-2 water. EcoLight-S is therefore valid for use in any water conditions, and there is no need to predetermine a water type for light and heating calculations. EcoLight-S also provides outputs of the remote-sensing reflectance, in-water spectral irradiances and upwelling radiance, which can be used to validate ecosystem predictions via remotely sensed or in-water optical measurements such as BGC-Argo (Fujii et al. 2007; Mobley et al. 2015).

4.1.1.2   Three-stream model

The models of Gregg and Casey (2007a) and Dutkiewicz et al. (2015a) both include a three stream-radiative transfer code (Gregg and Carder 1990). The three stream (downward direct, E_{dd}, downward diffuse, E_{ds} and upwelling, E_u) model follows Aas (1987), Ackleson et al. (1994), and Gregg (2002a). Downwelling (direct and indirect) irradiance just below the ocean surface are provided as forcing fields. The three streams are followed through the water column by the radiative transfer model. Converting these model outputs to remotely sensed reflectance requires some assumptions and simplifications. The upwelling irradiance, E_u, must be converted to radiance (L_u) using a distribution function (Q) which has values between 3 and 5 steradian (Morel et al. 2002) and depends on several variables, including inherent optical properties of the water, wavelength, and solar zenith angles (Morel et al. 2002; Voss et al. 2007). The subsurface values must also be adjusted to be above surface, to link to satellite-like measurements.

Gregg and Rousseaux (2017) convert to normalized upwelling spectral radiance (L_{wN}) using a formula that takes into account mean extraterrestrial irradiance just below the ocean surface, surface reflectance, and the index of refraction. Using 1-nm spectral resolution in the model simplifies the comparison with MODIS-Aqua L_{wN} by avoiding band mismatches. The model provided radiances that were within -0.074 mW cm^{-2} um^{-1} sr^{-1} of MODIS-Aqua radiances at 412, 443, 488, 531, 547, and 667 nm. The water leaving radiances from the model were significantly correlated to those from MODIS-Aqua (R = 0.71 p<0.05) and on average 10.4% lower than those from MODIS-Aqua with a mean semi-interquartile range of 0.08. These results suggested that this model has skill in simulating global water leaving radiances, and supports the use of these simulations for various applications including satellite mission design and analysis (e.g., PACE).

In Dutkiewicz et al. (2018, 2019) the below water reflectance diagnostic output was recalculated as if the sun was directly overhead (i.e., side-stepping the issues that the real satellite products have to correct). However, the same issues (as mentioned above) of converting irradiance to radiance, and from below to above water still needed to be parameterized. In these studies, model output was converted to above surface remotely sensed reflectance (R_{rs}) using the formula of Lee et al. (2002). The R_{rs} from the model was in 25 nm bands from 400 nm to 700 nm. These bands do not match those from satellite sensors, which measure over smaller bands and centered at different wavelengths (see Figure 2.7). Interpolation to satellite wavelength provide additional uncertainty when comparing directly to satellite products. However, there was good correspondence between the model remotely sensed reflectance and the satellite product.

4.1.1.3 Relationship to absorption and scattering

To avoid the significant computational cost of an explicit radiative transfer model, Baird et al. (2016) made a simpler calculation of remote-sensing reflectance using depth-resolved optical properties directly. First, the model calculates the depth-resolved absorption, total scattering and backscattering of different water constituents. From these, the depth-resolved ratio of backscattering to the sum of backscattering and absorption, $\mu = b_{btot}/(a_{tot} + b_{btot})$, is determined for each layer. The irradiance reflectance just below the surface, $R(\lambda, 0)$, is then calculated by weighting the influence of each layer using twice the fraction of light attenuated in the layer, and summing the weighted μ for the entire water column, and including the bottom reflectance, if light reaches the bottom. The remaining calculations converting irradiance reflectance to remote-sensing reflectance follow the procedure of Lee et al. (2002). The use of an optical depth-weighted μ gives an identical solution to a full radiative transfer model for waters in which the optical depth of water is shallower than the surface model layer thickness (for example with red light and/or turbid waters), but deviates when upwelling irradiance from lower layers becomes important.

Case Study: The Use of Reflectance in Models

Model water leaving irradiance and reflectance can be quantitatively compared at multiple wavelengths to assess the skill of the biogeochemical model. For qualitative but quite insightful comparisons in "colourful" water bodies, such as the shallow waters of the Great Barrier Reef, Baird et al. (2016) have shown the usefulness of comparing simulated true colour with observed true colour (Figure 4.3). Comparison of the two images immediately reveals some of the strengths and weakness of the biogeochemical model. The 1-km model sufficiently resolves submerged reefs so that individual features can be readily identified in both images, although on the shelf edge some reefs in the model are too deep to be seen from space. In both the observed and simulated true colour images, the macrotidal shallow coastal waters of Broad Sound (22°S) appear a similar hue of brown. This illustrates that the model both resuspends material in the correct locations, and that the resuspended material has mass-specific optical properties similar to the observed suspensions. But the waters off Fitzroy River are not sufficiently brown, possibly suggesting the suspended concentrations in the Fitzroy River flow are underestimated. Even in the relatively oligotrophic waters of the Great Barrier Reef, the patterns of algal blooms can be seen in both model and observations.

4.1.2 Optical properties

Uncertainty grows the further a satellite derived product gets from the light signal directly measured by a satellite radiometer (Figure 4.1). Inherent optical properties (IOPs) refer to water optical characteristics, including the absorption and scattering coefficients in the water column (see Section 2.3.4 in Chapter 2), and are the next tier in terms of level of derivation and uncertainty after R_{rs} (see Figure 4.1). As such, IOPs could be better comparison products than chlorophyll concentration or other products based on mass concentration that are further down the table of "derivedness". However, many models do not include these as output, and users of models might have less understanding of these quantities than for more biogeochemically relevant quantities such as Chl-a. Though for completeness, and since more models are likely to include some aspects of IOPs in the future, we include this section here.

IOPs provided from ocean colour satellite data are products derived from optical measurements. Both the measurements, and the algorithms used to derive the IOPS have uncertainties associated with them. The underlying optical relationships used in the IOP algorithms are dependent on empirically-derived constant spectral shape coefficients (see Chapter 2). As such, they do not capture the full range of temporal and spatial variability found in specific regions of the ocean, creating regionally varying uncertainty (see uncertainty maps of IOP products hosted by NASA: `https://oceancolor.gsfc.nasa.gov/`). Some new approaches (Werdell et al. 2018) are emerging that attempt to allow the coefficients to vary,
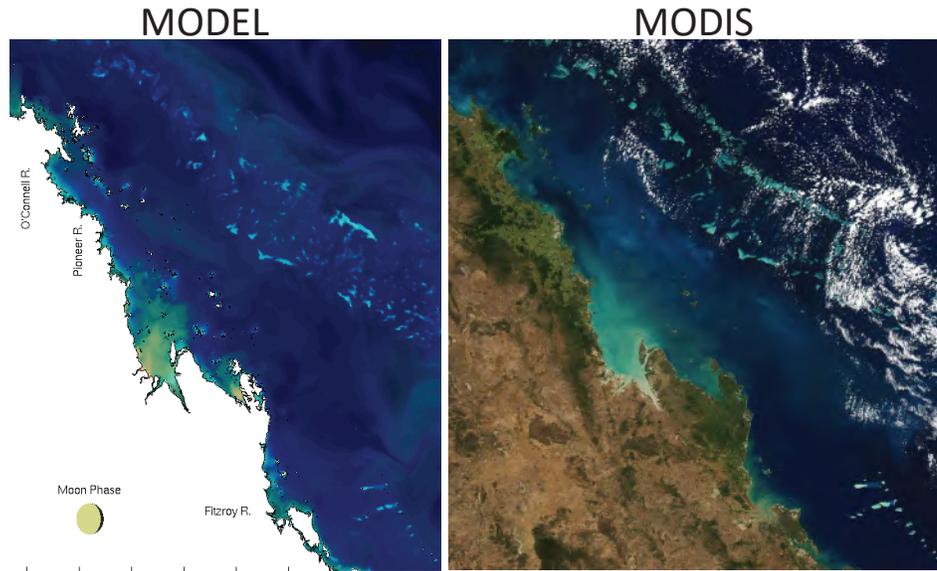
Figure 4.3 Comparison of simulated (left) and observed (right) true colour in the complex coastal waters of the Great Barrier Reef on 2 June 2018. Simulated true colour (left) obtained from simulated remote-sensing reflectance at 645, 555 and 470 nm of the ~1-km resolution eReefs simulation (www.eReefs.info, Baird et al. 2016). The ~250 m observed true colour (right) is from the MODIS Aqua sensor processed using the NASA Worldview application (worldview.earthdata.nasa.gov). Both images are generated using the NASA true-colour processing algorithm. Worldview uses MODIS Bands 1, 4, and 3, respectively corresponding to the red, green, and blue range of the light spectrum, assigned to the red, green, and blue channels of a digital image. The images include a brightening factor that depends on the graphics package rendering the image, and have been approximately matched.

reducing uncertainty.

Several studies have demonstrated the value of adding optics to biogeochemical models (e.g., Babin et al. 1993; Fujii et al. 2007; Sathyendranath and Platt 2007; Kettle and Merchant 2008) and newer models explicitly resolve IOPs (Fujii et al. 2007; Gregg and Casey 2007a; Xiu and Chai 2014; Dutkiewicz et al. 2015a; Baird et al. 2016). For instance, better optics provides a more realistic/unbiased light field for the ecosystem processes (e.g., phytoplankton growth). Moreover, including IOPs increases the amount and types of data that can be used to validate a model (see e.g., Fujii et al. 2007).

Total scattering, backscattering and absorption are a product of the different optically important constituents such as water molecules, phytoplankton, detrital matter, coloured dissolved organic matter (only important for absorption), viruses, minerals and salt. Models that include IOPs must make simplifications and rarely include all optically important components. In the models, each constituent must be related to the concentration of the corresponding model state variables (CDOM, detrital matter, and phytoplankton Chl-a or biomass) and to the absorption/scattering spectra assumed for each (see Figure 2.4).

Thus, while the satellite IOP products are high up in the table of "derivedness", the model IOPs are emergent properties of complex feedbacks involving many optical constituents, though some of which may be quite crudely represented. Here we consider absorption by CDOM (a_{CDOM}), as a case example of the difficulties between linking ocean colour satellite measurements, *in situ* measurements, and model output.

Case Study: Absorption by CDOM (Models, Ocean Colour and In Situ)

Coloured dissolved organic matter (CDOM), also known as gelbstoff or gilvin, is particularly important in absorbing in the blue portion of the spectra. However, CDOM is made up of many compounds and is not usually a state variable in models. Absorption by CDOM (a_{CDOM}) is typically measured as the difference in absorption of water samples before and after filtration through a 0.2 μm pore size filter. In contrast, most satellite products do not separate CDOM from absorption by detrital particles and the satellite-derived property is referred to as "CDM" (though see some GIOP products, Werdell et al. 2013). Though it is likely that detrital particles contribute <20% to CDM (Siegel et al. 2002), this is regionally variable. Thus, even *in situ* and satellite-derived products of a_{CDOM} are not directly comparable.

A few models have included a_{CDOM}, though at different levels of complexity. Fujii et al. (2007) included uniform a_{CDOM}, and in Gregg and Casey (2007b) a_{CDOM} was treated as a function of Chl-a. Dutkiewicz et al. (2015a) and Xiu and Chai (2014) included a CDOM-like tracer similar to Bissett et al. (1999a): it is assumed to have a source that is a fraction of DOM production, to have a long remineralization timescale, and to be bleached under high light conditions. a_{CDOM} is then related to the "CDOM" tracer. The absorption from CDOM calculated in the latter models are actually quite similar to those measured in the water column, and the comparison between model and these *in situ* observations is quite straightforward (see Figure 4.4e,f). However, comparing to satellite derived a_{CDOM} is far more complex, given that the ocean colour product is itself a "model" and the ocean product also includes detrital particles. Thus, in models that also capture absorption by non-living particles, the comparison should be between ocean colour a_{CDM} and model $a_{CDOM} + a_{DET}$.

There are a number of other considerations common to comparing model output to satellite-derived and *in situ* observations. Firstly, many numerical models do not resolve the terrigenous sources of a_{CDOM} that may have different turnover times and optical properties than a_{CDOM} produced in the ocean. Models also do not resolve the different turnover timescales of the different components of a_{CDOM}, including the potentially large pool of recalcitrant a_{CDOM} that would be captured by satellite (Hansell 2013).

Case Study: Vertical Attenuation (Models and Satellite)

The satellite product for vertical attenuation, $K_d(490)$ (m^{-1}), returns the diffuse attenuation coefficient for downwelling irradiance at 490 nm calculated using an empirical relationship derived from *in situ* measurements of $K_d(490)$ and blue-to-green band ratios of remote sensing reflectances (R_{rs}). Traditionally, biogeochemical models with simple optics have calculated a vertical attenuation rate based on the sum of vertically attenuating components (e.g., Fasham et al. 1990), accepting that, as recently demonstrated by Lee et al. (2018), attenuation is not simply the sum of individual constituents. More recently, sophisticated optical models calculate the absorption, scattering and backscattering of individual optically-significant components, and then calculate $K_d(490)$ depending on the light field, most critically on the solar zenith angle.

The vertical attenuation coefficient is fundamental to depth of solar heating and the vertical distribution of primary production, so strongly determines model behaviour. At the same time, the satellite products for $K_d(490)$ are more directly related to the phenomena they are quantifying than other IOP algorithms, such as chlorophyll concentration. Thus when Ciavatta et al. (2014) (see Chapter 6) used the mismatch between satellite-observed and simulated K_d (here for 443 nm, not 490 nm), they found better performance than assimilating satellite-observed chlorophyll concentration.

4.1.3   Chlorophyll-a

To highlight the points made by Figure 4.2, we first provide a brief description of uncertainties in *in situ* methods of measuring Chl-a. The most established approach for measuring Chl-a concentration in the field is by filtering water samples through GF/F filters and measuring the Chl-a fluorescence of
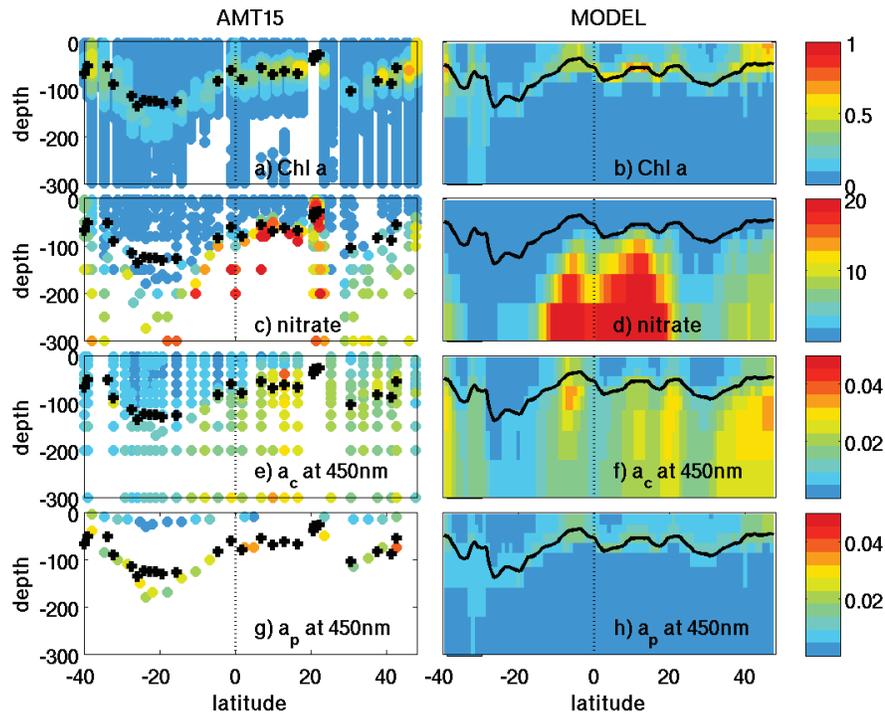
Figure 4.4 *In situ* observations and model comparisons along a transect through the Atlantic (Atlantic Merdional Transect 15). Observations on left are from (a) Chl-a; (c) nitrate; (e) absorption by CDOM, and (g) absorption by phytoplankton. The right column shows the model of Dutkiewicz et al. (2015a). (b) Chl-a with acclimation following Geider et al. (1998); (d) nitrate as a state variable; (f) absorption by CDOM, where CDOM is a state variable: an exponential function relates absorption to CDOM concentration at different wavelengths; (h) absorption by phytoplankton, where absorption is linked to each of the 9 modelled phytoplankton functional groups with a wavelength dependent spectral function determined from laboratory experiments. From Dutkiewicz et al. (2015a), Creative Commons Attribution 4.0 License.

the retained material resuspended in solvent (e.g., Welschmeyer 1994). However, Chl-a determined fluorometrically typically includes the effects of chlorophyll-b and phaeopigments in addition to Chl-a (Trees et al. 1985). An alternative approach using High Performance Liquid Chromatography (HPLC, e.g., Bidegare et al. 2002) better isolates and quantifies Chl-a, and is preferred for satellite validation (Mueller et al. 2003). HPLC measurements are, however, more time consuming and expensive so there is far more fluorometric data. Concentrations derived using the two methods are well correlated, with Chl-a typically 40% higher when measured fluorometrically compared to HPLC (Trees et al. 1985). A calibration is often used to scale fluorometric to equivalent HPLC values, but the relationship varies with cell physiology, and thus ocean conditions, so that cruise or region-specific calibrations are preferred (Mueller et al. 2003). Chl-a fluorescence can also be measured using various *in situ* sensors, including on profilers or autonomous vehicles. Again, the relationship between the measured fluorescence and Chl-a varies considerably, with non-photochemical quenching making the use of such data for remote sensing purposes particularly problematic (Roesler et al. 2017).

The most often-used algorithm for estimating Chl-a from ocean colour relies on the greater absorption in the blue range of the light spectrum than the green. The ratio of the amount of blue to green light leaving the ocean surface at any location therefore supplies information on the concentration of Chl-a. Using datasets of coincident radiometric observations and *in situ* Chl-a, a 4th order polynomial

can be constructed to estimate Chl-a from measured blue/green reflectance ratios (e.g., O'Reilly et al. 2000, see Section 2.3.1 in Chapter 2 for more in-depth discussion). This empirical algorithm is then used globally with satellite remote-sensing reflectance. The relationship is typically considered robust in open ocean conditions where the optical effects of phytoplankton co-vary with other optical constituents (including CDOM and detritus), so called Case-1 conditions (Smith and Baker 1978; Morel 1988; O'Reilly et al. 2000). Though even in these waters the mean error estimate is about 35% and rises even higher in coastal waters (Moore et al. 2009). It is likely that locally the errors are much larger, especially in areas of low Chl-a.

It is relatively well understood in the ocean colour scientific community that these ocean colour products have large errors and specific regional biases (Hu et al. 2000; Moore et al. 2009; Szeto et al. 2011; Blondeau-Patissier et al. 2014). However, there remain aspects of errors, biases and uncertainties that are poorly quantified, particularly in regions where there are little or no *in situ* data to compare to the satellite derived products. Further, many users of ocean colour products whose main expertise are in other arenas (e.g., numerical modellers) are less aware of these issues.

Most numerical models capture phytoplankton biomass in terms of carbon or nitrogen concentration (e.g., mmol C m^{-3}). Chl-a is rarely the main "currency" (though see for instance Gregg and Casey 2007b). Many models however diagnose a Chl:C ratio associated with the phytoplankton species (or types) captured in the model. Such Chl:C ratios are a function of the environment (nutrients, light, temperature). A frequently used formulation of the Chl:C ratio used in models is that of Geider et al. (1997, 1998), though potential improvements have been suggested (e.g., Baird et al. 2013; Jackson et al. 2017). The Chl-a content of the phytoplankton can either be a diagnostic value, or an explicit variable in the model that is followed over time with rates of synthesis and loss of chlorophyll determined via a parameterization such as Geider et al. (1998). It is advantageous to include Chl-a in a model as it improves the parameterization of photosynthetic rates and productivity. However, a secondary reason to include Chl-a in models is that this is the easiest and most ubiquitous variable with which to compare to observations. For instance, Chl-a is more often easier to measure in the real ocean than phytoplankton carbon, and model Chl-a fields are easy to compare to *in situ* measurements of Chl-a concentration (see Figure 4.4a,b). Moreover, the ocean colour Chl-a is better known than the ocean colour products of phytoplankton carbon, and likely have lower errors associated with them (Figures 4.1 and 2.8). Thus, global and seasonal evaluation of model Chl-a can be done with ocean colour Chl-a (see e.g., Chapter 5).

Ocean colour satellite-derived Chl-a is not only used as an evaluation product, but has also been used for data assimilation (e.g., Gregg 2008; Ciavatta et al. 2011; Rousseaux and Gregg 2012). The likely biases in the Chl-a estimates are often not appreciated by the modelling community. Modellers sometimes misinterpret mismatches as model defects that are actually potentially due to satellite product biases, or worse, have tuned their models or assimilated the products, to capture the ocean colour derived Chl-a even where it is likely biased. It should be made clear that Chl-a from satellites is merely an estimate of Chl-a and that there are likely (though not necessarily known) biases in these data.

Case Study: Modelling Actual and Satellite-Like Derived Chl-a

Dutkiewicz et al. (2018) explored the difference between Chl-a estimated from a blue-green reflectance ratio and the Chl-a that was explicit in the numerical model. They used a global 3-dimensional physical, biogeochemical, and ecosystem model that included nine phytoplankton functional types which had dynamically changing Chl-a concentrations. The sum of this explicit and dynamic Chl-a across all phytoplankton types can be referred to as model "actual" Chl-a (Figure 4.5a). When this model "actual" Chl-a is compared to satellite derived Chl-a (OC-CCI OC4 algorithm), large biases are evident, especially at high latitudes (Figure 4.5b).

The model also had radiative transfer of spectral irradiance in 25 nm bands between 400 and
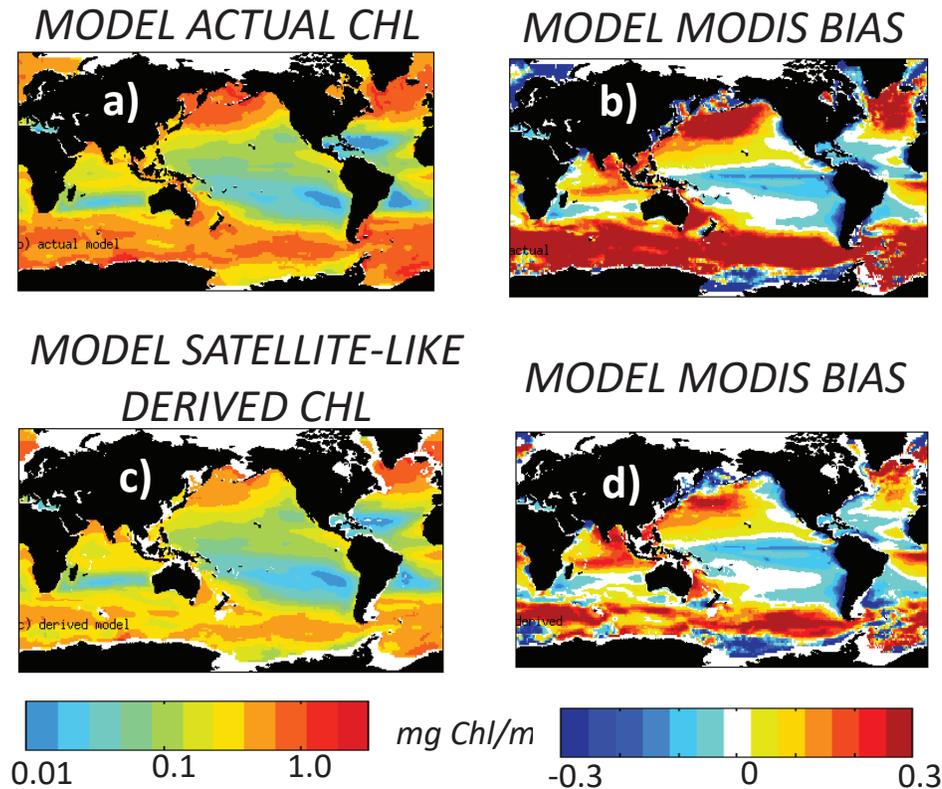
Figure 4.5 Comparison of model Chl-a to satellite-derived Chl-a. (a) Model annual average "actual" Chl-a. This is the explicit and dynamic Chl-a found from the phytoplankton in the model. (b) Bias between model and MODIS satellite derived Chl-a. Positive suggests that model over-estimates the Chl-a found from the satellite measurements. (c) product using the blue-green reflectance ratio and a 4th order polynomial to estimate Chl-a in the manner used in satellite products (e.g., NASA OC4). (d) Bias between model "satellite-like derived" Chl-a and the MODIS-derived Chl-a.

700 nm, that allowed the model to output spectral surface upwelling irradiance, output that is similar to measurements made by ocean colour satellites. Using these reflectances, the authors were able to construct an "ocean colour-like derived" Chl-a product. They first calculated the blue-green reflectance ratios and then used a 4th order polynomial (similar to the NASA OC4 algorithm) to estimate this "satellite-like derived" product (Figure 4.5c). When this satellite-like derived Chl-a product was compared to the real MODIS product its bias (Figure 4.5d) was less that the "actual" Chl-a bias (Figure 4.5b). If the authors had evaluated their model as in Figure 4.5c, they would have thought their model was performing far worse than when they compared more similar variables, that is, Chl-a calculated in the same way as the real satellite products are calculated. This study therefore showcases that ocean colour Chl-a itself has bias due to processes affecting the blue/green ratio that are not captured in the algorithm, and should be used with care in evaluation of (or assimilation into) a numerical model.

4.1.4 Carbon pools

Deriving information about the different carbon pools from ocean colour satellite data is a more recent development than deriving Chl-a (Legendre and Michaud 1999; Stramski 1999; Stramski et al. 1999; Sathyendranath et al. 2009, Chapter 2.3.3), largely due to the difficulties of measuring the different pools

of carbon *in situ* (e.g., dissolved, particulate, organic, inorganic, living, dead, autotrophic, heterotrophic pools). However, the carbon pools are the products that link to the more fundamental variables in numerical models (Figure 4.1). Carbon pools are comprised of a range of different compounds and entities, which vary globally. The components impact ocean colour via the absorption and/or scattering properties of the different constituents. The optical properties of each constituent are determined by its composition and structure, so that information on, for example, particle size is contained within the ocean colour signal (e.g., Vaillancourt et al. 2004; Hirata et al. 2008; Kostadinov et al. 2009). Consequently, a number of approaches to derive carbon pools from ocean colour have been developed (Section 2.3.3 in Chapter 2). The different constituents and their varied optical properties present a complex problem for retrieving information from ocean colour. Algorithms use a range of different approaches, assumptions and empirical datasets, even with diverse approaches to derive the same property.

The properties of carbon pools that are important for driving ocean colour (e.g., particle size) are often poorly resolved by biogeochemical/ecosystem models, with some pools (e.g., bacteria, detritus) resolved simply, or omitted entirely. Furthermore, any model that uses fixed elemental ratios will encounter a further potential source of mismatch to ocean colour products.

4.1.4.1 Particulate organic carbon

Field measurements of particulate organic carbon (POC) commonly capture the carbon associated with (living or dead) particles retained on a filter (Menzel and Vaccaro 1964), with the precise particle size depending on the method used. GF/F filters are typically used (with approximate 0.7 μm pore size) and samples may also be pre-screened to remove a large particle fraction.

In terms of IOPs, POC would fall under the absorption and scattering from both non-algal particles (NAP) and phytoplankton (Evers-King et al. 2017). Backscattering from the organic or inorganic pools cannot usually be discerned, and thus is most commonly defined as particulate backscatter, which incorporates all living and non-living particles. The relationships between satellite-derived particulate backscatter and POC (e.g., Stramski et al. 1999; Cetinić et al. 2012) account for the inclusion of the non-organic contribution to particulate backscatter through their empirical relationships.

A range of different assumptions and *in situ* datasets have been used to generate a variety of satellite POC products (Evers-King et al. 2017). Algorithms typically use satellite-derived Chl-a (where variability is driven by blue/green reflectance ratio) or satellite-derived backscatter, combined with empirical relationships between these observations and POC measured *in situ*. Some algorithms use additional information including satellite-derived K_d (Gardner et al. 2006). The spatial and temporal variability in POC satellite products is largely determined by the base measurement (reflectance ratio or backscatter) while the absolute magnitude is generally determined by the empirical relationships that scale the base measurement to POC. Satellite products will have different mismatches in variability and absolute magnitude when compared to model output (e.g., satellite products based on backscatter will have different variability to those based on Chl-a and blue/green reflectance ratio). The various satellite products are biased to different components of the POC pool, depending on the observations used in the algorithm. For example, products based on empirical relationships with *in situ* measurements of POC retained on GF/F filters are biased towards the variability of larger phytoplankton (>0.7 μm) and potentially small zooplankton that comprise this material. The satellite products are based on different assumptions and empirical datasets resulting in varied biases, and thus different mismatches when compared to model output.

Modellers often use the name "POC" to refer to the fraction of "dead" phytoplankton or zooplankton that exists as detritus, at least in models with an explicit detritus pool (Wild-Allen et al. 2010; Dutkiewicz et al. 2015a). Detrital carbon is lost to the dissolved (organic and/or inorganic) pools via simple functions representing degradation and mineralization. The description of detritus is often highly simplified, partly because it is hard to constrain, with the goal typically to be able to reproduce observed patterns

of POC decay with depth, rather than POC concentrations themselves. Furthermore, parameterizations of the bio-optical properties (mass-specific absorption and scattering) of POC used in coupled optical-biogeochemical models (Baird et al. 2016) are typically based on a small number of discrete water samples (e.g., Blondeau-Patissier et al. 2009). These discrete samples represent a specific mixture of POC sources that are unlikely to accurately represent the time-evolving combinations of POC sources that a model may endeavor to represent (i.e., changing fractions of labile and refractory detritus from dead phytoplankton and zooplankton). Models also omit small, slow turnover, detrital particles and heterotrophic bacteria which will likely affect the magnitude of the difference between modelled and satellite-derived POC but less so the variability, because those carbon constituents are relatively homogeneous globally (Ducklow 1999).

Model and satellite POC therefore differ in the components of the POC pool they represent. Most models explicitly resolve phytoplankton, zooplankton and detritus, but whether or not all of these pools should be combined when comparing to a particular satellite POC product is not straightforward. Nor is it straightforward to determine the effect of the model representation of heterotrophic bacteria, viruses or the myriad different sized other labile/refractory particles that influence the ocean colour product in a model-satellite comparison.

4.1.4.2   Phytoplankton carbon

The phytoplankton carbon pool is more clearly-defined than POC. Field measurements of phytoplankton carbon typically use cell enumeration by microscopy (cells larger than around 2 μm) and/or flow cytometry (cells smaller than around 5 μm) and then convert cell counts to carbon biomass using published conversion factors (e.g., Menden-Deuer and Lessard 2000; Tarran et al. 2006). Phytoplankton carbon may also be measured directly using carbon within the phytoplankton component separated using cell sorting flow cytometry (Graff et al. 2012; Graff et al. 2015). Most biogeochemical and ecosystem models have a clearly defined phytoplankton pool (e.g., Aumont et al. 2003; Le Quéré et al. 2005), though recent efforts to incorporate mixotrophy in models (e.g., Ward and Follows 2016) may make it more complicated to define the phytoplankton pool.

A range of approaches to derive phytoplankton carbon from satellite have been attempted (Martinez-Vicente et al. 2017). Products are typically based on satellite-derived Chl-a (the blue/green reflectance ratio) or particle backscattering derived semi-empirically from reflectance using information on absorption and scattering components (Behrenfeld et al. 2005; Martinez-Vicente et al. 2013). Some use information on allometry and additional combinations of satellite-derived and *in situ* measurements (Kostadinov et al. 2016; Roy et al. 2017). Most algorithms include empirical relationships between phytoplankton carbon and *in situ* POC. In a similar way to POC, implicit biases within the satellite products will affect the comparisons to model output. For example, products that are based on backscatter and/or flow cytometric data may be biased by the smaller (pico)phytoplankton, whilst those using microscopy data or incorporating POC measured on GF/F filters may be biased towards larger phytoplankton. Since phytoplankton carbon products are based on similar (or the same) base measurements as some POC products, and some use simple empirical relationships between POC and phytoplankton carbon as part of the algorithm, some satellite-derived phytoplankton carbon products are quite similar to those for POC.

Case Study: POC and Phytoplankton Carbon

Output from a global numerical model (Dutkiewicz et al. 2015a) was compared to selected satellite POC and phytoplankton carbon products as part of the ESA-funded "Pools of Carbon in the Ocean" project. The model resolved 3-dimensional physical, biogeochemical and ecosystem processes that included nine phytoplankton functional types and two zooplankton size classes. Model "POC" for comparison to satellite products was defined as phytoplankton, detritus and small zooplankton pools. Small (but not

large) zooplankton were included because they could contribute to the *in situ* POC used to develop the satellite algorithms. The model output selected for the comparison was monthly averages (to capture seasonal cycles but not temporal mismatches such as blooms occurring a few days too early or late), the time period 1998 – 2006 (to match ESA OC-CCI products) and water depths >1000 m (because the model was not designed to resolve coastal processes).

The model output was compared to four satellite POC products and three phytoplankton carbon products (Figure 4.6). For POC, the first product is based on an empirical relationship between a reflectance band ratio and POC (Stramski et al. 2008, method 1), and the second is based on backscatter (b_{bp}) analytically derived from ocean colour combined with an empirical relationship between b_{bp} and POC (Stramski et al. 2008, method 2). The third is based on b_{bp} derived from reflectance and K_d, and empirical relationships between total particle scattering (b_p), b_{bp} and POC involving Chl-a (Gardner et al. 2006). For phytoplankton carbon, C_{phy}, the first product is based on satellite Chl-a (i.e., a reflectance band ratio) and an empirical relationship between Chl-a and C_{phy} (Sathyendranath et al. 2009). The second is based on analytically derived b_{bp} and empirical relationships between b_{bp}, POC and C_{phy} (Behrenfeld et al. 2005) and the third is based on b_{bp} and empirical relationship between b_{bp} and C_{phy} (Martinez-Vicente et al. 2013).

The model overestimated the range in C_{phy} and POC between high and low biomass conditions relative to all the satellite products, attributable to known issues, for example, the model over predicts the range in phytoplankton biomass partly due to grazing and because the model resolution is too coarse to capture certain physical processes in the gyre regions (Dutkiewicz et al. 2015a). The similarity in the mismatches between model and satellite-derived phytoplankton carbon and POC products is explained by the ratio of phytoplankton carbon to POC in the model being almost constant. Furthermore, the model POC is dominated by phytoplankton, and the detritus and small zooplankton carbon co-vary closely with phytoplankton biomass. Some of the satellite approaches also assume a fixed contribution of phytoplankton carbon to POC within their algorithms (Behrenfeld et al. 2005; Sathyendranath et al. 2009).

The correlations between the model output and satellite products based on the reflectance band ratio were similar to one another, were more linear, and had lower RMSE compared to the correlations for products based on backscatter (Figure 4.6), demonstrating the importance of the base measurement for driving the variability within satellite products, and hence the variability in the model-satellite comparison. Since the overall magnitude of the satellite-derived carbon concentrations is largely determined by the empirical relationships within the algorithms, the magnitude of the difference between model and satellite is likely to stem from disparities between the model POC pools and the pools captured by the *in situ* measurement. In general, model POC was underestimated more than phytoplankton carbon, potentially because the model lacks many components of the POC measured *in situ* (such as refractory particles). The modelled ratio of phytoplankton carbon to POC was relatively high, and relatively constant, again potentially reflecting missing components of POC in the model (Smith and Baker 1978; Martinez-Vicente et al. 2013; Graff et al. 2015), though *in situ* observations of this ratio are sparse.

Whether or not large zooplankton were included in the model POC had a small effect on the difference between model and satellite-derived POC (on average <2% of the difference globally), with the impact mainly in high biomass conditions where large zooplankton carbon is most important.

4.1.4.3 Dissolved organic carbon

Dissolved organic carbon (DOC) is one of the largest pools of carbon in the ocean. DOC is an important component in biogeochemical and ecosystem models, and is usually a state variable (i.e., it is a field explicitly stepped through time in the model, see Section 3.2.1 in Chapter 3) and as such, the concentration at any location at any time is part of the model. In biogeochemical models, DOC is usually parameterized as being formed from lysis and mortality of phytoplankton and zooplankton
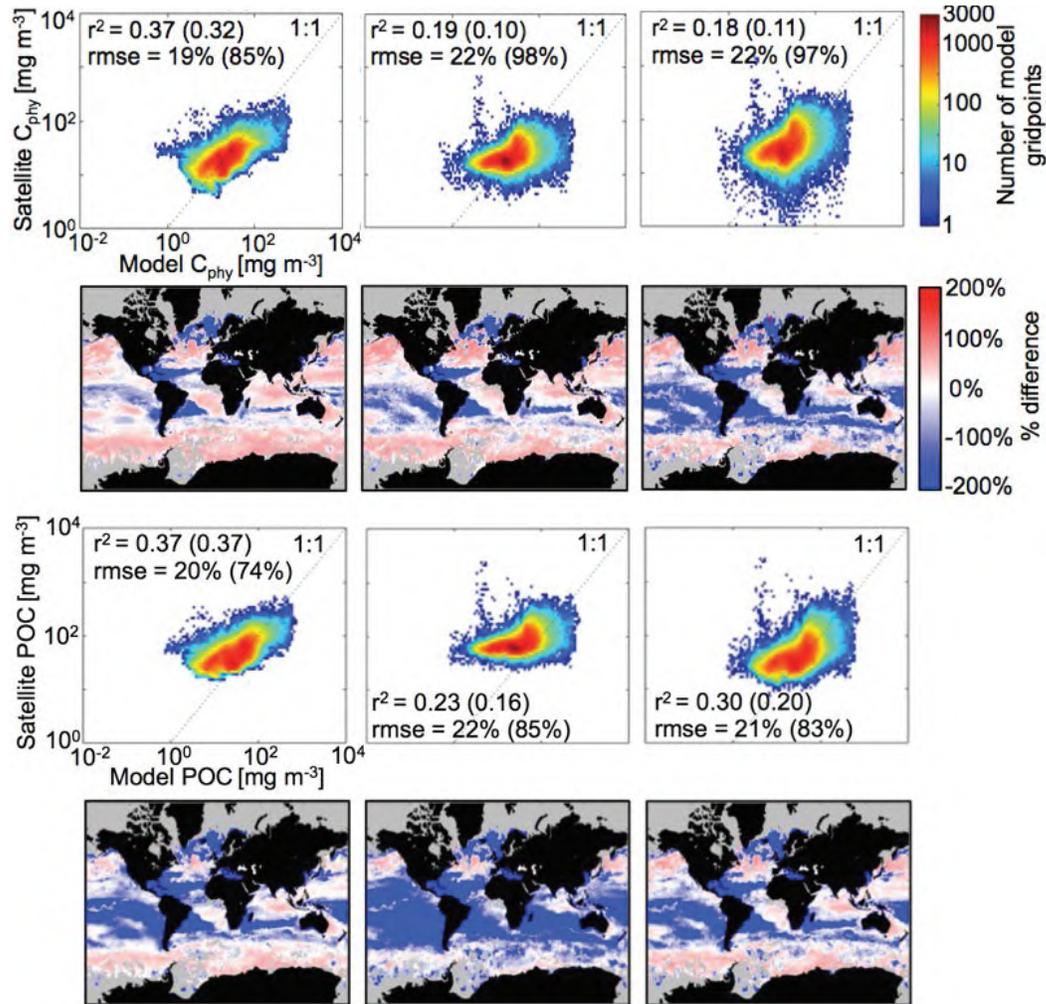
Figure 4.6 Pools of carbon from ocean colour products and models. C_{phy} (top two panels) and POC (bottom two panels) from the model compared to C_{phy} and POC derived from selected satellite approaches. For each comparison the top panels are surface monthly-mean values at all model grid points from 1998–2006 compared to values from each satellite-derived product, where colours indicate the density of points. The bottom panels are the percent difference between model and satellite-derived values for monthly-mean for March 1998 ((*model - satellite*) / model). Correlation statistics (r^2 and the root mean square error as percentage of model values) are provided for the log-log relationship and linear relationship (in parentheses). For C_{phy}, the satellite products are from left to right: Sathyendranath et al. (2009) (left panels), Behrenfeld et al. (2005) (middle panels) and Martinez-Vicente et al. (2013) (right panels). For POC the satellite approaches from left to right are: Stramski et al. (2008, method 1), Stramski et al. (2008, method 2), and Gardner et al. (2006). We acknowledge help from Hayley Evers-King and Victor Martinez-Vicente with the satellite-based calculations shown in this figure.

at a size-dependent proportion to detritus, and as a result of sloppy feeding. DOC is remineralised to dissolved inorganic carbon (DIC) usually with a very simple constant timescale parametrization, although some models do include explicit heterotrophic bacteria (Fasham et al. 1990; Stock et al. 2014b; Zakem et al. 2018). In general, the DOC in models does not include the refractory pool.

Limited *in situ* data hinders the accurate representation of DOC concentration in both models and satellite products. There is very little *in situ* data on either the total DOC or the proportion that absorbs light (CDOM, discussed in Sections 2.3.3 and 4.1.2). Rather, *in situ* observations are mainly of a_{CDOM} (e.g., Nelson and Siegel 2013). This is due to the ease of capturing the coloured fraction (i.e., CDOM) of DOM with optical approaches that can be applied to satellite and autonomous observations (Siegel et al. 2005; Xing et al. 2012). CDOM is particularly important optically, absorbing strongly in the blue portion of the spectrum (see Figure 2.4), although CDOM is only a small part of the DOM pool.

Models typically do not resolve CDOM (though see Bissett et al. 1999b; Xiu and Chai 2014; Dutkiewicz et al. 2015a; Baird et al. 2016, and discussion in Section 4.1.2). When they do, it is usually because they include an optical component that requires this important water constituent. Such models do not use CDOM to evaluate or tune their DOC fields, as these are not comparable quantities.

There is no global algorithm for obtaining DOC from satellite measurements. In coastal waters, however, CDOM absorption tracks the spectral shape of DOC due to the presence of a strong gradient of relatively degraded, terrestrial-derived material and conservative mixing (Stedmon and Markager 2003; Fichot and Benner 2011; Mannino et al. 2014; Cory and Kling 2018). These consistent relationships between a_{CDOM} and DOC do not hold up in open ocean waters due to relatively low production rates and strong photodegradation in surface ocean waters (Nelson et al. 2010; Helms et al. 2013).

4.1.4.4   Particulate inorganic carbon

Particulate inorganic carbon (PIC) is produced by marine calcifying organisms including corals, foraminifera, pteropods and coccolithophores. Coccolithophores, a major producer of PIC, are surrounded by a shell made of coccoliths. The coccoliths can detach from the cells thereby largely determining the PIC concentration of seawater. Balch et al. (2005) developed an algorithm to quantify the amount of PIC using normalized water-leaving radiance at 440 and 550 nm. Models may have a representation of coccolithophores, or PIC, or both. It is therefore important to keep this difference in mind when comparing satellite data and model output. Coccolithophores and PIC are not the same. Moreover, coccolithophores are not the only source of PIC. For instance, other marine organisms (e.g., foraminifera) produce PIC, and there are terrestrial sources. Thus models and ocean colour algorithms such as Balch et al. (2005) underestimate actual PIC. The difference between total PIC, and that produced by coccolithophores alone, need to be remembered, especially since PIC production affects model alkalinity.

Case Example: Modelling of PIC Production

Few models incorporate explicit PIC. In the model of Gregg and Rousseaux (2017), PIC is produced by the growth of coccolithophores and is lost via sinking and dissolution. PIC is produced as a fraction (25%) of the coccolithophore growth rate (Gregg and Casey 2007a) minus respiration. The PIC sinking rate is represented as an exponential function of concentration, assuming that large concentrations of PIC are associated with larger coccolith size. Dissolution follows Buitenhuis et al. (2001), except that no dissolution is allowed for depths shallower than the calcium carbonate compensation depth, which was defined as 3500 m. The PIC from the model was compared to the global PIC using the algorithm developed by Balch et al. (2005). They found that the model PIC distributions largely corresponded to the satellite distributions, although local maxima in the southern central North Pacific and the Greenland are subdued in the model (Gregg and Rousseaux 2017).

4.1.5   Phytoplankton types/groups

Satellite approaches to estimate phytoplankton functional types (PFTs) began with identifying dominant single groups and have evolved to capturing multiple groups. The identification and mapping of the

distributions of coccolithophore blooms (Brown and Yoder 1994; Brown and Podesta 1997), *Tricho-desmium* (Subramaniam et al. 2002), harmful species (Hu et al. 2008; Tomlinson et al. 2009; Stumpf et al. 2016), and diatoms (Sathyendranath et al. 2004; Jackson et al. 2011) were based on spectral ratios, thresholding, and the unique spectral signatures (absorption and backscattering) associated with each phytoplankton group. The detection of multiple groups, such as phytoplankton size clas-ses, phytoplankton taxonomic composition and particle size distribution (that includes all particles, not phytoplankton alone) have several different approaches (IOCCG 2014; Bracher et al. 2017; Mouw et al. 2017, see Section 2.3.7 and Table 2.1). Abundance-based approaches rely on phytoplankton concentration (expressed as chlorophyll concentration or absorption coefficient, e.g., Uitz et al. 2006; Aiken et al. 2007; Brewin et al. 2010; Hirata et al. 2011) as an indicator of phytoplankton community composition. Radiance-based approaches classify multiple PFTs based on the spectral bounds of the shape and/or magnitude of satellite-observed spectral remote sensing reflectance (e.g., Alvain et al. 2005, 2008). Absorption-based approaches rely on the spectral shape or magnitude of phytoplankton absorption to estimate phytoplankton size composition (e.g., Ciotti and Bricaud 2006; Hirata et al. 2008; Mouw and Yoder 2010; Brewin et al. 2011; Devred et al. 2011; Fujiwara et al. 2011; Roy et al. 2011; Bricaud et al. 2012) and phytoplankton taxonomic groups (Bracher et al. 2009; Sadeghi et al. 2012). Scattering-based approaches utilize particulate backscatter to estimate particle size distribution (Montes-Hugo et al. 2008; Kostadinov et al. 2009; Kostadinov et al. 2010). Finally, some authors also rely on other information such as blending physical data in addition to bio-optical information from ocean colour to distinguish between different phytoplankton groups (e.g., Raitsos et al. 2008; Palacz et al. 2013).

Phytoplankton composition in biogeochemical/ecosystem models are mostly based on biogeoche-mical functions. Numerical models in general encompass limited diversity of phytoplankton. They rarely (if ever) capture a single species, but usually parameterize groups of phytoplankton as a single entity, e.g., "diatoms" (which will have a silica requirement) and "non-diatoms" (which do not). Many models include phytoplankton functional groups such as diatoms, pico-cyanobacteria, and diazotrophs (e.g., Baretta et al. 1995; Bopp et al. 2005; Gregg and Casey 2007a). Some models capture explicit size classes instead (e.g., Ward et al. 2012; Ward and Follows 2016). A few studies have assimilated PFT ocean-colour products into ecosystem models (Ciavatta et al. 2018, 2019; Skákala et al. 2018, see Chapter 6).

Studies suggest that model and satellite estimates of phytoplankton biogeography often do not compare well in terms of phytoplankton dominance patterns (Vogt et al. 2013; Ciavatta et al. 2019), and in the timing of blooms (Hashioka et al. 2013; Kostadinov et al. 2017). These discrepancies may suggest that the numerical models are not sophisticated enough, or the satellite products are not well developed, or (more likely) that the groupings that each capture are different enough that the comparison is ill-posed. This latter has been identified as a major gap in current ocean colour phytoplankton diversity studies (Bracher et al. 2017), which recent studies are tackling by coordinating the definition and computation of PFTs from ocean-colour algorithms and marine ecosystem models (e.g., the regional algorithms by Brewin et al. 2017a and the ERSEM model in Ciavatta et al. 2018).

Case Study: Phytoplankton Size Class Phenology Across Satellite Imagery and Models

Satellite PFT algorithms have a variety of phytoplankton classes, units, and satellite product outputs, precluding direct comparison of algorithm performance. Instead, Kostadinov et al. (2017) compared phenological cycles (bloom timing) between PFT algorithms and several CMIP5 models to identify spatial patterns of agreement and disagreement (Figure 4.7). The timing, amplitude and duration of blooms of microplankton were compared. The standard NASA chlorophyll concentration product was used as reference. While the phenology of the PFT algorithms on global scales agrees to a first order, there were significant differences. This points to a need for a more extensive comparison effort with a

comprehensive harmonized *in situ* data set incorporating co-located radiometric, bio-optical (pigment concentration, IOPs) and derived biological quantities (phytoplankton Chl, size structure, etc.). This will allow the ability to resolve whether differences are due to algorithm uncertainties or to actual biogeophysical differences (Bracher et al. 2017). The comparison of phenological parameters in CMIP5 model output (diatom carbon biomass) to those of satellite imagery suggest the models represent the timing of the bloom well, but biases are found in other phenological characteristics in CMIP5 models. The biases include: 1) more pronounced seasonal variability in the models, 2) clear seasonality is evident in the satellite data along a zonal band at 30° latitude, but this feature is not reflected in the models, and 3) models exhibit a single annual biomass peak over most of the ocean, except for the Equatorial band, whereas the satellite data capture secondary blooms in temperate latitude zonal bands that are not reflected in the climate models. These biases are attributed to an over-simplification of processes in models and a damped response to interannual variability than in reality (Cabré et al. 2016).
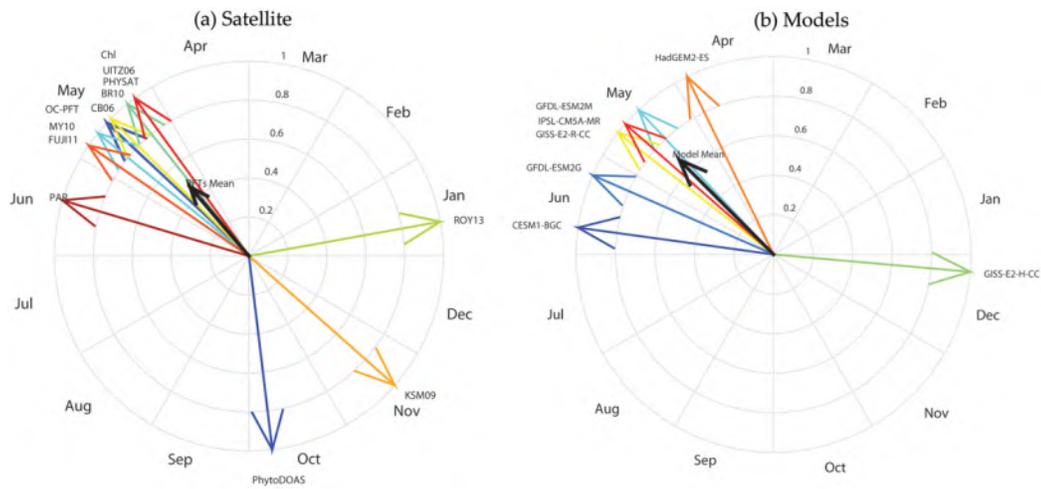
Figure 4.7 Comparison of satellite and modelled microplankton phenology in Longhurst (1998) North Atlantic Drift province. (a) Months of maxima of the satellite PFT algorithms; (b) Months of maxima of the models' diatom carbon biomass for the North Atlantic Drift Region. Image adapted from Kostadinov et al. (2017) with permission from Elsevier.

4.1.6   Primary production

The most widely used satellite primary production products are based on satellite derived Chl-a. Such products use an empirical model that derives the primary production from Chl-a, surface irradiance, light attenuation (often derived from Chl-a) and/or mixed layer depth and the photosynthetic efficiency described by photo-physiological parameters (often derived from SST, e.g., the VGPM model, Behrenfeld and Falkowski 1997b). The empirical models to quantify primary production from this information are well established (e.g., Ryther and Yentsch 1957; Platt 1986; Sathyendranath et al. 1989, 1991; Antoine and Morel 1996; Behrenfeld and Falkowski 1997b). Phytoplankton change their cellular Chl-a content to acclimate to their nutrient and light environment. To reduce the uncertainties in primary production estimates inherent in using Chl-a, a carbon-based approach was developed. The carbon-based primary production models utilize satellite-estimated b_{bp} to estimate phytoplankton carbon biomass which is related to growth rates from chlorophyll-to-carbon ratios (Behrenfeld et al. 2005; Westberry et al. 2008). Further, the "Carbon, Absorption, and Fluorescence Euphotic-resolving"(CAFÉ) model has been developed to take advantage of several recent remote sensing advances allowing for the use of phytoplankton absorption, particulate backscattering and phytoplankton carbon to estimate

net primary production (NPP). NPP is calculated by the CAFÉ model as a product of energy absorption and the efficiency by which absorbed energy is converted into carbon biomass, while growth rate is calculated as NPP normalized to phytoplankton carbon (Silsbe et al. 2016).

Primary production algorithms have additional uncertainties relative to some of the other products discussed earlier (they are more "derived", see Figures 4.1 and 2.8). These include, but are not limited to, the assumption of the temperature dependence of primary production, the relationship between chlorophyll and carbon, the relationship between surface and depth integrated carbon/chlorophyll, the assumption of the light dependence of photosynthesis, and light attenuation in the water column/euphotic zone. Carr et al. (2006) argue that differences in the temperature dependence of primary production underlie the primary differences in the spatial patterns in the primary production algorithms. They also highlight the high uncertainty in the predicted global net primary production magnitude, which likely arises from the collective effects of all the other factors.

The satellite products are validated against *in situ* observations, with the most established technique to quantify primary production in the field being the assimilation rate of radio-labelled carbon (^{14}C) during incubations of water samples at simulated *in situ* conditions (Steemann Nielsen 1952). However, the approach is time consuming and has large uncertainties that remain difficult to quantify (Marra 2002). For example, whether the incubations measure gross or net primary production (which includes nighttime respiration) depend on the incubation timescale (typically a few hours to 12 h or 24 h), with incubations typically representing something between the two. There are also artifacts that mean there is a large degree of variability between studies (Marra 2002). Additional physiological information is required by many satellite algorithms, including the maximum quantum yield and/or maximum Chl-a-specific carbon fixation rate, which are obtained from short-term (typically 2–4 h) ^{14}C incubations over a light gradient (photosynthesis vs. irradiance experiments).

Other aspects of phytoplankton photosynthesis and community production can be observed using active Chl-a fluorescence and other isotope (e.g., oxygen) techniques, but some of these are difficult to relate to phytoplankton carbon fixation (Suggett et al. 2009; Quay et al. 2012). Although there are numerous different ways of measuring the various metrics of phytoplankton and community production, ^{14}C-derived measurements remain the benchmark because it is the most longstanding technique with reasonable data coverage (Regaudie-de-Gioux et al. 2014).

In one satellite product inter-comparison (Saba et al. 2011), the average root-mean-square-error for all algorithms and regions was around 30%, with roughly half of the mismatch between satellite products and *in situ* observations stemming from uncertainties on the input variables (in particular Chl-a), with a further fifth attributed to uncertainties in the *in situ* observations within the primary production dataset. For the 21 products, the more complex depth or wavelength resolved algorithms performed no better than simple ones (Saba et al. 2011). The high uncertainty in satellite-derived primary production leads to the large range in predicted global NPP magnitude (Carr et al. 2006). Given the large uncertainties in the primary production products, space agencies do not provide these production products. However, such products are available elsewhere (e.g., `https://www.science.oregonstate.edu/ocean.productivity/`).

Models also use parameterizations for primary production that relate the biomass (or Chl-a), light and primary production, but typically depend on nutrient availability as well (such as the Geider et al. 1998 formulation). Model primary production is a particularly important property to validate against observations because it is often the rate within the biogeochemistry and ecosystem processes that can be most reliably constrained. However, the large uncertainties in *in situ* data and satellite products make model validation problematic. Discrepancies are to be expected if the satellite algorithm and model include different assumptions about photophysiology, including the quantum yield, and different assumptions about light penetration and/or predict different mixed layer depth or vertical distribution of Chl-a.

Variability between different models and satellite products have been investigated in a series of round robin inter-comparison exercises (Campbell et al. 2002; Carr et al. 2006; Friedrichs et al. 2009; Saba et al. 2010; Lee et al. 2015a). Mismatches between the models, satellite products and *in situ* data

remain large (Lee et al. 2014). When compared to *in situ* data, ocean colour products have been shown to perform no better than biogeochemical models (Friedrichs et al. 2009).

4.2 Temporal and Spatial Mismatch of Model Outputs and Ocean Colour Products

There are differences in the spatial and temporal resolution of the model variables and the ocean colour products. For instance, most global models are often on a coarser grid (around 100 km) than the ocean colour measurements (<10 km), while on the other hand coastal models may have resolution at much finer scales (<1 km) than that provided by ocean colour products. Moreover, ocean colour products typically only cover the globe every few days, giving incomplete instantaneous coverage. There are issues comparing model temporally averaged variables (which include every time step and every location) and satellite "climatologies" (which contain incomplete instantaneous coverage and can miss data from clouds and, in high latitudes, when there is insufficient sunlight). Additionally, ocean colour products are provided from two-dimensional measurements, while model variables are three-dimensional. These issues make comparing the variables that models output difficult to compare to ocean colour products. Here we discuss just a few of these issues, and leave others for Chapter 5.

4.2.1 Temporal gaps in ocean colour measurements

Models have regular grids, and temporally, the output is usually statistical "means" over days, months or years that includes input from every timestep: there is no missing data. On the other hand, satellite measurements are not continuous, with missing data due to orbital periods, clouds, aerosols etc. Satellite products are potentially misleadingly provided as "climatologies", where all available data are averaged together. One potential solution to this is to limit data with gaps, such as satellite data, to seasonal climatologies (mean for several years, but where the months are kept separate e.g., a mean of all Januaries, all Februaries, etc.). The word "climatology" to a modeller usually refers to a long term mean (e.g., over several years), with continuous output (i.e., a true mean). Thus there can be distinct confusion between the two communities.

In high latitudes, a satellite Chl product could have as little as six months of data for an annual "climatology", due to low sun angle, and likely further reduced by cloud coverage. Such annual averages cannot be compared to model "climatology" which would include all the model output i.e., output from all months, and no missing data due to cloud cover (though note newer technologies such as CALIOP will provide more complete coverage of the globe all year). Gregg and Casey (2007b) showed, for example, that persistent clouds, especially in the high latitudes such as in the North Pacific, lead to an overestimate of chlorophyll concentration from the satellite data since the satellite "sees" the chlorophyll only when there are no clouds (which generally corresponds to a period of higher growth for phytoplankton). In the North Pacific, clouds can limit the number of available days of satellite ocean colour observations to <5 days per month. This sampling problem can lead to biases between 6–13% in chlorophyll concentration estimates for the months of May through July, a period of high growth for phytoplankton in this region (Gregg and Casey 2007b). Similarly, aerosols can lead to underestimates in chlorophyll concentration from ocean colour by ~30% by selective sampling in lower aerosol thickness periods, which correspond to lower phytoplankton growth periods. High solar zenith angle, clouds, aerosols, sun glint, inter-orbit gaps and sensor tilt changes can each represent a source of gaps in ocean colour data (See Chapter 8, Case Study 8.1.1 for more details on this study). The seasonal and interannual variability of these factors can be large, and further complicate the quantification of ocean colour-derived data as well as any seasonal and trend analysis. It is therefore crucial to keep these gaps in mind when using any satellite ocean colour data for research purposes, and in particular, for model validation (see Chapter 8, Case Study 8.1.2 for more on this topic).

4.2.2 Matching in time

The dynamic nature of the optical properties of the ocean means that any mismatch in the timing of observations and models needs to be minimized. For calibration of satellite algorithms such as OC3M, satellite and *in situ* observations usually occur within 3 h in the open ocean, and 1/2 h in the coastal ocean. This short time window is to account for changing chlorophyll concentration and the movement of fronts, among other processes. A similar maximum time mismatch should be used for model-observation match-ups.

A second reason to ensure a tight match-up in time of models and satellite observations is the changing solar angle. Most satellite measurements are apparent optical properties (AOPs) meaning that they depend on the solar angle (though note that geostationary satellites will help provide corrections for these issues). The satellite measured remote-sensing reflectance and normalized water leaving radiances account for the effect of solar angle on incoming solar radiation per m^2 by normalizing to the solar radiation at zero zenith value. But this normalisation does not account for the effect of solar angle on the vertical attenuation of light through the water column. Thus satellite observations at high zenith angles will be measuring properties closer to the surface than those at a zero zenith angle. The error in the mismatch in time due to the changing solar zenith angle will be greatest at high latitudes, as well as when the time between the local noon and the observation is greatest.

4.2.3 Biases in ocean colour products due to spatial resolution of *in situ* measurements

The spatial and temporal sampling of the *in situ* data used to develop the algorithms also lead to biases in satellite products. *In situ* datasets may be biased to coastal rather than open ocean regions, or vice versa, with empirical relationships often omitting data from particular ocean basins. Some regions may be strongly biased to spring and summer months. Some regions may not conform to the empirical relationships used in the algorithms. A recent model study suggested significant deviations from the Chl-a vs. blue-green reflectance ratio used in the OC4 Chl-a algorithm in some under-sampled regions including the Southern Ocean (Dutkiewicz et al. 2018, see Chapter 8, Case Study 8.2.2). Assessing the uncertainties in regions with no data is clearly problematic and should be born in mind when validating model output for these regions.

4.2.4 Mismatches due to depth resolution: comparing 2- and 3-dimensional quantities

Biogeochemical models produce a range of three-dimensional variables (e.g., nutrients, phytoplankton biomass, Chl-a) that are properties of the volume of each individual grid cell (see Section 3.2.2. in Chapter 3). Pixelated ocean colour observations are determined from the two-dimensional water leaving irradiance fields with an orientation aligned with the satellite path, and with contributions from a range of depths in the water column.

A number of methods have been undertaken to convert a 3-dimensional model variable to the 2-dimensional field. The simplest, and most common approach, is to assume the surface model field is equivalent to the ocean colour product. However, this is a large assumption unless the water column is well mixed to the greatest depth of returning light. In a slightly more sophisticated approach (Gordon and Clark 1980), a 2-D ocean colour-like product (C_{2D}) can be determined from the model 3-D field, $C(z)$, using an exponential weighting function, $f(z)$, that accounts for arriving irradiance having been attenuated differentially at each depth.

$$C_{2D} = \frac{\int_0^z C(z)f(z)dz}{\int_0^z f(z)dz}$$

$$f(z) = \exp\left(-\int_0^z 2K_d dz\right)$$

This approach also has limitations, in particular in assuming that the downwelling and upwelling pathlengths are the same. However, it has been used to calculate remote-sensing reflectance from depth-resolved model IOP fields (Baird et al. 2016) and depth-resolved chlorophyll fields (Moline and Prezelin 2000).

A further complication of this approach for ocean colour products using band ratios is to determine what K_d value to use. In fact, this illustrates an inherent limitation of band ratio algorithms: since bands have a different profile of vertical attenuation, the ocean colour product is based on the return of light from two different depth profiles. In the case of satellite-derived chlorophyll, it is the ratio of green light returned from, on average, shallower depths than the returned blue light. This inconsistency highlights the advantages of calculating satellite-like ocean colour products from model remote-sensing reflectance (see e.g., Baird et al. 2016; Gregg and Rousseaux 2016; Dutkiewicz et al. 2018, see also case study in Section 4.1.3). The satellite-like product will be different to the model variable its name suggests it should represent, but nonetheless more comparable to the satellite-derived ocean colour product.

4.2.5 Mismatches due to uncertainty of *in situ* measurements

The uncertainty and availability of different data varies considerably, so that some measurements may have high uncertainty but abundant data and vice versa. For instance, there are far more fluorometrically-determined Chl-a measurements than those measured via HPLC, although HPLC determination of Chl-a is more accurate. Both data types are more abundant and have lower uncertainty than phytoplankton carbon or primary production. The uncertainties on many types of *in situ* measurements remain difficult to quantify (a notable example is primary production, Williams and Purdie 1991). Moreover, when models or satellite algorithms use empirical relationships or observations within them, then some aspects of the methodological errors will be incorporated into those models or satellite products (e.g., McKee et al. 2014, 2015b). Obviously, presence of uncertainties does not mean that we should stop using *in situ* data (or the satellite products informed by them), but it is important to be aware of the existence of, and potential implications of, such uncertainties.

4.3   Summary and Recommendations

In this chapter we have highlighted how ocean colour products, *in situ* observations and biogeochemical model output do not compare cleanly (schematically shown in Figure 4.2). There are mismatches in what is understood by a variable's name, uncertainties are not well understood or quantified, and there are biases linked to missing data. These mismatches are not well understood and provide a hindrance to the best use of ocean colour (and models). Miscommunication linked to the lack of dialogue between ocean colour product producers and the users (in this case the modelling community) can lead to incorrect use of products. Both models and ocean colour products contribute to our understanding of the biogeochemical and ecological processes and variability in the ocean, but in different ways. A combined use of the two tools can be valuable (see e.g., Chapter 7), but require a better understanding of the advantages and disadvantages of each.

We have stepped through the ocean colour products in order of their "derivedness" (see Figure 4.1), that is the level of uncertainty that goes with each step away from what is actually measured by the satellite sensor, and the number and complexity of algorithms needed to reach a specific product (i.e., from remotely sensed reflectance, to Chl-a, to primary production, which takes the derived Chl-a as an input). The different information used, including the observed R_{rs}, theoretical and empirical

relationships, and the necessary assumptions and averaging, can each contribute to mismatches in magnitude and/or variability of properties when compared to output from numerical models.

4.3.1   Questions to consider when comparing model and satellite output

❖ Which model variables compare most directly to the satellite ocean colour product?

❖ Are uncertainties in the ocean colour product quantified (rarely), and if not, what is the level of derivedness of the product (assuming the uncertainties get larger with addition derivation)?

❖ Are there underlying biases in model or satellite products due to assumptions, optical properties or biases in *in situ* data?

❖ Which products are the priorities to 'get right' (e.g., Chl-a or carbon)?

❖ What are the biases from missing satellite measurements?

❖ What are the biases from limited *in situ* data to train or validate the ocean colour products?

It is important to consider separately the biases associated with 1) the satellite measurement that provides the spatial and temporal variability (e.g. reflectance ratio, analytically-derived backscatter) and 2) the (normally empirical) relationships that scale the variability to the property of interest (e.g., reflectance ratio vs. Chl-a, backscatter ratio vs. POC). In the case of primary production, it is clear that satellite-model comparisons are model-to-model: an empirically-based satellite model and a (typically more) mechanistic simulation model. Careful consideration should be given to what information is actually being gained from the satellite product and how to interpret comparisons.

4.3.2   Should we bring model output closer to ocean colour or ocean colour products closer to model outputs?

Models could (and likely will) continue to be developed to resolve ocean colour at its least derived form (e.g., water leaving radiances) as these products are likely the most robust field to validate and assimilate (Jones et al. 2016). Though there is already a remarkable fidelity in the model connections to fields such as R_{rs} and normalized water radiance (e.g., Baird et al. 2016; Gregg and Rousseaux 2016; Dutkiewicz et al. 2018, 2019) there are still uncertainties with these model outputs. Such models will likely continue to resolve greater complexity in the optical constituents to better compare to the ocean colour observed from space. Such increased complexity could include:

❖ resolving size partitioning of particulate carbon (phytoplankton, zooplankton, detrital material);

❖ resolving additional optically important water constituents such as bacteria and viruses, and their optical make up;

❖ resolving CDOM more adequately through including terrestrial sources and a variety of turnover times.

However, to compare model output to the least uncertain ocean colour-derived quantity (e.g., reflectance) may often not be possible given the additional computational costs and added complexity of the model. Moreover, stakeholders are more familiar with quantities such as Chl-a than reflectance or IOPs. It is likely that both approaches will be needed in the future: models that link closer to ocean colour, and newer ocean colour products that are closer to the needs of more traditional biogeochemical models (e.g., that do not explicitly include optics or radiative transfer components). Such products will likely improve on the capturing of specific pools of carbon. Communication between modellers and remote sensing scientists is key to decide "where best to meet in the middle".

4.3.3   Recommendations

Models, *in situ* observations and ocean colour products are different tools to help us understand the ecological and biogeochemical processes in the ocean. At some level we can ask if we do, in fact, need to bring all three into a format where we can match them exactly, or instead use them for their best application and rather merge the understanding that comes from them. This, in the end, is a worthy outcome, especially as non-scientist stakeholders would prefer one answer (rather than three different ones). However, such integrated results will need a multi-way discussion.

We recommend that modellers should have better information about using satellite data as well as the development of new satellite approaches. Agencies providing ocean colour products should examine and explore how additional information can be presented alongside satellite products to help modellers make informed choices and interpretations. On the other hand, the needs of modellers, and the potential usefulness of models, for instance for algorithm development (see Chapter 8), should be borne in mind with new ocean colour product development.

The needs of the modelling community will change as ecosystem models become more complex (such as resolving mixotrophs, bacteria, viruses, other carbon pools). Simply developing satellite algorithms for deriving each of these (a reactionary approach) is not recommended. Consideration is needed of users' requirements, *in situ* and satellite observational constraints, and of the assumptions, errors and uncertainties of different approaches to assess the most robust comparisons between satellite products and model output.

<div style="border:1px solid #aaccee; background:#dde8f5; padding:10px; width:50%;">

Chapter 5

</div>

Ocean Colour for Model Skill Assessment

Charles Stock and Stefano Ciavatta

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

5.1   Introduction

Underlying the development of most marine ecosystem models is a desire to make inferences beyond observed patterns. This can occur spatially, across ecosystem properties, or as predictions of future ocean states. In all cases, the question arises: how much should we trust the model predictions?

In the Earth sciences, compelling attempts to confront this question arise for weather predictions and climate change projections. For weather, comparison of past predictions against observed outcomes over many years provide highly developed probabilistic frameworks for assessing and communicating prediction skill (e.g., Stanski et al. 1989; von Storch and Zwiers 2002; Jolliffe and Stephenson 2003; Mason and Stephenson 2008; Wilks 2011). Extensive retrospective forecast experiments also provide a foundation for assessing the skill of seasonal to multi-annual climate predictions (Becker et al. 2014; Kirtman et al. 2014; Stock et al. 2015). Century-scale climate change projections, however, do not have the luxury of a large catalog of past anthropogenic climate change experiments with which to quantify skill. Confidence in model projections must instead be built upon an assessment of the model's capacity to reproduce current and past climate states (Randall et al. 2007; Cheung et al. 2016).

Recent years have seen rapid growth in the use of marine ecosystem models to inform decisions in ways analogous to weather and climate predictions (Clark et al. 2001; Allen et al. 2015; Gehlen et al. 2015; Tommasi et al. 2017b). This burgeoning transition of marine ecosystem models from research, to applications, to operations, has increased the impetus for rigorous model skill assessment. In this context, a flawed assessment of the question "how much do we trust an ecosystem prediction?" does not just lead to flawed scientific conclusions, but to flawed decisions that can impact property, livelihoods and lives.

The primary obstacle in assessing the skill of ocean ecosystem model simulations and predictions is the availability of relevant observations at relevant scales. As we will describe in this Chapter, the revolutionary capacity of satellite-based ocean colour measurements to reveal ecosystem properties over a broad range of spatial and temporal scales has ensured satellite measurements a central role in ecosystem model skill assessment. The decision about which ocean colour patterns to assess models against should be based on the intended use of the model. That is, skill assessments should be "fit to purpose". A model intended to simulate the evolution of large-scale ocean biomes on decadal to century scales (e.g., Polovina et al. 2011) would be well served to capture both mean biome states and their historical inter-annual variation. In contrast, a model intended to explore the integrated effect of mesoscale or sub-mesoscale processes should faithfully simulate the spatiotemporal statistics of such features (Powell et al. 2006).

In this Chapter, we begin our discussion of ocean colour-based model skill assessment with a brief review of skill assessment metrics in Section 5.2. Section 5.3 then highlights ocean colour comparisons intended to support a range of model applications across a range of spatiotemporal scales. Examples are drawn mainly from chlorophyll comparisons and emphasize 1) the challenges posed by uncertainty and regional biases in ocean colour estimates; 2) strengths, weaknesses, and alternatives to point-to-point comparisons between models and satellite data; and 3) the need for comparisons and retrospective forecast evaluation at the space and time scales of interest for each application.

In some examples, uncertainty in satellite-based estimates are large enough to make skill assessment difficult, even when comparing quite different models. This uncertainty arises from the gap between what most satellites measure (remote-sensing reflectance at select wavelengths) and what we use these measurements to infer (e.g., plant pigment concentration). It also underscores a fundamental aspect of model skill assessment: what we observe and compare models against is an imperfect measure of the true ocean state, which is impossible to know exactly (Lynch et al. 2009, Figure 5.1).
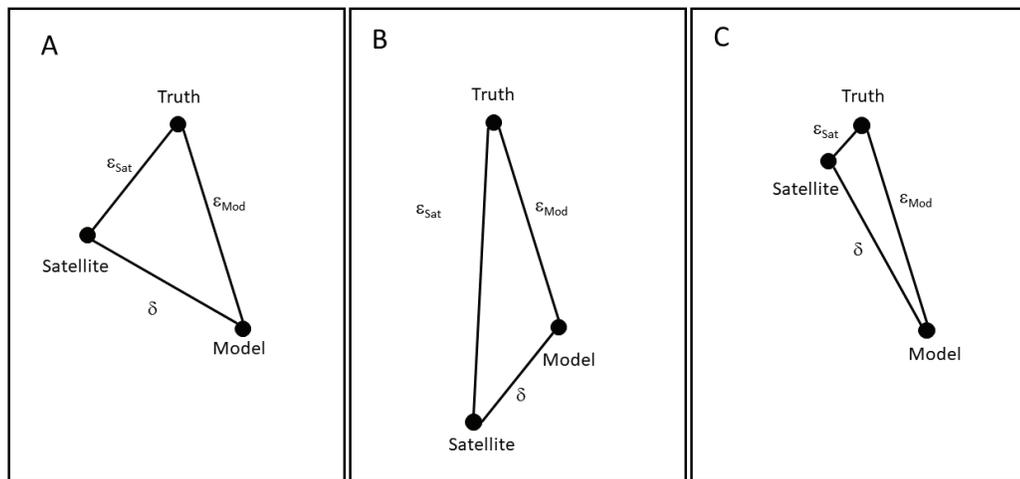
Figure 5.1 Truth, error and misfit in comparing ocean colour estimates and models (following Lynch et al. 2009). Both satellite-based measurements and simulations approximate of the true ocean conditions (e.g., chlorophyll), with errors characterized by ε_{Sat} and ε_{Mod}, respectively. When assessing model predictions, one is ultimately interested in ε_{Mod}, but the true state of the ocean cannot be known exactly. One must instead use the misfit (δ) between the model and satellite-based estimates. The panels illustrate the implications of this for three different levels of satellite accuracy. Panel A, shows a case where both model and satellite estimates have significant uncertainty, but the satellite is slightly more accurate. In this case, perhaps typical of chlorophyll in open ocean regions (Gregg and Casey 2004), δ and δ_{Mod} are similar: the measure of skill provided by the comparison with satellite is a reasonably robust measure of skill relative to the true ocean state, but should still be viewed with some caution. In Panel B, both the satellite estimate and the model are subject to large uncertainties. This case is comparable to skill assessments in the coastal ocean (Ciavatta et al. 2011), or comparing against primary production (Carr et al. 2006; Friedrichs et al. 2009) and the misfit δ relative to any single NPP algorithm may not be a good approximation of ε_{Mod}. Finally, panel C represents a case where the satellite estimate is an excellent measure of the ocean state such that there is only a small difference between δ and ε_{Mod}.

Chlorophyll estimation has long been a focal point of ocean colour measurements (Gordon et al. 1983). It has relevance to diverse potential model applications, including understanding seasonal variations and fish recruitment (Platt et al. 2003), predicting fisheries catch (Friedland et al. 2012; Mcowen et al. 2015), and understanding ocean habitats, including bycatch avoidance (Hazen et al. 2018). Chlorophyll is not, however, "fit to purpose" for building confidence across all model applications.

Section 5.4 thus discusses comparisons between models and satellite-based estimates beyond chlorophyll (e.g., primary production, the distribution of sizes and functional types in the phytoplankton community). Such properties are generally more "derived" (*sensu* Zheng and DiGiacomo 2017, Chapter 2) than chlorophyll. That is, the number and boldness of assumptions and approximations required to connect native satellite measurements to the estimated quantity are greater than those required to estimate chlorophyll. This often results in greater differences between satellite estimates and the true ocean state for the more derived quantity than for chlorophyll (Figure 5.1b). Alternatively, direct comparison between models and satellite-measured remote sensing reflectance (e.g., Baird et al. 2016; Dutkiewicz et al. 2019) reduces the distance between what a satellite actually measures and what the model is compared against (Figure 5.1c). Simulating remote sensing reflectance, however, requires biophysical models to include more complex (and potentially computationally expensive) optical and radiative transfer components than most biophysical models currently include. Also, while comparing directly with remote sensing reflectance may be natural for some studies (Baird et al. 2016), it may not be "fit to purpose" for other model applications.

Section 5.5 concludes with a forward-looking discussion of advances by the modelling and ocean colour communities that would further build confidence in marine ecosystem predictions and projections. On the modelling side, there has been tremendous progress since the literature review of Arhonditsis and Brett (2004) highlighted meager skill assessment efforts for marine ecosystem models. However, more work is needed to ensure that skill assessment is "fit to purpose". The utility of ecosystem predictions is often linked to capturing subtle anomalies around strong mean spatial and temporal patterns. More focused assessments on these decision-relevant scales, including formal retrospective prediction experiments (Séférian et al. 2014; Rousseaux and Gregg 2017; Park et al. 2019) are needed to further strengthen model skill assessment. Ocean colour is unique in its capacity to support such efforts, as it provides a global, multi-scale, decades long, time series of ecosystem state. The utility of ocean colour would be further strengthened by reducing the uncertainty in ocean colour based estimates of ocean conditions (i.e., striving for Figure 5.1c), continuing to develop new metrics for societally critical ecosystem phenomena, and rigorously quantifying uncertainties. Efforts on all these fronts are pivotal to realizing the aspiration of translating ecosystem predictions and projections into improved societal resilience.

5.2 Model Skill Assessment Metrics

Diverse metrics are used in model skill assessment. This chapter does not attempt a full review, and readers interested in a more complete treatment are directed to numerous texts and papers on the topic (Burnham and Anderson 2003; Stow et al. 2009; Wilks 2011). A few of the most common metrics are provided in Table 5.1. The reader will note that Table 5.1 does not use "model" and "observed" designations. This reflects the complexities expressed in Figure 5.1: we are comparing two estimates of the same ocean quantity, with one derived from a physical-biological model and the other derived from satellite-observed ocean colour. We use y to designate the satellite-based estimate, and $\hat{y}$ to indicate the estimates arising from a model. The differences between the model and satellite-based estimates ($\delta = y - \hat{y}$) are then referred to as "misfits" or "residuals". In comparisons against satellites, all quantities ($\delta, \hat{y}, y$) are generally gridded values in space and time, interpolated or averaged onto a common comparison grid.

The metrics in Table 5.1 measure different aspects of the model-satellite comparison. The correlation coefficient, for example, measures pattern consistency: it quantifies agreement (or disagreement) in the relative highs and lows on a normalized (-1 to 1) scale. It does not, however, contain information on the magnitude of misfits as does, for example, the root mean squared difference (RMSD). The RMSD can be further decomposed into a mean bias (B) and a contribution due to misfits remaining after the mean bias is subtracted (RMSD'). This decomposition is useful for understanding how much of the

Table 5.1 Common skill metrics used in comparisons of models with satellite data. In all cases, δ is the misfit between the model-based estimate ($\hat{y}$) and the satellite-based estimate (y), and n is the number of comparison points. The standard deviation of the model and satellite data are indicated as $\sigma_{\hat{y}}$ and σ_y, respectively.

Metric	Formula	Interpretation
Root Mean Squared Difference (RMSD)[*]	$RMSD = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\delta_i^2}$	The standard deviation (dispersion) of the model-satellite misfits
Bias (B)	$B = \dfrac{1}{n}\sum_{i=1}^{n}\delta_i$	The mean model-satellite misfit
Correlation coefficient (r)[**]	$r = \dfrac{\frac{1}{n}\sum_{i=1}^{n}((\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y}))}{\sigma_{\hat{y}}\sigma_y}$	Normalized measure consistency of relative spatiotemporal variations
Unbiased Root Mean Squared Deviation (RMSD')	$RMSD' = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(\delta_i - B)^2}$	RMSD after subtracting out mean bias
Normalized standard deviation	$\sigma^* = \sigma_{\hat{y}}/\sigma_y$	Ratio of variability in model to that of the comparison data
Nash-Sutcliffe model efficiency (NSE)[***]	$NSE = 1 - \dfrac{\sum_{i=1}^{n}\delta_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Measure of performance relative to a model capturing just the mean; fraction of the variance explained when > 0; perfect model has NSE = 1

[*] Also referred to as the root mean squared error (RMSE). We adopt "difference" herein for reasons explained in Figure 5.1

[**] This is the Pearson correlation coefficient. Another commonly encountered variant, Spearman's rank correlation, is discussed in Table 5.2

[***] See text for discussion of the relationship between NSE and the coefficient of determination (R^2)

misfit is due to a constant error across an ocean region, and how much may be due to smaller-scale spatial mismatches between simulated and satellite-estimated patterns.

A common model skill benchmark can be defined by comparing the misfit variance to those arising from a reference model that represents only the mean of the data ($\bar{y}$). This can be seen in the Nash-Sutcliffe Model Efficiency (NSE, Table 5.1). The NSE formula is identical to that for the coefficient of determination (R^2) commonly applied in statistical regression analysis. Values > 0 can be interpreted as the fraction of variance explained (just like R^2), and a perfect model has an NSE of 1. NSE values < 0 indicate that the model does worse than a model that represents only the mean of the data. There are many variants of such metrics, and many variants for reference models. Skill relative to a persistence forecast, for example, is a common metric used in forecasting (e.g., Stock et al. 2015; Jones et al. 2016).

The comparison metrics in Table 5.1 can be calculated with raw values or after logarithmic transformation. Many satellite properties, including chlorophyll, vary over several orders of magnitude (Campbell 1995; Van Oostende et al. 2018). Logarithmic transformation has the advantage of weighting misfits equally across orders of magnitude. That is, a misfit between a modelled value of 1 mg Chl m^{-3} and a satellite-based estimate of 0.1 mg Chl m^{-3} is given the same weight as a misfit between 10 mg Chl m^{-3} and 1 mg Chl m^{-3}. If values are not log-transformed, skill metrics are much more strongly influenced by values at the high end of chlorophyll's dynamic range. On a global scale, the desire to match chlorophyll patterns across their dynamic range often argues for log-transformation

(see examples in Section 5.3.3). A strongly skewed misfit distribution is a good indication that log-transformation will give a more balanced assessment of model performance across the range of the quantity being estimated.

Skill metrics generally have probabilistic underpinnings that can be harnessed for significance tests and model weighting. For example, the RMSD is broadly used because minimizing the RMSD is equivalent to maximizing the probability that a model explains an observed pattern, if the misfits are independent and normally distributed. That is, the model that minimizes the RMSD under these conditions is also the "maximum likelihood" model. This probabilistic interpretation, which can be expanded to non-normal distributions, enables use of a diverse hypothesis testing, model selection and inference tools (Burnham and Anderson 2003).

The metrics in Table 5.1 are more difficult to interpret if residuals remain strongly non-Gaussian even after log-transformation. They are highly sensitive to outliers, raising the possibility that a judgement of poor (or good) skill may actually rest on just a few points. Furthermore, many of the significance tests assume that residuals are approximately Gaussian. Other transformations can be applied to achieve a Gaussian distribution of residuals (Simon and Bertino 2009; Friedland et al. 2012), but an alternative (or complementary) approach is to use "robust skill assessment metrics" (Daszykowski et al. 2007). These have similar interpretations to the metrics in Table 5.1 but do not require assumptions about the distributions of the data. These metrics rank the values of the distributions and use the median and the inter-percentile ranges to characterize their central positions and dispersion, instead of the mean and standard deviations. Table 5.2 provides examples of robust metrics and their interpretation.

Table 5.2   Robust skill metrics used in comparing models with satellite data. In all cases, δ is the misfit between the model-based estimate ($\hat{y}$) and the satellite-based estimate (y), and n is the number of comparison points. The interquartile ranges of the model and satellite data are indicated as $IQR_{\hat{y}}$ and IQR_y, respectively. The ranked values of model and satellite data are indicated as $\hat{y}_r$ and y_r, and the standard deviation of the model and satellite data are indicated as $\sigma_{\hat{y}r}$ and σ_{yr}.

Metric	Formula	Interpretation
Median absolute deviation (MAD)	$MAD = median(|\delta|)$	The dispersion of the model-satellite misfits
Robust bias (B^*)	$B^* = median(\hat{y}) - median(y)$	The distance between model and satellite medians
Spearman's rank correlation coefficient (ρ)	$\rho = \dfrac{\frac{1}{n}\sum_{i=1}^{n}((\hat{y}_{r,i} - \bar{\hat{y}}_r)(y_{r,i} - \bar{y}_r))}{\sigma_{\hat{y}r}\sigma_{yr}}$	Normalized measure of monotonic consistency of relative spatiotemporal variations
Unbiased Median Absolute Deviation (MAD')	$MAD' = median(|\delta - B^*|)$	MAD after subtracting out the robust bias
Normalized interquartile range	$IQR^* = IQR_{\hat{y}}/IQR_y$	Ratio of variability in model to that of the comparison data

Summary diagrams combine multiple metrics from Table 5.1 into a single visual depiction of model skill. A commonly applied diagram in geoscience is the Taylor diagram (Taylor 2001), which is a polar coordinate plot combining correlation (angular position) and the normalized standard deviation (radial

position). Satellite ocean colour examples can be found in numerous published works (e.g., Gruber et al. 2006; Doney et al. 2009; Friedrichs et al. 2009; Moore et al. 2013; Aumont et al. 2015). Jolliff et al. (2009) noted that Taylor diagrams lacked information on model bias and proposed the "target diagram", which combines RMSD and bias information, to address this (e.g., Doney et al. 2009; Friedrichs et al. 2009; Saba et al. 2010, 2011; Lee et al. 2015a, 2016). Summary diagrams can be constructed with robust statistics (Laufkötter et al. 2015; Ciavatta et al. 2016). Section 5.3 provides an example application of these diagrams to compare the ability of Earth System Models contributed to the 5th Coupled Model Inter-comparison Project (CMIP5) to simulate annual mean chlorophyll observed via satellite.

A common probabilistic skill assessment alternative to the metrics in Tables 5.1 and 5.2 arises in the case where the model is designed to predict discrete outcomes e.g., can a model predict high biomass blooms? (Allen et al. 2007, 2008). Frameworks have been developed to assess the predictive power of binary (or otherwise discrete) classification systems based on the distributions of correct positives, correct negatives, false positives and false negatives (Brown and Davis 2006). A common means of summarizing the skill of such a system is a Receiver Operator Characteristic (ROC) curve, which summarizes the skill of correctly predicted blooms while avoiding false negative (e.g., Shutler et al. 2011). Such an approach is a good example of "fit to purpose" for testing the robustness of prediction systems for high risk events, such as its original application for differentiating friendly and enemy aircraft during the Second World War (Brown and Davis 2006). Further discussion of applications of this approach for ocean ecosystem models can be found in Allen et al. (2007) and Stow et al. (2009).

Lastly, it is notable that all but one of the skill assessment metrics in Table 5.1 require close spatial matches between modelled and satellite-based estimates for a model to rate highly (the one exception being the normalized standard deviation). This is a stringent (i.e., challenging) test, especially given the satellite-driven revelation that chlorophyll and other optical properties vary dynamically across a full range of spatiotemporal scales (Abbott and Zion 1985; Denman and Abbott 1988, 1994). Substantial ocean colour variation is associated with physical phenomena (e.g., eddies, fronts) that may be resolved by models, but may not occur in precisely the same location as those observed. This may be an issue if the goal of a simulation is spatially precise ocean state estimation and/or 1–3 day forecasts, but process-oriented models are more concerned with capturing phenomena faithfully and may be ill-served by metrics requiring a strict spatial match. Likewise, seasonal to multi-annual predictions are generally expected to provide information on larger-scale anomalies, not the precise position of mesoscale fronts or eddies.

The simplest approach to relaxing the requirement of strict spatial agreement for small-scale features is averaging prior to comparison. Such comparisons, however, fail to assess whether smaller-scale features arise in a manner consistent with those observed in even a statistical sense. There are, however, other approaches to address this by relaxing requirements for a strict spatial match, or focusing on statistical or spectral properties (Powell et al. 2006; Rose et al. 2009; Stow et al. 2009; Shutler et al. 2011; Saux Picart et al. 2012). These will be discussed further in Section 5.3.

5.3   Ocean Chlorophyll Comparisons Across Scales

Satellite ocean colour measurements provided the first truly global perspective on ocean ecosystems. Chlorophyll estimates revealed variations in the convergence of nutrients and light in the surface ocean that drive phytoplankton production (Figure 5.2a): elevated equatorial chlorophyll arising from upwelling currents induced by easterly winds sweeping toward the tropical convergence zone, highly stratified oligotrophic subtropical gyres with surface chlorophyll below 0.1 mg m^{-3}, chlorophyll rich temperate and subpolar regions fueled by the upward mixing of nutrient-rich deep waters each winter, and richly productive coastal ecosystems where dynamic circulation across complex bathymetry generates chlorophyll in excess of 10 mg m^{-3}. Climate variability alters these confluences of nutrients and light (Chavez et al. 2011; Taboada et al. 2019) and climate change is expected to do so (Bopp et al.

2013; Laufkötter et al. 2015). Thus, a foundational test of global and regional ecosystem models is whether they can capture mean chlorophyll distributions.

A: GlobColour Chl, mg Chl m^{-3}

B: ESM2M-COBALT, mg Chl m^{-3}

C: Ratio of Model to Satellite
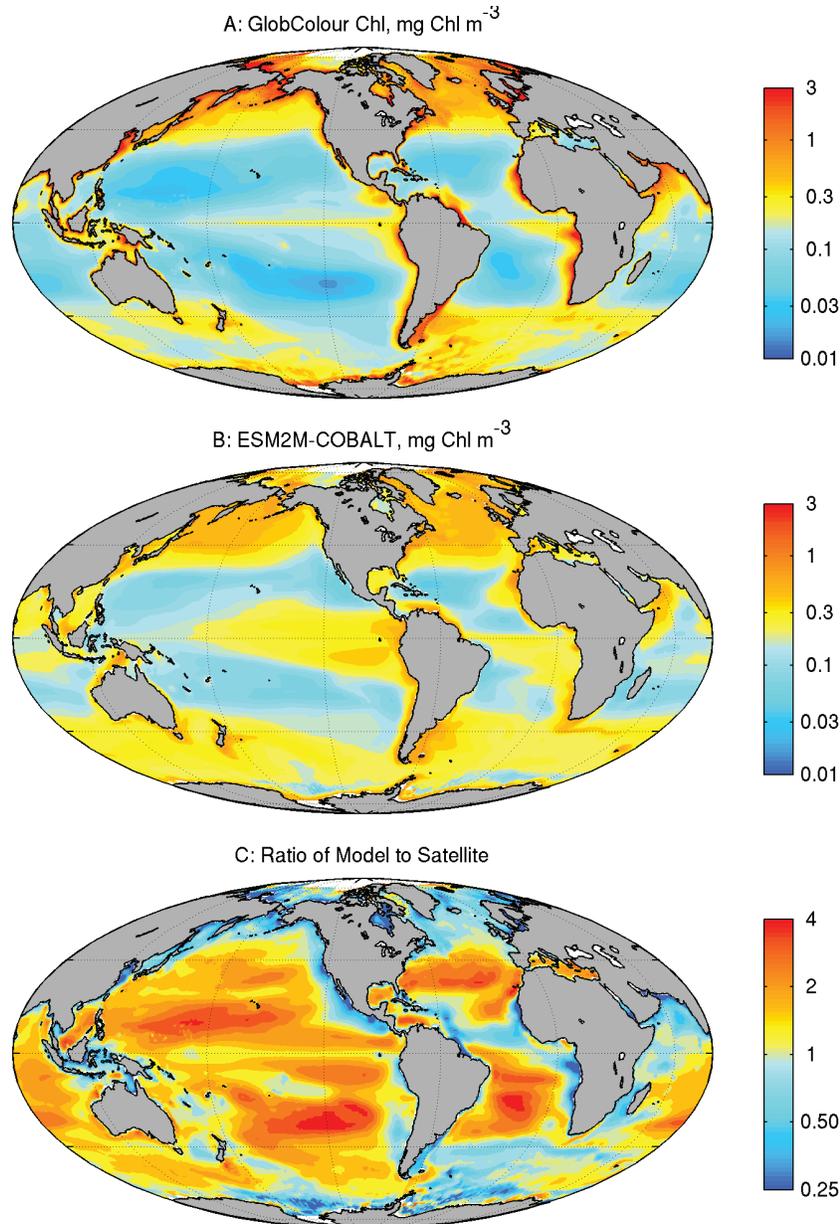
Figure 5.2 Panel A: The mean annual chlorophyll concentration estimated from a 20 year average of GlobColour data (mg Chl m^{-3}, 1998–2017). Means at high latitudes reflect non-winter months with satellite returns. Panel B: The annual chlorophyll concentration from GFDL's ESM2M-COBALT climate model (mg Chl m^{-3}). High latitude values only include data from months with satellite returns to allow for a consistent comparison with satellite data. Panel C: The misfit, which has been transformed from a difference between the log-transformed values to an equivalent ratio of the modelled to the satellite estimate (=10^{δ}) to ease comparison. The values of the skill metrics in Table 5.1 for this comparison are: RMSD = 0.34, B = 0.06, RMSD' = 0.34, r = 0.69, NSE = 0.48, and σ^{*} = 0.58.

An example comparison for the Geophysical Fluid Dynamics Laboratories global Earth System Model, ESM2M (Dunne et al. 2013) running with the Carbon, Ocean Biogeochemistry and Lower Trophics (COBALT, Stock et al. 2014a,b) is given in Figure 5.2. The model output is compared against the GlobColour chlorophyll estimates (http://www.globcolour.info/, ACRI-ST GlobColour, 2017). Model outputs were taken from 20 years of the late 20th century historical period of a climate projection. That is, the model's concentration of greenhouse gases, aerosols and radiation from the sun are consistent with recent years. All other aspects of the atmospheric, ocean, terrestrial physical, chemical and ecosystem responses are internally determined by the model (i.e., there is no data assimilation, "nudging", or incorporation of observed winds). Such loosely constrained runs are critical for understanding and projecting climate changes, but can lead to regional climate biases and drifts. The model's ocean resolution is coarse (~1°) to enable computationally expensive multi-century climate change projections (readers are directed to Stock et al. 2011 for further discussion of global climate and Earth system models).

The model output is compared against 20 year annual mean chlorophyll from GlobColour. Both the GlobColour satellite estimate used in Figure 5.2 and the model output have also been mapped onto a standardized 1° grid. As discussed in Section 5.2, this spatiotemporal averaging focuses the comparison on fidelity with large scale features (i.e., chlorophyll variations across ocean biomes). The simulated chlorophyll is furthermore averaged over the top grid cell of the model, which for ESM2M is 10 m. As noted in Chapter 4, this is a simplification often made in model comparisons with satellite data and ignores that the depth sampled by satellites varies with wavelength and optical constituents. While this does not compromise the cross-biome variations of interest in Figure 5.2, more advanced approaches are warranted in other circumstances (Gordon and Clark 1980; Moline and Prezelin 2000; Baird et al. 2016). Both the satellite and model data have been plotted on logarithmic scales to highlight similarities and contrasts across the dynamic range of ocean chlorophyll. We use the chlorophyll comparison in Figure 5.2 to provide an intuition for the skill assessment metrics in Table 5.1.

The correlation coefficient (r) between the log-transformed model and satellite-based surface chlorophyll estimates is 0.69, indicating generally consistent relative spatiotemporal chlorophyll variations. The difference between the model and the data, however, can be greater than a factor of 4 in some places (Figure 5.2c). The model under-estimates chlorophyll in many coastal areas and over-estimates chlorophyll in sub-tropical gyres (i.e., the "ocean deserts" indicated by the deep blue colours in the satellite data). The largest over-estimates often occur along the edges of the sub-tropical gyres, where the smaller size of the biomes and moderate displacements in biome boundaries can lead to very large misfits. This highlights the challenge posed by the strict spatial fidelity required by the skill metrics in Table 5.1. The misfit RMSD is 0.34, or roughly a factor of 2 variation. While this may seem large, it should be viewed relative to the orders of magnitude variation in chlorophyll dynamic range. The mean bias is modest (B=0.06) such that RMSD' is also 0.34 to within rounding error. The NSE is 0.48, indicating that the model captures roughly half of the variance in the data. The normalized standard deviation (σ^*= 0.58) confirms the clear visual impression that the model chlorophyll variation is damped relative to observations. Thus, the model has little bias, is robustly correlated with observed chlorophyll, captures about half the observed variance, but under-represents the dynamic chlorophyll range and can have large regional biases due to displaced biome boundaries and under-representation of very high chlorophyll coastal areas.

Comparison of satellite-based and simulated surface chlorophyll for ESM2M-COBALT and the suite of ESMs contributed to the 5th Coupled Model Intercomparison Project (CMIP5) provides an opportunity to highlight the utility of Taylor (Taylor 2001, Figure 5.3a) and target (Jolliff et al. 2009, Figure 5.3b) summary diagrams. A very broad range of model skill across CMIP5 is immediately apparent. This in part reflects wide variation in model comprehensiveness that arises from varying research foci and computational constraints across modelling centers. It should also be noted that models in Figure 5.3 are coarsely calibrated, to the limited extent that computational capacity allows, to satellite-observed chlorophyll patterns. For example, the highly uncertain iron scavenging rate in ESM2M-COBALT was

calibrated to get reasonable mean levels of iron, chlorophyll and nitrate in iron-limited High Nutrient, Low Chlorophyll regions (Stock et al. 2014b; Tagliabue et al. 2016). Since Figure 5.3 is intended solely as an example of a summary diagram, and we cannot fully discuss variations in model complexity and calibration explaining differences in skill, we refrain from identifying individual models aside from ESM2M-COBALT (blue dot).
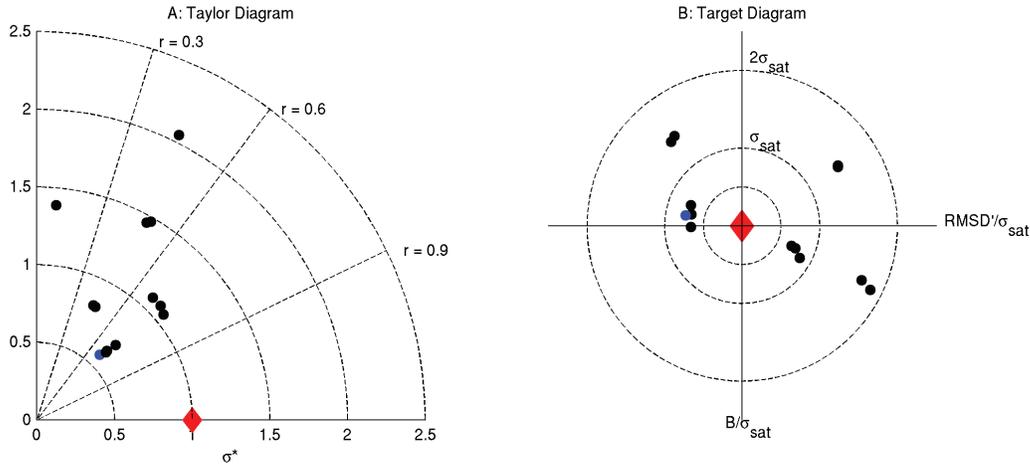
Figure 5.3 Summaries of the skill of ESMs contributed to IPCC-AR5 at capturing global mean chlorophyll (i.e., Figure5.2). Model output was downloaded from the NOAA Climate Change Web Portal (`https://www.esrl.noaa.gov/psd/ipcc/`, Scott et al. 2016). In the Taylor diagram (left), the polar coordinate is the correlation coefficient (r) and the radial coordinate in the normalized standard deviation (σ^*). The distance between each model and the perfect model ($r = \sigma^* = 1$; red diamond) is proportional to RMSD'. In the Target diagram (right), the x-coordinate is used for the unbiased RMSD and the y-coordinate is used for the bias. Since RMSD' is always positive, the sign of the x-axis is used to indicate whether the standard deviation of the model data is larger ($x > 0$) or smaller ($x < 0$) than the standard deviation of the satellite data.

In the Taylor diagram (Figure 5.3a), ESM2M-COBALT produces one of the better correlation coefficients (the polar coordinate) but has the lowest normalized standard deviation (radial coordinate). A perfect model would produce $r = \sigma^* = 1$ (the red diamond). The relatively short distance from ESM2M-COBALT to the red diamond indicates low unbiased RMSD relative to most models but, as Jolliff et al. (2009) point out, the Taylor diagram contains no information about the relative biases between models. The target diagram (Figure 5.3b) emphasizes the low bias of ESM2M-COBALT relative to other models. The target circles are drawn for $\frac{1}{2}$, 1 and 2 times the standard deviation of the satellite chlorophyll. None of the models fall within the inner target, but 7, including ESM2M-COBALT, fall within the second ring. The negative value on the x-axis indicates that the simulated standard deviation of the model data is less than that of the satellite data (i.e., $\sigma^* < 1$).

Fidelity with mean spatial chlorophyll gradients on global or regional scales provides a valuable baseline model skill assessment, but it is just a starting point. In the remainder of this section, we briefly highlight two challenges for ocean colour comparisons and one area for priority development where ocean colour has a prominent role to play. The first challenge is posed by uncertainties in satellite-based chlorophyll estimates, particularly persistent regional biases relative to *in situ* observations. The second challenge is posed by the difficulty of point-to-point comparisons in an ocean where chlorophyll often varies at very fine scales. The priority development focuses on the need for more focused comparison of models against ocean colour variations at the space and time scales of intended use, including formal retrospective forecast experiments (Séférian et al. 2014; Rousseaux and Gregg 2017; Park et al. 2019).

5.3.1 Regional biases in satellite chlorophyll measurements

As discussed in Chapters 2 and 4, satellite chlorophyll estimates and other satellite-based estimates must be derived from remote sensing reflectance at a limited number of wavelengths. This ultimately results in reliance on mechanistically-motivated empirical relationships between R_{rs} and chlorophyll concentration that have considerable overall fidelity with observed global patterns, but may be inaccurate regionally and/or temporally. Persistent regional or temporal biases could lead to false conclusions or mis-calibration of model dynamics.

A prominent biome-scale observation bias, for example, occurs in the Southern Ocean where global satellite-based algorithms consistently under-estimate *in situ* chlorophyll observations (Gregg and Casey 2004). Figure 5.4 compares mean SeaWiFS surface chlorophyll estimates (NASA 2014) in the Southern Ocean for December–February with the Southern-Ocean SeaWiFS algorithm of Johnson et al. (2013). The standard SeaWiFS algorithm is typically a factor of 2 lower than chlorophyll estimated from Johnson et al. (2013). Similar results were found for global MODIS and GlobColour algorithms (Johnson et al. 2013). Such discrepancies can significantly change the conclusions drawn from a comparison between simulated and satellite-estimated chlorophyll and create challenges when merging satellite products (GlobColour 2017). The chlorophyll comparison versus GlobColour in Figure 5.2, for example, gives the impression that ESM2M-COBALT overestimates chlorophyll over much of the Southern Ocean away from the coast of Antarctica, while Figure 5.4b suggests the opposite. Continued work toward globally robust satellite products, effective communication of regional limitations where they exist, and streamlined access to regionally refined products from the same portals providing global data would reduce risks of misinterpretation.
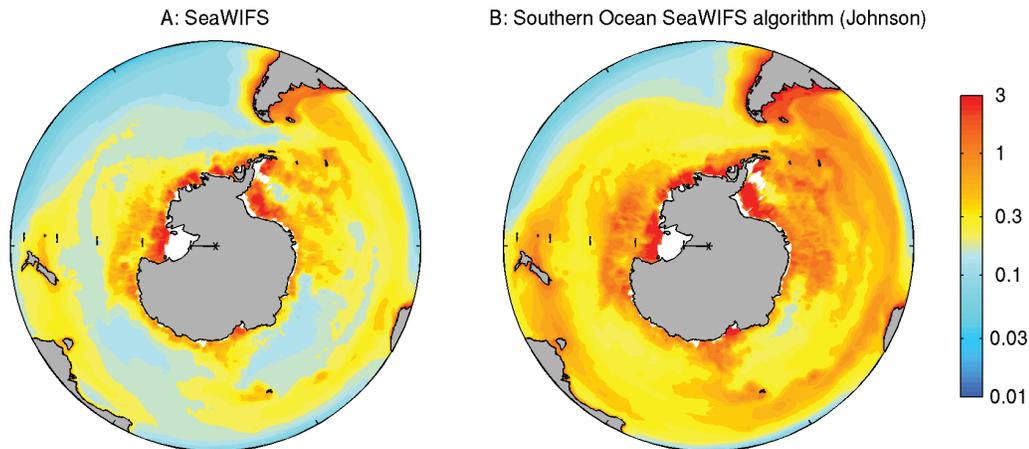
Figure 5.4 Comparison of chlorophyll estimated via the SeaWiFS algorithm (panel A, obtained from the Ocean Productivity Page, `science.oregonstate.edu/ocean.productivity/index.php`) and chlorophyll estimated by the Southern Ocean chlorophyll algorithm of Johnson et al. (2013) (panel B). Units are mg Chl m^{-3} in both panels.

Another uncertainty that challenges the use of satellite-based chlorophyll estimates in skill assessment are high uncertainties in optically-complex coastal waters (Schofield et al. 2004; Dierssen 2010), which can generate discrepancies similar in magnitude to the Southern Ocean example in Figure 5.4 (Gregg and Casey 2004). An example of the interpretative difficulty posed by optically-complex coastal waters is provided in Figure 5.5, which shows results from a decadal reanalysis of the biogeochemistry in the North West European shelf (Ciavatta et al. 2016), compared to the ocean colour product of the Climate Change Initiative of the European Space Agency (ESA CCI OC, `http://www.esa-oceancolour-cci.org/`; Sathyendranath et al. 2017). The largest deviations between model (Figure 5.5a) and satellite product

(Figure 5.5b) were found at the coast (higher RMSD and lower correlations in Figure 5.5c and Figure 5.5d). Here, however, both the model and the satellite derived estimates are likely to have the largest divergence from the truth (Figure 5.1b). Riverine input variability is poorly represented by climatology data in the model, while terrestrial dissolved organic matter and sediment resuspension increase the uncertainty in the ocean colour product. This makes a skill assessment inconclusive by means of ocean colour only, and supports augmenting comparisons against ocean colour-based measurements with *in situ* data.
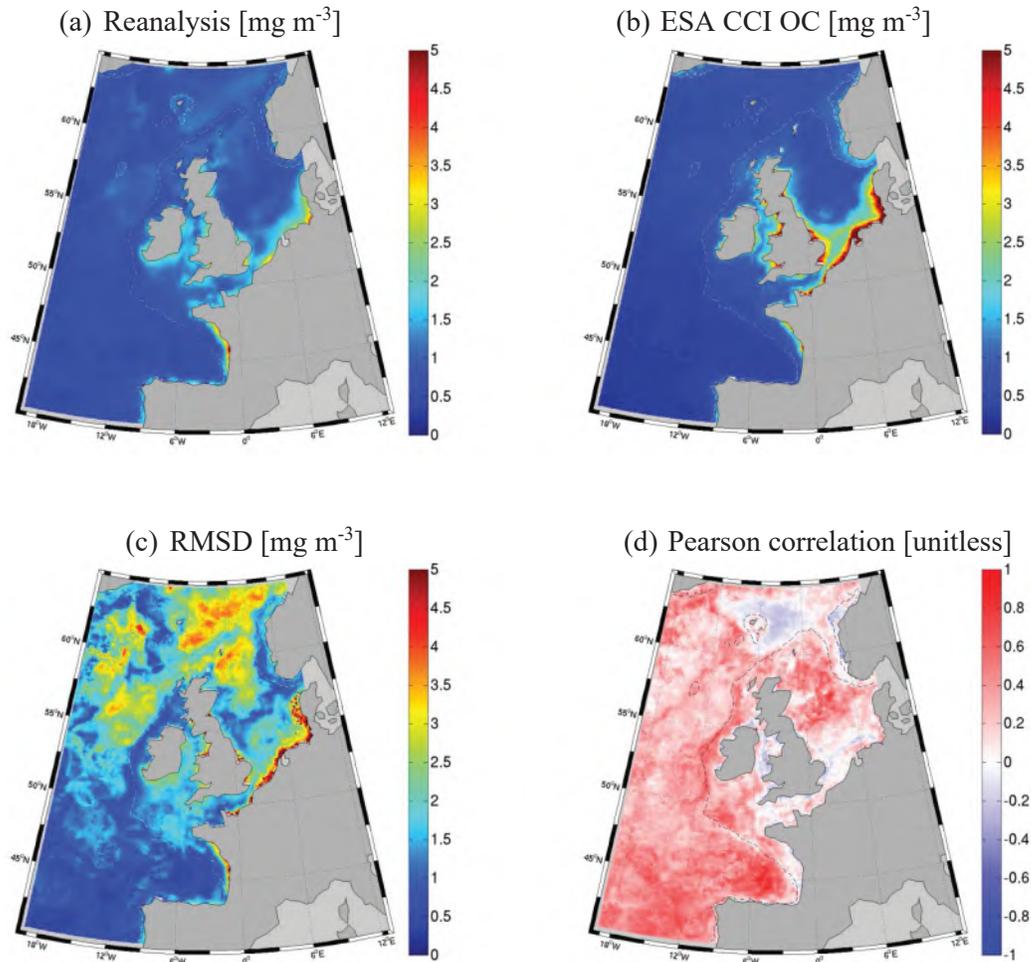
Figure 5.5 Skill assessment of a decadal reanalysis simulation of chlorophyll in the North East Atlantic in the years 1998–2009. (a) Average reanalysis chlorophyll concentrations; (b) Average ocean-colour-based chlorophyll concentrations (ESA's CCI product). (c) Root-mean-square-deviation, and (d) Pearson correlation defined in Table 5.1 were computed using time series of monthly data at each model grid point. Figure reproduced from Ciavatta et al. (2016), Creative Commons Attribution 4.0 International (CC BY 4.0).

The effect of coastal biases can also impact global-scale comparisons. Stock et al. (2017), for example, assessed the capacity of a high-resolution Earth system model (10-km ocean resolution) to capture mean chlorophyll patterns across globally distributed coastal "Large Marine Ecosystems" (LMEs, Sherman and Alexander 1986). Initial comparison including all ocean depths suggested that simulated chlorophyll is correlated with observed patterns, but under-estimates observed values (i.e., has a low

bias) of roughly a factor of 2. However, subsequent analysis revealed that much of this bias arose from gross under-estimates of chlorophyll in waters < 50 m (Table 5.3). The model under-predicts high chlorophyll in coastal areas, but the initially severe assessment should be tempered by its strong dependence on less reliable satellite-based chlorophyll estimates in optically-complex shallow waters.

Table 5.3 Comparison of simulated chlorophyll in GFDL's high-resolution ESM2.6 global Earth system model with SeaWiFS chlorophyll across coastal Large Marine Ecosystems (adopted from Stock et al. 2017). The arithmetic average across each LME was taken, and then $\log_{10}$-transformed before comparison. For the bias, we also report $F_{med} = 10^B$, which is interpretable as the characteristic ratio of the modelled to the observed value (Campbell et al. 2002). Note the pronounced effect of excluding waters that are < 25 m and < 50 m deep on model skill.

| Depths	R	Bias (F_{med})	RMSD*
All	0.74	-0.23 (0.59)	0.32
> 25m	0.84	-0.12 (0.76)	0.22
> 50m	0.86	-0.08 (0.83)	0.19

* Note that an RMSD of 0.3 implies a factor of 2 variability around the modelled value; while a value of 0.18 implies a factor of 1.5.

5.3.2   Challenges of point-to-point comparisons in a heterogenous ocean

As discussed in Section 5.2, many broadly applied skill metrics assess the match between models and satellite data "point-to-point". That is, they place high value on a direct spatial and/or temporal match between model and data. Snapshots of ocean colour data and the Northwest European Shelf pre-operational simulations of Skákala et al. (2018) illustrate the difficulties posed by this criteria (Figure 5.6). The ocean colour snapshot (Figure 5.6, "EO" = Earth Observations) reveals fronts throughout the region, where chlorophyll can vary by a factor of 5 over fine spatial scales. Note, for example, the thin lines of high chlorophyll extending eastward from the English coast, or the small elevated chlorophyll feature displaced offshore east of Scotland. The precise timing of the generation and subsequent movement of such features may be difficult for a simulation to capture and even slight displacement can lead to a large misfit. A coupled physical-biological simulation without data assimilation, for example, creates fine-scale features of qualitatively similar magnitude and scale (Figure 5.6, "reference"), but errors in the placement of these features, along with some larger-scale biases, lead to a substantial bias and RMSD relative to satellite-based chlorophyll estimates (Figure 5.6, bottom right, red dot). Assimilation of plankton functional types reduces both persistent biases and improves the spatial match to fine-scale features (Figure 5.6, bottom left), significantly reducing bias and RMSD at the model initialization and over 5 day prediction experiments (Figure 5.6, bottom right, blue stars).

More precise spatial matches with fine-scale features may be critical for short-term (1–5 day) forecasts of phenomena with acute local impacts (e.g., Skákala et al. 2018), but expecting a precise spatial match at fine scales for seasonal to century-scale predictions is less reasonable. The goal of such efforts is generally prediction of regional anomalies or trends, not the precise location of each eddy, meander or front. Furthermore, a model diagnosing the net effect and internal dynamics of fine-scale features need not capture their precise timing and locations at each point in time. Generating realistic fine-scale fronts at realistic frequencies and distributions may be enough.

The challenges of "point-to-point" comparison are not limited to fine-scales. Figure 5.2 highlights the impact of displaced fronts on skill metrics requiring a strict spatial match. The displacement of ocean biomes (e.g., sub-tropical gyres) creates large misfits, at times exceeding a factor of 4, along biome-boundaries. It is arguable that such large penalties for small displacements may be inappropriate
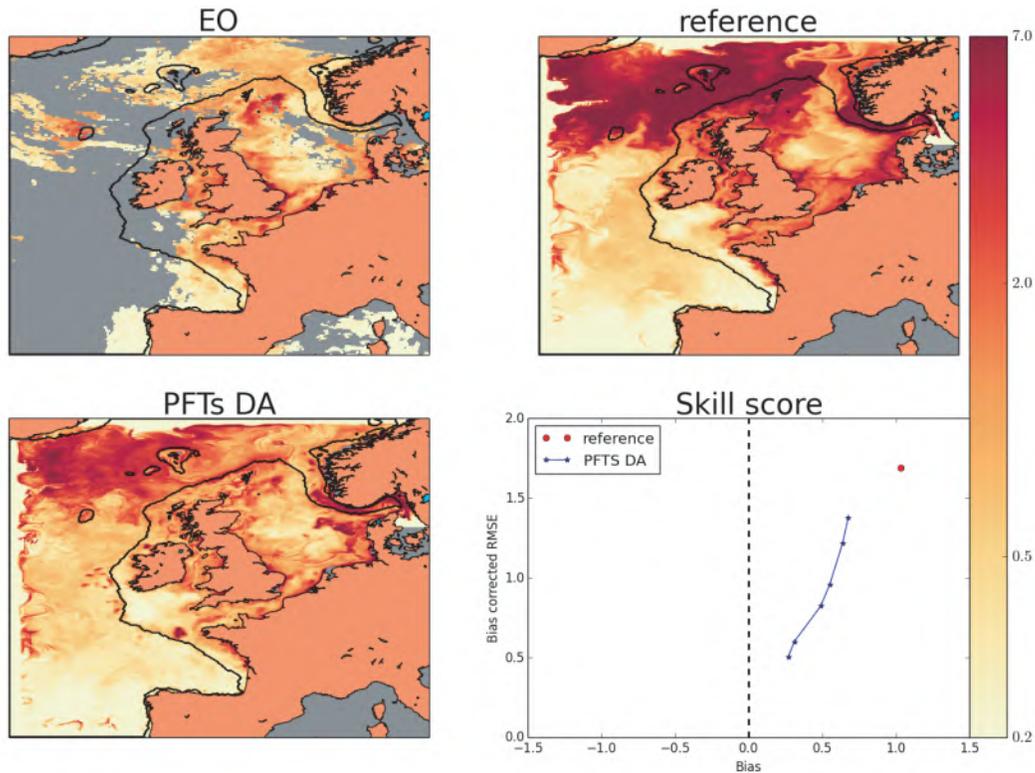
Figure 5.6 An example of chlorophyll comparison over fine scales redrawn from Skákala et al. (2018). Top left: a snapshot of satellite-estimated chlorophyll (mg Chl m^{-3}) from the Northwest European Shelf taken on May 28, 2010; EO = Earth Observation. Top right: Simulated chlorophyll during the same period from for a pre-operational model of the North West European shelf. Bottom left: The snapshot after assimilating ocean colour products into the model, improving the fit to mesoscale features. Bottom right: Skill metrics for the reference simulation (red dot), the simulation after data assimilation (star with the lowest RMSE' and bias), and for a 5 day forecast (5 subsequent stars with increasing RMSE' and bias). Note that the skill stays above that of the reference simulation throughout the forecast. The simulations were obtained from the pre-operational model NEMO-ERSEM of the North West European shelf and the PFT products were derived from the ocean colour product of the ESA CCI programme, within the European Copernicus Marine Environment Monitoring Service project TOSCA (Skákala et al. 2018), reproduced with permission from John Wiley and Sons.

for a model attempting to track changes in the scale of ocean-biomes across centuries (e.g., Polovina et al. 2011).

As discussed briefly in Section 5.2, a variety of approaches have been applied to relax the requirement of exact spatial mismatches, the simplest being spatial and temporal averaging onto relatively coarse grids prior to comparison (e.g., Figure 5.2). Methods have also been devised to allow features to "slip or slide" in limited ways to avoid over-penalizing small spatial misfits across sharp frontal features (Rose et al. 2009). Yet another approach involves isolating the features of interest in both the model and data and comparing feature properties. Empirical orthogonal functions and eddy detection algorithms, for example, are often used to isolate ocean colour signals associated with large-scale modes of variation (e.g., see Taboada et al. 2019 for a recent example) and eddies (e.g., Chelton et al. 2011), respectively. Such features can be compared as composites across all occurrences (e.g., Chelton et al. 2011; Turi et al. 2018), or skill can be broken down by scale (Shutler et al. 2011; Saux Picart et al. 2012). Yet

another approach to filtering out the impact of small-scale spatial ocean colour mismatches is eschewing spatial comparison altogether in favor of the statistical distributions of ocean properties over broad regions. This can be done with histograms or "violin plots" of modelled versus satellite-observed points (Van Oostende et al. 2018, Figure 5.7), or through ecologically interpretable distributional parameters (Cael et al. 2018).
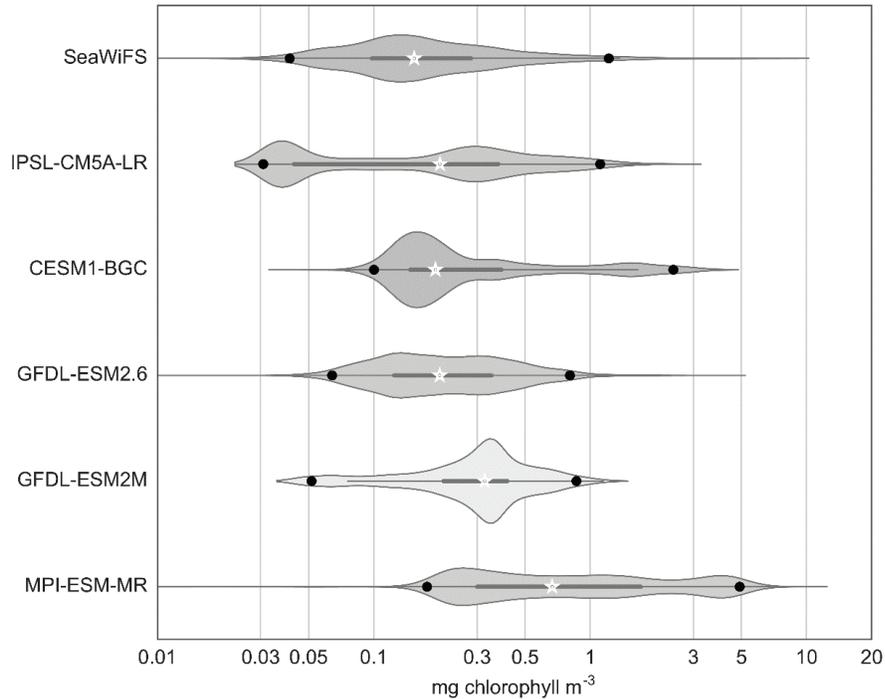
Figure 5.7   Violin plots comparing the probability distributions of annual mean chlorophyll from various global Earth System Models (ESMs) against chlorophyll estimated from SeaWiFS (figure redrawn from Van Oostende et al. 2018). The width of each violin is proportional to the number of points in each bin. Note that spatial agreement is not assessed in this comparison, only the distribution of chlorophyll values across the range of global values.

5.3.3   Making the space and time scales of skill assessments "fit to purpose"

The skill assessment examples in the preceding sections focused on assessing models against a variety of spatial, temporal and statistical ocean colour patterns. Successful comparisons strictly support model-based analysis of the drivers underlying the patterns to which the model was compared. The objectives of modelling efforts, however, often extend beyond dominant spatial or temporal patterns. Relatively subtle seasonal anomalies or trends, for example, may be far more important than comparing mean states across biomes or seasons. Furthermore, successful simulation of past ocean colour variations does not imply a capacity to predict future changes. Thus, while comparison against dominant spatiotemporal patterns is a good starting point, it is critical that skill assessments match the intended applications of a model as closely as possible. That is, skill assessment metrics should be "fit to purpose".

As an example, the biome-scale comparison provided in Figure 5.2 supports further analysis of the plankton dynamics underlying biome-scale ocean colour variation (e.g., Stock et al. 2014b). It does

not, however, demonstrate the model's capacity to simulate seasonal to inter-annual variations within biomes. Tight linkages between seasonal and inter-annual ecosystem anomalies and marine resources (Lehodey et al. 2006) has led to closer scrutiny of seasonal chlorophyll anomalies (Doney et al. 2009; Park et al. 2018). The more subtle anomaly patterns present a higher bar than stark spatial gradients or seasonal changes (Figure 5.8, left panel), but are ultimately essential before diagnosing drivers of seasonal to inter-annual ocean colour anomalies. Additional examples of assessment of the capacity of ecosystem models to capture relatively subtle inter-annual anomalies and fluctuations can be found in analyses of bloom phenology (Henson et al. 2009).

Building confidence in seasonal to inter-annual chlorophyll predictions presents a further challenge. Successful simulation of past ocean colour variability does not imply successful forecasts of future ocean colour variations. The latter relies not only on a skillful ecosystem model, but also a robust estimate of current ocean conditions and skillful atmosphere and ocean forecasts. Assessing the skill of ecological forecasts requires computationally-intensive retrospective forecast experiments. Our capacity to predict seasonal atmospheric and oceanic temperature anomalies, for example, has been assessed with over 30 years of retrospective forecast experiments (Becker et al. 2014; Kirtman et al. 2014; Stock et al. 2015): dozens of forecasts are initiated each month over the past 30 years and assessed for their capacity to reproduce observed patterns. Similar retrospective forecast experiments are now being conducted for ecosystem models using the ocean colour time series to assess forecast skill. This is being done at global (Séférian et al. 2014; Rousseaux and Gregg 2017; Park et al. 2019) and regional scales (e.g., Skákala et al. 2018).
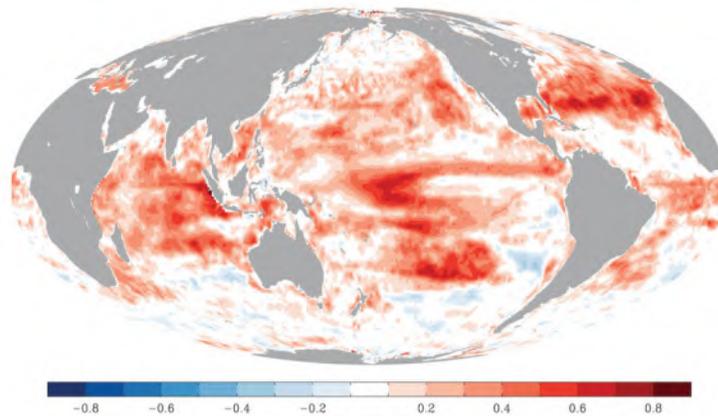
Figure 5.8 Anomaly correlation coefficients between ESM2M-COBALT simulated chlorophyll and GlobColour from a simulation integrating COBALT dynamics with physical data assimilation (Park et al. 2018). Anomaly correlation coefficients are based on data from 1997–2017. The monthly chlorophyll climatology was removed before calculating anomalies, thus focusing the comparison on monthly anomalies rather than the monthly mean pattern. Reproduced from Park et al. (2018), Creative Commons Attribution Non-Commercial No Derivatives License (CC BY-NC-ND).

Climate change presents a particularly challenging time-scale for building confidence. There is still substantial uncertainty surrounding the impact of climate change on ocean productivity (Bopp et al. 2013; Laufkötter et al. 2015). Climate change signals on ocean ecosystems are generally relatively subtle trends underlying pronounced variations that can take decades to manifest (Rodgers et al. 2015; Henson et al. 2017). In most cases, satellite ocean colour time-series are not yet long enough to reliably isolate climate change signals from those associated with variability (Henson et al. 2010). Confidence in climate projections must thus lean on the mechanistic underpinnings of the model and its capacity to represent past fluctuations, often on shorter time scales (Randall et al. 2007; Cheung et al. 2016).

Recent work has linked modelled ocean productivity responses to El Niño with the strength of a model's climate change response (Kwiatkowski et al. 2017). This relationship, together with satellite-based estimates of past El Niño responses, has improved constraints on ocean productivity change. A capacity to compare directly with climate change trends, however, would further build confidence. Avoidance of gaps in satellite coverage is crucial for this goal (Beaulieu et al. 2013).

5.4 Beyond Chlorophyll: Assessing Skill Against Other Satellite-Derived Ecosystem Properties

Comparison with satellite-based chlorophyll estimates account for the majority of uses of ocean colour data for model skill assessment. However, as useful as chlorophyll may be as an indicator of ecosystem state, the end goal of most modelling efforts is not solely simulation of chlorophyll. For example, a model's intended purpose may be simulating changes in sediment, primary production, phytoplankton community composition, the export of carbon between the surface and depth, or hypoxia. Skill assessment of robust satellite-based estimates of such quantities would be more "fit to purpose" than chlorophyll alone.

The challenge with satellite-based estimates of such quantities is that they often require additional assumptions to get from remote sensing reflectance to the quantities of interest (Zheng and DiGiacomo 2017, Chapters 2 and 4). This, combined with the relative scarcity of direct measurements of these properties, leads to an uncertainty between satellite values and the true ocean that is comparable to, or larger than that of coupled physical biological models (i.e., Figure 5.1c). Many such comparisons are effectively "model-to-model" (Doney et al. 2009).

Figure 5.9 provides one example of the large spread between more "derived" satellite-based estimates. The top row gives primary production estimated from the Vertically Generalized Production Model (VGPM, Behrenfeld and Falkowski 1997a). This formulation used a limited set of *in situ* primary production estimates to posit that there is a peak in maximum chlorophyll-specific carbon fixation near 20°, which is reflected in productivity maxima at mid-latitudes in Figure 5.9a. Figure 5.9b is identical, except that the original temperature scaling of maximum carbon fixation is replaced with a monotonic increase inspired by, though not equivalent to, observed relationships between the maximum phytoplankton growth rate and temperature (Eppley 1972; Bissinger et al. 2008). The productivity maximum is much more skewed toward equatorial regions as a result. Other NPP algorithms tend to cluster around these two highly distinct patterns (Carr et al. 2006), yet *in situ* NPP observations still have difficulty determining which is more accurate (Friedrichs et al. 2009; Saba et al. 2011). This partly reflects difficulties measuring NPP (e.g., Barber and Hilting 2002) and partly uncertainties associated with the relationship between satellite measurements and NPP. While concerted efforts to resolve these uncertainties continue (Buitenhuis et al. 2013; Silsbe et al. 2016), they have yet to result in a community consensus on ocean productivity patterns.

Similarly significant spreads in estimates apply to other highly derived quantities, including phytoplankton sizes and functional types (Brewin et al. 2017b; Mouw et al. 2017). This is not meant to imply that comparing models against more "derived" satellite estimates is not useful. Chlorophyll, and indeed any observation (Figure 5.1), is subject to uncertainty. However, modellers should carefully consider the range of uncertainty when assessing models. Table 5.4 provides some useful references for characterizing uncertainty for different satellite-based ocean property estimates. Efforts to further characterize and, where possible, reduce these uncertainties would be highly beneficial to ocean colour applications in model skill assessment (see Section 5.5). In some cases, diagnoses of the simulated relationships between surface observed and derived properties in models may offer new perspective for the design and improvement of ocean colour algorithms (Stock 2019).
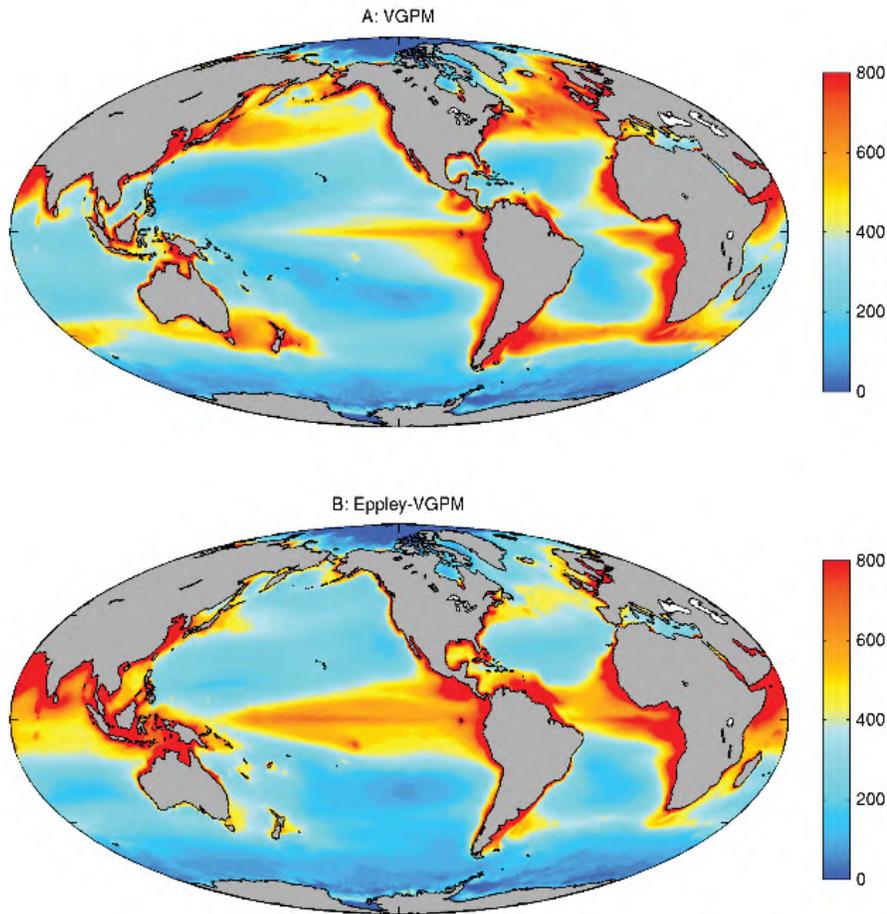
Figure 5.9 Primary production (mg C m^{-2} day^{-1}) simulated by the VGPM (panel A) and Eppley-VGPM primary production algorithms (panel B).

Table 5.4 Some critical references for understanding and quantifying uncertainties in satellite-based ocean property estimates.

Property	References
Chlorophyll	Gregg and Casey (2004); Moore et al. (2009); Jackson et al. (2017)
Carbon	Evers-King et al. (2017)
Primary Production	Campbell et al. (2002); Carr et al. (2006); Friedrichs et al. (2009); Saba et al. (2010, 2011); Buitenhuis et al. (2013); Lee et al. (2015a, 2016); Silsbe et al. (2016)
Phytoplankton size/functional type	Mouw et al. (2017); Brewin et al. (2017a)

5.5 Conclusions

The primary obstacle to assessing the skill of ecosystem models is the availability of observations. The spatially continuous, global, multi-decadal picture of ocean ecosystems provided by satellite-based

ocean colour measurements has thus played a crucial role in building confidence in marine ecosystem models and the predictions and projections they provide. This holds for models designed to project multi-decadal changes across ocean basins, those simulating inter-annual fluctuations in seasonal transitions and anomalies, and those simulating fine-scale dynamics along mesoscale fronts and eddies. The continued extension of the ocean colour time series, improvements to existing products, and development of new products have furthered this role. There are, however, challenges that still must be confronted by the modelling and ocean colour communities to continue on this positive trajectory.

For the modelling community, work by Arhonditsis and Brett (2004) highlighted the need for increased skill assessment in ecosystem modelling studies, noting that only 30% of 153 modelling studies reviewed reported a measure of goodness-of-fit. Since then, the model analysis community has embraced expanded model skill assessment against a range of ocean observations with improved approaches (e.g., Lynch et al., 2009). Calls for ecosystem predictions to inform decisions (e.g., Hobday et al. 2016; Tommasi et al. 2017a), however, place added pressure on skill assessment to provide an accurate and quantitative answer to the question: "how much should we trust the model prediction?". This requires skill assessment metrics to be closely fit to the space and time scales of interest. In cases of ecological forecasting, retrospective forecast experiments that accurately reflect the conditions under which current forecasts are being made (e.g., Séférian et al. 2014; Rousseaux and Gregg 2017; Skákala et al. 2018; Park et al. 2019), are essential.

The heightened need for rigorous skill assessment, including retrospective forecast experiments, further enhances the importance of ocean colour. The unique combination of spatial and temporal coverage of ecosystem responses that ocean colour offers enables applications that match the space and time scales of the evaluation with those of the intended prediction. The multi-decadal duration of the ocean colour time series supports retrospective forecast experiments on seasonal to inter-annual time scales. These become more robust with each year of new data. The continued lengthening of the satellite time series will also enable more direct testing of climate change trends (Beaulieu et al. 2013; Sathyendranath et al. 2017). Advances in a number of areas, however, would be highly beneficial. First, reduction in uncertainty, particularly regional biases, would greatly assist skill assessment by making model-satellite misfits more robust metrics of the difference between models and the true ocean state (i.e., Lynch et al. 2009, Figure 5.1). Second, improved characterizations of uncertainties (Table 5.4) would ensure proper interpretation of misfits, in addition to enabling formal integration of models and satellite data in assimilative frameworks.

<div style="border:1px solid #999; display:inline-block; padding:8px 40px 8px 12px; background:#dfe6f0;">

Chapter 6

</div>

Assimilation of Ocean Colour

Mark Baird, Emlyn Jones, Stefano Ciavatta, Cecile Rousseaux, Marjorie A. M. Friedrichs, Daniel E. Kaufman, Igor Shulman, Sergey Frolov and Christopher A. Edwards

---

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

The challenge of quantifying the spatially- and temporally-resolved biological state of the ocean has often split into two distinct endeavours: analysis of observations (including ocean colour remote sensing) and biogeochemical modelling. This split produces two results, sometimes difficult to reconcile, while failing to take advantage of the relative strengths of the two approaches. The assimilation of ocean colour into biogeochemical (BGC) models, like the assimilation of remotely-sensed physical variables into hydrodynamic models, provides a rigorous method to include both observations and models into one unified output that incorporates the information from both sources. As such, ocean colour assimilation is highly-sought after both by the scientific (Ciavatta et al. 2016) and environmental management (Jones et al. 2016) communities.

The earliest example of a formal assimilation of data into a BGC model used *in situ* nutrient and plankton observations to optimize parameter values in 3-, 4- and 7- component configurations of a NPZD model (Matear 1995). A history of BGC assimilation is available in Matear and Jones (2011). This Chapter will concentrate on examples of BGC data assimilation using ocean colour products. The earliest example of the use of ocean colour for data assimilation involved "nudging" a BGC model to the observed value from the Coastal Zone Color Scanner (CZCS, Armstrong et al. 1995). More recent studies investigating the benefits of assimilating ocean colour products, predominantly SeaWiFS-derived Chl-a, into BGC models include those of Carmillet et al. (2001) and Natvik and Evensen (2003), with a comprehensive review of algorithms used and observations assimilated detailed in Gregg (2008) and Edwards et al. (2015). There are now examples of operational and pre-operational global systems that routinely assimilate Chl-a products (Ford et al. 2018; Fennel et al. 2019). Additionally, there has been further experimentation with assimilating alternative remotely-sensed products such as phytoplankton functional types (Ciavatta et al. 2018, 2019), apparent optical properties (AOPs) such as the vertical attenuation coefficient at 443 nm, $K_d(443)$ (Ciavatta et al. 2014) and inherent optical properties (IOPs) such as phytoplankton absorption (a_{ph}), as described in Shulman et al. (2013).

This Chapter describes the basic concepts used in the assimilation of ocean colour into BGC models, and provides recent case studies that illustrate a number of recent successful applications.

6.1 Basics of Assimilation/Types of Assimilation Models

Numerical ocean models predict the evolution of a set of prognostic variables in time and space determined by a set of governing equations. In the case of ocean general circulation models (OGCMs), the prognostic variables are typically sea-surface elevation, temperature, salinity and currents. In the

case of biogeochemical models, the prognostic variables vary, but typically include phytoplankton biomass, nutrients, as well as organic and inorganic particles. Data assimilation (DA) is the process of using observations to reduce the error between observations and the model's prognostic variables (see, e.g., Moore et al. 2019. To align the language used in this chapter with that used in the DA literature, we refer to the model "state" vector as a vector of model prognostic and diagnostic variables that are mapped from the full 2D/3D reference space into a single column vector.

There are a variety of DA techniques that have been successfully applied to BGC models. In the broadest sense, these can be considered as state estimation, parameter estimation and joint (state and parameter) estimation. State estimation DA algorithms can then be further categorized as either sequential approaches (Figure 6.1, dark blue line) or variational approaches (Figure 6.1, brown line), each with their own distinct advantages and disadvantages. Gregg (2008) presents a broad review of the techniques applied until 2008, and Edwards et al. (2015) discuss advances since 2008, while Friedrichs et al. (2007) examined how well observations are able to constrain differing complexities of models. A key theme that has emerged in BGC data assimilation literature is an acknowledgment that the structure and equations of BGC models are uncertain, and the quality, sparsity and relationship between BGC observations and the BGC model state variables is challenging (Parslow et al. 2013; Dowd et al. 2014; Jones et al. 2016; Ciavatta et al. 2016; Ford and Barciela 2017).
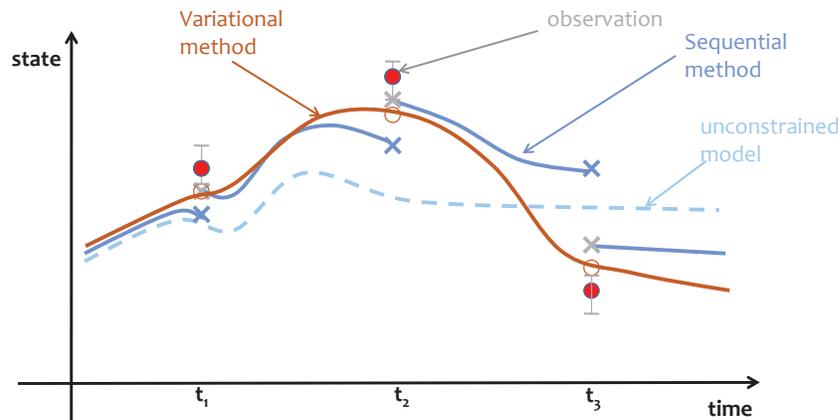
Figure 6.1 Schematic of data assimilation techniques. Red dots represent observational data (e.g., satellite derived Chl) with an estimate of uncertainty. Blue dashed line is an unconstrained model (i.e., a model without data assimilation). Dark blue lines represent a data assimilated model that uses a sequential method to nudge the model at specific times toward the observations. Brown line indicates model using the variational method that has parameters/initial conditions adjusted to best match the observations over the full course of the time period.

6.1.1 Variational methods

Derived from optimal control theory, variational data assimilation methods seek a solution that minimizes a cost function, typically consisting of the weighted sum of squared model data differences. Non-linear models (such as BGC models) generally result in cost functions with multiple, local minima. Linearization of the dynamics yield a quadratic cost function with a single minimum that can be found iteratively in complex, multi-dimensional problems using tangent linear and adjoint models. The tangent linear model steps forward in time, a (small) perturbation to the BGC state using dynamics linearized about the prior solution (obtained from the nonlinear BGC model). The adjoint model steps the linearized model backward in time to obtain the gradient of the cost function with respect to the

control vector. Arguably, the greatest challenge for variational methods, and one that has likely reduced its broader application, is the necessary construction and maintenance of the adjoint and tangent linear models. Automated tools exist to build these complex codes (e.g., Giering and Kaminski 1998), but laborious hand construction is also accomplished (Song et al. 2016a). Recently, Mattern and Edwards (2017) have developed a method appropriate for BGC models to approximate the tangent linear and adjoint models using finite differences, and this method may be sufficiently efficient to apply to a wide range of BGC models of intermediate complexity.

The variational approach has been applied to parameter estimation applications in one (vertical) dimension (e.g., Friedrichs et al. 2006; Ward et al. 2010). In these cases, the control vector consists of the set of model parameters, generally numbering 10 or more parameters. Variational methods applied to 1-dimensional problems have shown limited success in constraining parameters of simple BGC models. Parameter estimation using non-variational methods (Monte Carlo and genetic search algorithms applied to the full nonlinear model) in realistic, 3-dimensional configurations have shown promise (Mattern and Edwards 2017), but application of the variational approach to this 3-dimensional problem has not yet been demonstrated.

Rather, variational methods in realistic OGCM configurations have focused on BGC state estimation in which the control vector consists of the BGC ocean state (i.e., all prognostically-modelled BGC elements). This work has built on the extensive penetration and maturity of these approaches in physical oceanography. Two methods are common. In 3-dimensional variational (3D-Var) data assimilation, the ocean is considered static during the interval that a set of observations is made. In contrast, methods that exploit ocean dynamics to connect observations collected through time during a finite assimilation cycle are referred to as 4-dimensional variational (4D-Var) methods. Variational BGC state estimation methods have been developed using 4D-Var.

In 4D-Var methods, results depend sensitively on choices made for the weighting coefficients, embodied in the observation and background error covariance matrices. Although these matrices are poorly known, Desroziers et al. (2005) offered diagnostic consistency tests for portions of these error co-variances, yielding objective methods to constrain their values. These methods have been applied in the California Current System to improve 4D-Var BGC estimates by Mattern et al. (2018).

Two challenges specific to the BGC assimilation problem are that variables are positive definite, and the error statistics of BGC variables are non-Gaussian. Campbell (1995) demonstrated that surface chlorophyll concentrations in the ocean are reasonably described by log-normal statistics. Fletcher and Zupanski (2006a,b) and Fletcher (2010) developed a theoretical approach to 4D-Var for log-normally distributed variables for applications in meteorology. Song et al. (2016a,b,c) and Mattern et al. (2017) applied this method to a realistic regional configuration in the California Current System. There have been a variety of approaches trialed that span the simple tuning of parameters through to the full use of an adjoint and tangent linear model in 4D-Var (Mattern et al. 2017).

6.1.2   Sequential methods

The most common sequential methods used in BGC data assimilation are the Ensemble Kalman Filter (EnKF), and other related Kalman Filter variants (e.g., Ciavatta et al. 2011; Jones et al. 2016). Sequential methods rely on the approximation of the probability distribution of state of the prognostic variables at time t, $\mathbf{X}_t$, based on observations of state up until time t, $\mathbf{Y}_t$, $p(\mathbf{X}_t|\mathbf{Y}_{1:t})$. The probability is calculated via a discrete approximation generated using multi-simulation (ensemble) methods. The most common method to estimate $p(\mathbf{X}_t|\mathbf{Y}_{1:t})$ are the various flavors of the Kalman Filter. For large-scale non-linear BGC models, EnKF is the most common sequential assimilation method. Additionally, there are a growing number of applications using non-parametric methods including the Particle Filter (PF) (Jones et al. 2010; Mattern et al. 2012; Parslow et al. 2013) and emulator approaches (Mattern et al. 2012; Margvelashvili et al. 2013).

Sequential methods employing full dynamic ensemble techniques are easier to implement than

adjoint methods because they do not require the substantial investment in developing adjoint and tangent linear models. Furthermore, because the sequential method algorithms operate on the model output of existing non-assimilating codes, generic data assimilation software has been developed that can be applied to models' numerous research fields (Sakov 2018). The downside of sequential methods relates to their computation burden, often requiring a large number of ensemble members (10s – 100s) to sample the probability distribution of the state, $p(\mathbf{X}_t|\mathbf{Y}_{1:t})$.

6.1.3   Common requirements of observational data sets

Regardless of the data assimilation method chosen, there are shared requirements common to all data assimilation systems. Most state variables in biogeochemical models are measured in an indirect way, which is especially true for observations derived from ocean colour remote sensing. Therefore, the observation operator that quantifies the relationship between the observation and model state variable can be quite complex, with error and uncertainty arising from many sources. Jones et al. (2016) identified three sources of error when relating observations to the modelled state variables:

1. **Analytical** — errors arising from precision error in the measurement technique;

2. **Difference-in-kind** — errors arising from the conversion of an observed quantity to and from a modelled quantity (e.g., the OC3M algorithm that relates remote-sensing reflectance to HPLC derived Chl-a concentration);

3. **Spatial representation** — errors caused from models representing an average quantity over a model grid cell being compared to a measurement with a different spatial scale in a complex spatial structure. Discussed in detail in Oke and Sakov (2008).

With respect to ocean colour observations, there is an additional error that stems from scene wide atmospheric correction that may cause large spatially-coherent errors in the form of bias. Additionally, cloud clearing algorithms can, at times, allow small regions within a scene to have a large locally-correlated error. The handling of difference-in-kind error, atmospheric correction problems, and cloud clearing all have the potential to play havoc with both sequential and variational data assimilation methods. Bad-batch and outlier detection can filter some of these errors out, but if the correlation/covariance structure of the errors can be specified, then the assimilation system can be configured such that the analysis fields are less sensitive to such errors.

6.1.4   Parameter estimation

Estimating optimal values for parameters that are difficult to measure is greatly aided by data assimilation procedures that objectively search for values resulting in an optimal match between simulation output and observations. Data derived from satellite ocean colour measurements are well suited for these procedures (Hofmann and Friedrichs 2001; Friedrichs and Kaufman 2019), as they provide surface concentration estimates of chlorophyll and other properties with high spatial and temporal coverage. Assimilation of ocean colour satellite data was a goal for coupled physical-biogeochemical models for nearly a decade (Abbott, 1992) before the first ocean colour data assimilation experiments (Friedrichs 2002). Since then, parameter optimization efforts have expanded to include a wide variety of methodologies. Below we describe examples in the literature, including the early assimilation experiments of SeaWiFS data, as well as more recent developments in utilizing a variety of novel satellite products and assimilation techniques. We also describe how the lessons learned from ocean colour assimilation are more recently being adopted to assimilate bio-optical data, not only from space, but also from *in situ* autonomous measurement platforms.

In order to optimize parameter values for a one-dimensional biogeochemical model developed for the equatorial Pacific, Friedrichs (2002) conducted a series of experiments assimilating different subsets

of satellite and *in situ* data. The experiments specifically compared the assimilation of *in situ* cruise data with the assimilation of daily and 8-day composites of SeaWiFS ocean colour data over an ~8 month period. An adjoint of the ecosystem model was constructed following the method of Lawson et al. (1995), and was used to optimize parameters governing processes such as recycling, remineralization, mortality, growth, and grazing. Nearly identical parameter sets were obtained regardless of whether 8-day satellite composites or *in situ* data were assimilated. However, assimilating daily satellite data instead of 8-day composites resulted in unsuccessful parameter optimization experiments with no optimal parameter sets being identified. These results illustrate that when assimilating satellite-derived data into marine ecosystem models for the purpose of parameter estimation, modelled temporal scales must closely match those of the assimilated data.

Xiao and Friedrichs (2014b,a) optimized parameters for a 1D biogeochemical model at four continental shelf sites in the Mid-Atlantic Bight and investigated the effect of assimilating different types of satellite-derived data products. Assimilation experiments revealed differences in model-data misfits when assimilating satellite-derived total chlorophyll, size-fractionated chlorophyll, and/or particulate organic carbon (POC). Assimilating POC without assimilating chlorophyll generated substantially worse modelled chlorophyll estimates; in contrast, assimilation of chlorophyll without assimilating POC resulted in almost no decrease in model-data misfit. The best model-data misfits were achieved when both size-fractionated chlorophyll and POC were assimilated together.

Due to the high computational expense of running a model multiple times over a large 3D domain, parameter optimization is often carried out in 1D, as in the previous examples (though sometimes optimal parameters generated from a 1D assimilative application are then used in a 3D implementation, e.g., Oschlies and Schartau 2005; McDonald et al. 2012; St-Laurent et al. 2017. Identifying efficient and effective means of optimizing 3D biogeochemical models directly is a key challenge to more widespread use of parameter estimation, and overcoming this challenge requires experiments comparing diverse assimilation setups to provide guidance for further optimization efforts. Mattern and Edwards (2017) compared the performance of four different parameter estimation methods based on evolutionary algorithms (heuristic search routines) in a 3D model domain of the coastal eastern Pacific. Experiments were conducted for the year 2000 using assimilated SeaWiFS imagery, for two models of different complexities. It was found that the evolutionary algorithms, on average, converged to solutions quicker than random, uninformed searches. Multiple cost function formulations were tested, and correlations between the various cost functions enabled a more informed selection of an appropriate cost function definition.

The effectiveness of parameter estimation through data assimilation has improved through the considerable developments in effective utilization of ocean colour, as well as through development and deployment of new measurement platforms, which complement the synoptic views available from satellites. Efforts to assimilate data from novel platforms benefit from the lessons learned from assimilating ocean colour in studies like those mentioned above, and in recent years bio-optical observations from autonomous platforms have been effectively assimilated as well. For example, Bagniewski et al. (2011) assimilated chlorophyll, oxygen and nutrient observations from a Lagrangian float using a variational adjoint during the North Atlantic spring bloom to optimize parameters in three different 1D ecosystem models. Their efforts resulted in well-constrained parameters with small estimated uncertainties, however they concluded that subsurface measurements are important for estimating vertical carbon export. More recently, chlorophyll and POC observations derived from a buoyancy-driven glider were assimilated to optimize parameters in a 1D biogeochemical model for the Ross Sea (Kaufman et al. 2018). Parameter optimization experiments utilizing various subsets of the glider observations demonstrated some potential differences between assimilation of *in situ* versus satellite-derived data, and the importance of high-resolution data when attempting to constrain biogeochemical models in such remote and variable marine environments.

6.2 Role of Ocean Colour and Model Structural Uncertainties

6.2.1 Model structure uncertainty

Model structural uncertainty refers to incomplete or non-representative model equations that lead to errors in model predictions. These errors are, of course, inevitable, not least because biogeochemical models are a simplification of a complex web of plankton and nutrients, driven by imperfect hydrodynamic models. Although model formulations vary, the representation of processes like phytoplankton growth tend to be relatively insensitive to model changes. In contrast, models can be very sensitive to changes in zooplankton processes. In a numerical experiment, Baird (2010) used a size-resolved plankton model to investigate the effect of small changes in the structure of models on the model outputs. The clear result was that plankton models are extremely sensitive to the predator-prey links. That is, the addition of just one zooplankton class that persists, and consumes a unique diet of prey, could significantly change the model. This is one reason why biogeochemical models, which often have multiple nutrient and phytoplankton functional types, typically keep the number of zooplankton and higher trophic level types to a minimum.

6.2.2 Ocean colour data uncertainty

As for any other data, it is necessary to define the error co-variances associated with ocean colour observations for the assimilation into biogeochemical models. Ideally, such definitions should quantify the different sources of error (analytical, representation and difference-in-kind errors) at each model grid point and assimilation step. In practice, these sources are unknown and, typically, observation errors have been assumed uncorrelated and constant in space and time. For example, Natvik and Evensen (2003) fixed the chlorophyll percentage error to 35% of the concentration value, i.e., to the pre-launch accuracy target for SeaWiFS (Hooker et al. 1992). This error was scaled to different constant values in the global ocean basins by Nerger and Gregg (2007), taking account of the spatial variability of the accuracy of SeaWiFS chlorophyll data compared to *in situ* data (13% – 56% with global average of 31%; Gregg and Casey 2004). Constant percentage errors were applied also when assimilating optical data retrieved from ocean colour (Ciavatta et al. 2014).

Each ocean colour missions' sensors has a finite operational lifetime, and sensors differ in their design, with different band locations, band widths, radiometric sensitivities and orbit. Therefore, it is difficult to combine data from different ocean colour missions to produce a consistent time series of chlorophyll concentration for data assimilation over longer than a single mission period. Several studies have observed the effects of these discrepancies (e.g., Gregg and Casey 2010; Mélin 2016; see also Section 2.3.7). The advantages of defining spatial-temporal variable errors for improving the performance of ocean colour assimilation have been argued by approximating the variability as a function of i) the distance of the assimilated data from land (Ciavatta et al. 2011); ii) the variance of the data in the assimilated composites (Teruzzi et al. 2014) and, iii) mismatches between model predictions and assimilated observations (Ford et al. 2012).

Ocean colour products that provide data with per-pixel error values (GlobColour, Blower et al. 2009; ocean colour from the Climate Change Initiative (OC-CCI) of the European Space Agency, Brewin et al. 2013; Brewin et al. 2017a Jackson et al. 2017; Groom et al. 2019) offered new opportunities to characterize observation error in data assimilation systems. For instance, Ciavatta et al. (2016) used per-pixel bias and root-mean-square-error data of the OC-CCI product to compute unbiased values and variance of assimilated chlorophyll composites, in a decadal reanalysis of the North East Atlantic biogeochemistry.

Per-pixel error information helps to define the analytical error of the assimilated data more accurately, but defining the contributions of difference-in-kind and the representation error is less well understood. Jones et al. (2016) showed that modelling and assimilation of the remote sensing

reflectance (i.e., the "real" parameter observed by satellites) by-passed the difference-in-kind error between ocean colour chlorophyll products and biogeochemical model chlorophyll concentration, improving the simulation of Case I and Case II waters. Quantification of representation error and inclusion of observation error correlations remain open challenges in the framework of ocean-colour data assimilation.

6.3 Examples Studies

In this section we describe published examples of biogeochemical data assimilation, providing both a summary of the studies themselves, and also insights from the authors on the pitfalls and successful approaches to the particular stream and technique of assimilation.
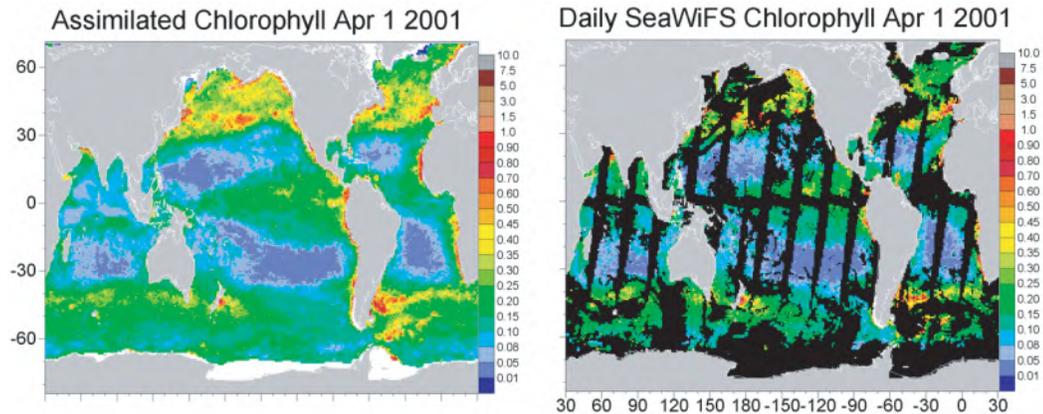
6.3.1 Assimilating chlorophyll

Data assimilation techniques can be used to constrain a model to track observation time series. Carmillet et al. (2001) used a singular evolutive extended Kalman filter to assimilate observations into a model for the North Atlantic and were able to constrain phytoplankton as well as fields such as nitrate. Natvik and Evensen (2003) used satellite chlorophyll assimilation to improve surface phytoplankton concentration and to reduce the variance of surface nitrate concentration. Gregg (2008) further demonstrated the potential of chlorophyll assimilation to improve model results by assimilating SeaWiFS chlorophyll data in a three-dimensional global ocean model. They found that this assimilation decreased the bias (21.0%) and uncertainty (65.3%) observed in the free-run compared to the assimilation (bias of 5.5% and uncertainty of 10.1%). Hemmings et al. (2002) assimilated satellite chlorophyll to refine model parameters in the North Atlantic and found that the assimilation improved chlorophyll concentrations in some areas but seasonal variability was poorly represented.

The assimilation of satellite chlorophyll is also known to reduce spatial and temporal biases when compared to satellite data. The assimilation of satellite chlorophyll produces estimates within 0.1% bias and 33.4% uncertainty as compared to *in situ* data (Gregg 2008, Figure 6.2) thereby decreasing the bias of satellite data (-1.3%) while keeping the same level of uncertainty (32.7%, Gregg 2008). Rousseaux and Gregg (2014) compared the chlorophyll concentration from a free-run with those from a run that assimilates satellite chlorophyll and found that the assimilation consistently improved the chlorophyll estimates in all major oceanographic regions. They found that the assimilation improved the chlorophyll concentration in the regions with the highest bias by 10–20%.

6.3.2 Assimilating remotely-sensed diffuse attenuation coefficient using localized ensemble Kalman filter (EnKF)

Modelling and assimilation of water optical properties can be useful for improving the simulation of biogeochemical variables in oceanic and coastal areas (see e.g., Fujii et al. 2007). Bio-optical modelling exploits the causal links between optical properties and key biogeochemical variables and processes, enhancing the simulation realism for marine ecosystems (Fujii et al. 2007). In addition, remotely-sensed optical properties can be retrieved with lower errors than chlorophyll, in particular in shelf-seas affected by terrestrial inputs of coloured matter (see, e.g., Lee et al. 2005a; Saba et al. 2011; Zhao et al. 2013). Ecosystem models that include the description of water optical properties have been proposed and successfully applied in marine studies (e.g., Gregg and Walsh 1992; Bissett et al. 1999b, 2005; Manizza et al. 2005; Fujii et al. 2007; Mouw et al. 2012. However, the assimilation of satellite-derived optical data into ecosystem models is still in its infancy. In this case study we show the benefits of assimilating data of diffuse light attenuation coefficient from ocean colour (Ciavatta et al. 2014).

Figure 6.2 Comparison of chlorophyll (mg m^{-3}) for April 1, 2001 from (left) the assimilation model and (right) the daily SeaWiFS Chlorophyll for April 1, 2001. Reprinted from Gregg (2008) with permission from Elsevier.

### 6.3.2.1 Study region and model configuration

The study area is the western English Channel (WEC), a shelf-sea region that includes both Case I and Case II waters (Groom et al. 2009). The model of the WEC ecosystem (Lewis and Allen 2009) couples the hydrodynamic model POLCOMS (Proudman Oceanographic Laboratory Coastal Ocean Modelling System; Holt and James 2001) with the biogeochemical model ERSEM (European Regional Seas Ecosystem Model; Blackford et al. 2004) and includes a bio-optical module (Ciavatta et al. 2014). ERSEM describes the dynamic of four phytoplankton functional types, three zooplankton types and one bacteria type, characterized by variable internal stoichiometry. The model includes the dynamics of five inorganic dissolved nutrients (carbon, nitrate, ammonia, phosphate and silicate), dissolved oxygen, particulate and dissolved organic matter. The bio-optical module describes the spatio-temporal variability of the spectral diffuse attenuation coefficients, $K_d(\lambda)$, at three optical bands (blue λ = 443 nm; green λ = 555 nm; red λ = 670 nm). $K_d(\lambda)$ is a function of the inherent optical properties (IOPs) of the optically active compounds simulated by ERSEM (Lee et al. 2005b) and, in turn, the IOPs are computed using bio-optical equations and parameterizations from the scientific literature (see Ciavatta et al. 2014 for details).

6.3.2.2 Data assimilation system

Diffuse attenuation coefficient in the blue band, $K_d(443)$, measured by SeaWiFS, was assimilated into the WEC ecosystem model by using the Ensemble Kalman filter (EnKF) (Evensen 1994, 2003; Ciavatta et al. 2011). The EnKF was applied with: i) a localized configuration (with spatially variable radius); ii) perturbation of the observations; iii) analysis with log-transformation of state and observations; and iv) 100 ensemble members. Model error is accounted for through the stochastic perturbation of the diffuse attenuation coefficients in the EnKF forecast (Gaussian perturbation with standard deviations equal to 10% of the simulated $K_d(\lambda)$ values). Model error was also added to all the forecasted variables prior to the analysis step (10% of the value of the variables). The observation error was assumed uncorrelated and constant, i.e., equal to 20% of the assimilated $K_d(443)$ values. This system was applied in a one-year long reanalysis, spanning the year 2006, by assimilating five-day composites of $K_d(443)$ into the ecosystem model, with a weekly frequency. The performance was also compared to a reanalysis were total chlorophyll from SeaWiFS was assimilated into the same system.

6.3.2.3   Skill improvement of the system

The data assimilation performance for year 2006 was assessed using independent, weekly data of 18 biogeochemical and optical time series measured at station L4 of the Western Channel Observatory (WCO), a NERC National Capability at PML (www.westernchannelobservatory.org.uk). The comparison between the reference model run (without assimilation) and the assimilative run showed that assimilation reduced the root mean square error (RMSE) and improved the correlation with the (unassimilated) L4 data for most variables (see Figure 6.3a). Moreover, $K_d(443)$ assimilation performed better than the assimilation of total chlorophyll data (Figure 6.3b).
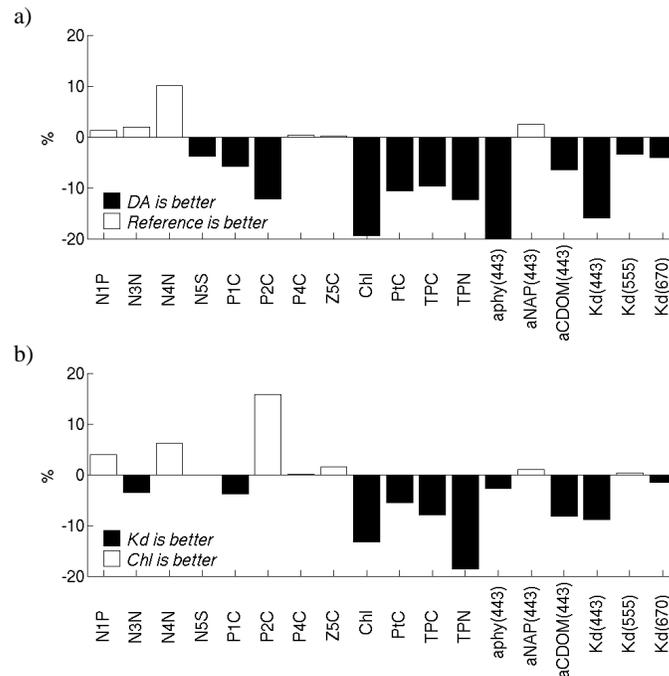
Figure 6.3   a) Skill of $K_d(443)$ assimilation in estimating the not-assimilated, *in situ* data. The bars represent the percentage differences between RMSE of the assimilation (DA) versus the reference simulation; b) Skill for $K_d(443)$ versus chlorophyll (chl) assimilation in estimating the not-assimilated, *in situ* data. The bars represent the differences between the RMSE obtained in the two assimilative runs. The x-axis lists model variables: N1P - phosphate; N3N - nitrate; N4N - ammonium; N5S - silicate; P1C - diatom carbon biomass; P2C - nanophytoplankton; P3C - picophytoplankton; P4C - dinoflagellates; Z5C - micro-zooplankton; Chl - chlorophyll; PtC - total phytoplankton carbon biomass; TPC and TPN - total particulate carbon and nitrogen, respectively; $K_d(\lambda)$ - diffuse light attenuation coefficient at the wavelength (λ); $a_{CDOM}(\lambda)$, a_{ph}, a_{det} - absorption of CDOM, phytoplankton and detritus, at the wavelength λ = 443 nm. Figure modified from Ciavatta et al. (2014), Creative Commons Attribution License (CC BY).

Assimilation of $K_d(443)$ also impacted the simulation of the carbon fluxes within the plankton community (Figure 6.4a-c), resulting in a shift of the simulated food chain from the herbivorous food chain towards the microbial loop (see the conceptual diagram in the lower part of Figure 6.4). This shift contributed to improving the simulation of total particulate carbon (TPC in Figure 6.3, which is a diagnostic model variable computed by summing up eleven model variables across the whole simulated trophic system).
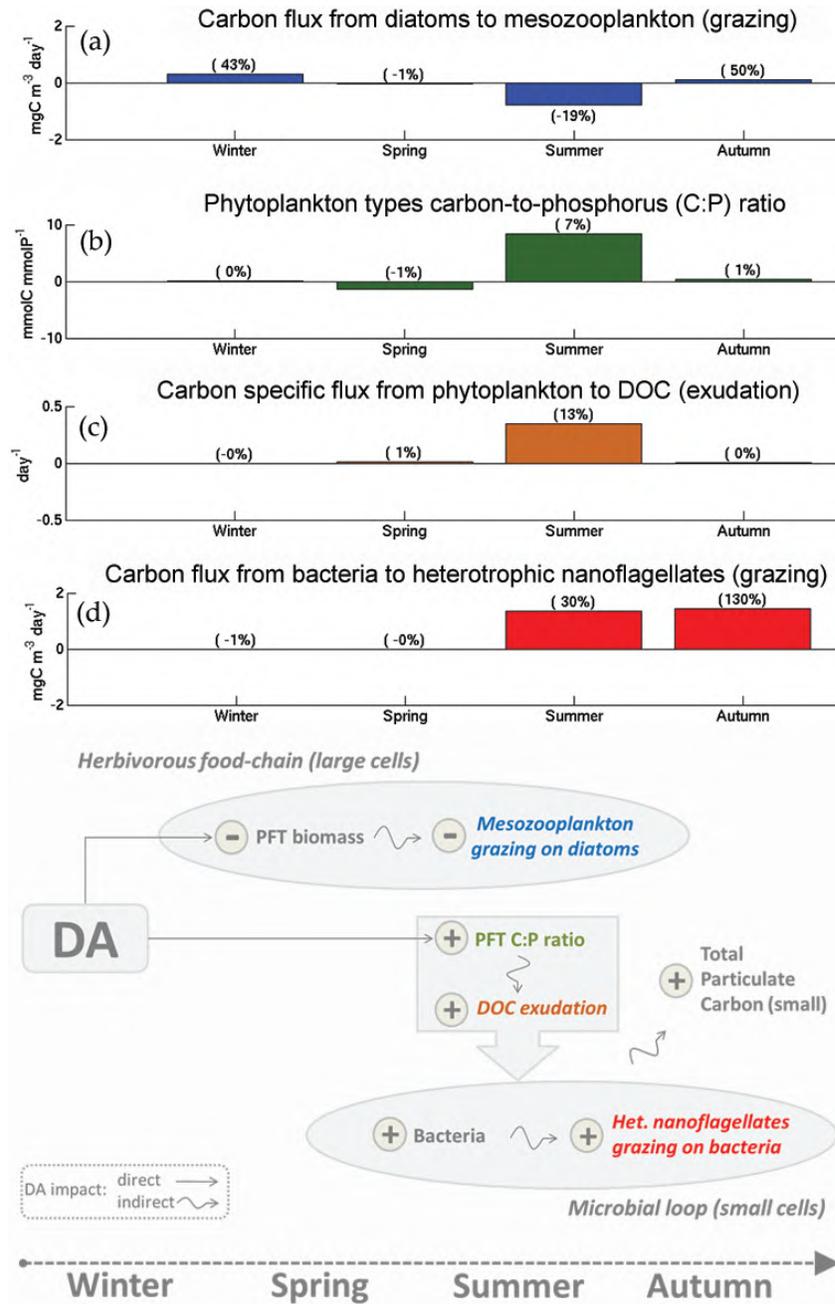
Figure 6.4 Scheme of the direct and indirect effects of data assimilation on the simulation of the trophic dynamic at the *in situ* monitoring site. In the graphs, the bars represent the differences between selected carbon fluxes resulting from the data assimilation and reference simulations. The differences refer to winter (January – March), spring (April – June), summer (July – September) and autumn (October – December) 2006. The percentage values in brackets are the differences normalized by the mean seasonal values from the reference simulation. Reproduced from Ciavatta et al. (2014), Creative Commons Attribution License (CC BY 3.0).

6.3.2.4   Lessons learned

In this case study, the assimilation of $K_d(443)$ not only improved the reference model simulation of unassimilated biogeochemical data, but it also outperformed the assimilation of total chlorophyll. To explain this result, the lower error of the satellite optical products should be considered, in particular in coastal Case II waters (compare, e.g., Saba et al. 2011; Zhao et al. 2013), which is advantageous because it can drive the assimilation corrections closer to the "true values" of the assimilated variable. Secondly, the light attenuation coefficient assimilated in this case study represents a bulk property of marine systems, summing up the effects of optically active variables ranging from inorganic suspended solids to phytoplankton, while chlorophyll approximates just the last mentioned compound. Thus, in principle, the assimilation of bulk properties could drive direct corrections to a larger set of model variables. In practice, this requires further efforts to properly parameterize and model the contributions of the different optical compounds, so that multivariate assimilation can partition the corrections among the model state variables properly.

Finally, this case study shows that ocean-colour assimilation can impact the simulation of emergent properties of the ecosystem, such as the shift of the food web through the continuum pathways proposed by Legendre and Rassoulzadegan (1995). This was allowed by the complexity and plasticity of the model applied in the case study (ERSEM), i.e., by the adaptability of the model functional groups to the changes in the variables imposed by the data assimilation system. In particular, modelling variable carbon-to-chlorophyll and carbon-to-nutrients allowed ocean colour assimilation to impact the simulation of the trophic processes, and eventually, to simulate better ecosystem indicators and fluxes.

6.3.3   Assimilating remote-sensing reflectance using a deterministic ensemble Kalman filter (DEnKF)

In some of the above approaches, one source of error in the assimilation system was the need to use an IOP algorithm to calculate the observed state variable against which the model variable was compared. This error has been reduced in some studies by using a more basic IOP (i.e., vertical attenuation K_d, Section 6.3.2) instead of the concentration of optically-activity constituents such as chlorophyll. In this case study (described in Jones et al. 2016 with on-going improvements, in preparation) the data assimilation system goes one step further — the assimilation used the mis-match between simulated and observed remote-sensing reflectance, R_{rs}, directly, thus avoiding the need for an IOP algorithm.

6.3.3.1   Study region and model configuration

The biogeochemical model used data from the Great Barrier Reef off the northeast coast of Australia, a region characterized by complex coastal waters influenced by terrestrial runoff. The waters are also shallow, with benthic reflectance influencing the remotely-sensed observations. The biogeochemical model contained >20 optically-active constituents (including phytoplankton, CDOM, sediments, bottom substrates) that contribute to the simulated remote-sensing reflectance. The optical model calculated R_{rs} for each horizontal pixel based on an optical-depth weighting of the ratio of the vertical profile of the backscatter to the absorption plus backscatter. Analysis of model output showed that the error between the observed and simulated R_{rs} was primarily due to errors in the dynamical prediction of the time- and space- varying optically-active constituents, not the calculation of the simulated R_{rs}. Thus, in correcting R_{rs} through updating optically-active state variables, the data assimilation model improved the prediction of the state variables themselves.

6.3.3.2   Data assimilation system

The study employed a deterministic ensemble Kalman filter (DEnKF) constrained by the mismatch between the observed R_{rs} and the simulated R_{rs}. The observed R_{rs} was calculated using local atmospheric

correction, a necessity in the complex coastal waters. The ensemble size employed was 108 members, and used the software package EnKF-C. Care was taken to match the time of the simulated and observed R_{rs} to within 30 minutes.
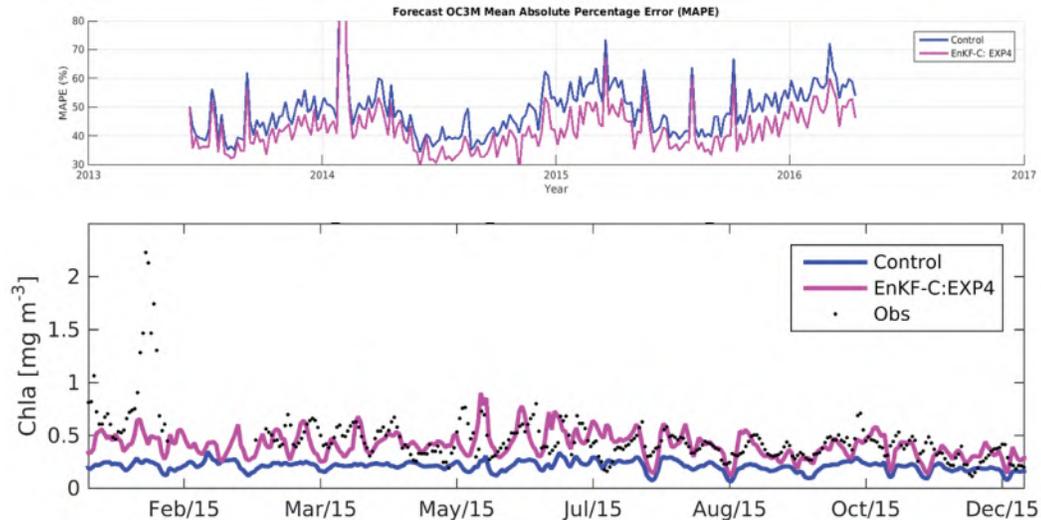
Figure 6.5   Top panel: Comparison of the skill of the unassimilating (control) run and the mean of the ensemble, quantified as a mean absolute percent error of the mismatch between the simulated and observed R_{rs}. Bottom panel: Comparison of the chlorophyll concentration measured by a moored fluorometer (Obs), and the control and assimilating runs at one site (Double Cone off Airlie Beach, central Queensland) at 5 m depth.

6.3.3.3   Skill improvement of the system

The assimilation reduced the mean absolute percent error (MAPE) of the prediction of the variable being assimilated, OC3M (the MODIS algorithm based on the blue-green band ratio) from 50 % to 42 % (Figure 6.5, top). Further, the assimilation system improved the comparison to *in situ* data at 13 of the 14 observations sites. The system was able to correct both the bias, and add variability that matched the observed variability (Figure 6.5, bottom).

6.3.3.4   Lessons learned

The assimilation of remote-sensing reflectance, instead of IOPs, introduces the possibility of a large observation state vector composed of 8 bands, in the case of MODIS ocean colour bands, or more for future hyperspectral sensors. In reality, it was found that some remote-sensing reflectance bands (observed state) either co-varied with optically-active constituents (model state), or were uninformative. A non-exhaustive investigation into combinations of remote-sensing reflectance bands for the purpose of constraining the model found the ratio of the blue and green bands provide the best constraint. Further, if this ratio was quantified by the OC3M algorithm, the assimilating observed state took a form that was more familiar to biogeochemical modellers (i.e., units of mg m^{-3}) than remote-sensing reflectance. Thus, while assimilation of hyperspectral remote-sensing reflectance has the promise of vastly larger observational data sets, and therefore better system performance, the law of diminishing returns, and the careful selection of spectral bands in the existing moderate resolution sensors to capture chlorophyll, suggests little will be gained from future hyperspectral satellites for data assimilation.

A second lesson learned from the assimilation of like-for-like variables was the simplicity with which observed and model states can be compared. Firstly, there is no need for a locally-calibrated IOP algorithm. But perhaps more importantly, ocean colour is inherently a 2-dimensional field, while the model state is quantified by a 3-dimensional field of optically-activity constituents such as phytoplankton. By comparing the calculated remote-sensing reflectance, there is no need to determine a depth-weighting function for the integral of the 3-dimensional field to reduce for comparison to 2-dimensional ocean colour.

6.3.4   Assimilating phytoplankton functional type (PFT) data

The size of the phytoplankton groups composing the ocean plankton community plays a key role in marine ecology and biogeochemical cycling, because the size impacts growth, nutrient assimilation, sinking rate, and predation (Chisholm 1992). State-of-the-art marine ecosystem models describe size-based phytoplankton functional types (PFTs) to increase the realism of marine ecosystem models (a non-exhaustive list of such models includes ERSEM (Baretta et al. 1995; Butenschön et al. 2016), PISCES (Aumont et al. 2003), NOBM (Gregg et al. 2003), BFM (Vichi et al. 2007), the MITgcm-biogeochemical (Follows et al. 2007), PlankTOM (Le Quéré et al. 2005), and MEDUSA (Yool et al. 2013). However, the simulation of phytoplankton types remains challenging, mainly due to the difficulty of formulating their processes (Shimoda and Arhonditsis 2016). The assimilation of PFT data is a promising new approach to improve the simulation of PFTs and their biogeochemical feedbacks. Xiao and Friedrichs (2014a) showed that assimilating ocean-colour PFT data improved the simulation of two plankton groups with a testbed one-dimensional model of the Mid-Atlantic Bight. The case study presented here shows the benefits of assimilating a regional product for four PFTs into an ecosystem model of the North East Atlantic in a multiannual reanalysis (Ciavatta et al. 2018).

6.3.4.1   Study region and model configuration

The study area is the North East Atlantic (NEA), including the North West European Shelf (NWS). The model of the NEA ecosystem (Artioli et al. 2012) couples the hydrodynamic model POLCOMS (Proudman Oceanographic Laboratory Coastal Ocean Modelling System; Holt and James 2001) with the biogeochemical model ERSEM (European Regional Seas Ecosystem Model; version 15.06: Butenschön et al. 2016). ERSEM has been described in Section 6.3.3 as a model capable to describe the dynamics of four phytoplankton functional types (diatoms, dinoflagellates, nanoplankton, picoplankton), three zooplankton types and one type of bacteria, characterized by variable internal stoichiometry. This case study used a novel parameterization of ERSEM which enhanced the differences among the PFTs, i.e., the differences among PFT rates and feedbacks to the ecosystem biogeochemistry (Butenschön et al. 2016).

6.3.4.2   Data assimilation system

Ocean-colour PFT data for the NEA region — diatoms, dinoflagellates, nanophytoplankton, and picophytoplankton, parameterized as a function of sea-surface temperature and provided with per-pixel errors by Brewin et al. (2017a) (EU CMEMS TOSCA and ESA's CCI Ocean Colour projects) — were assimilated into the NEA ecosystem model using the ensemble Kalman filter (EnKF) (Evensen 2003; Ciavatta et al. 2018). The EnKF was applied here with i) a localized configuration (with spatially variable radius); ii) perturbation of the observations; iii) analysis with log-transformation of state and observations; and iv) 100 ensemble members. Model error was accounted for through the stochastic perturbation of the solar irradiance at sea-surface (Gaussian perturbation with standard deviations equal to 20% of the irradiance values). Model error was also added to all the forecasted variables (10% of the value of the variables) prior to the first annual analysis steps. The observation error was assumed uncorrelated and was derived from the per-pixel values of bias and root-mean-square-deviation provided with the PFT

data (Ciavatta et al. 2016). Five-day composites of the ocean-colour PFTs were assimilated monthly in a reanalysis simulation spanning the years 1998–2003. A reanalysis assimilating total chlorophyll data from ocean colour was also performed for years 1998–2003, for comparison.

6.3.4.3 Skill improvement of the system

The reanalysis outperformed the reference model simulation in estimating not only the assimilated plankton functional types, but also the non-assimilated total chlorophyll (Figure 6.6, first row, Ciavatta et al. 2018). In general, the reference model simulation overestimated the PFTs on the ocean and shelf regions, in particular diatoms and picophytoplankton in the Northern region (not shown). The reanalysis reduced the overestimation by systematically correcting the PFTs towards the lower observed concentrations, reducing the model RMSD up to 50% for dinoflagellates (Figure 6.6, first row). The reanalysis improved the model estimation of nanophytoplankton on the shelf, though this was deteriorated on the ocean part of the domain.
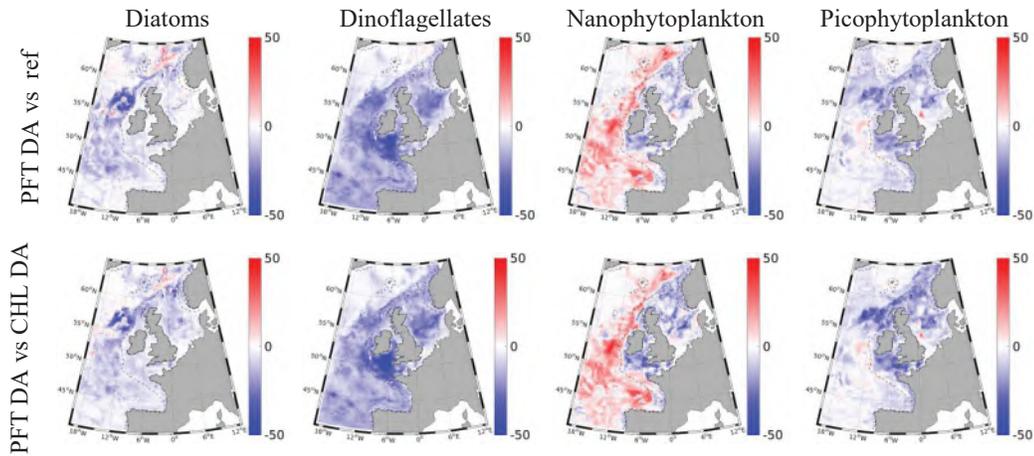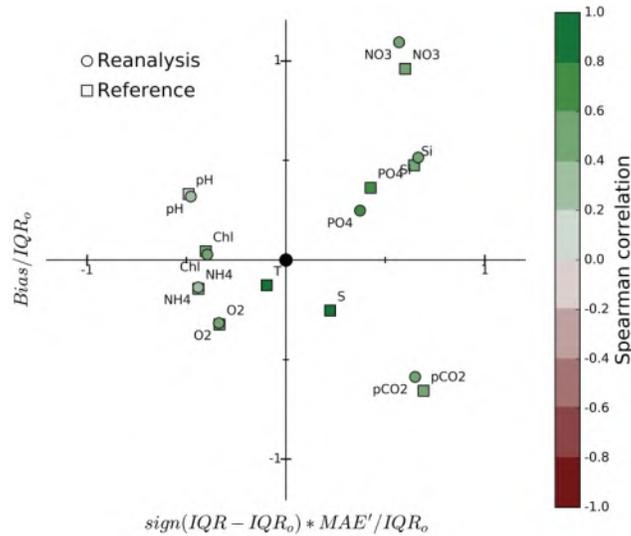

Figure 6.6 Skill of the reanalysis in simulating the ocean-colour phytoplankton functional types. The maps show the percentage differences (%) between the data-to-output root-mean square-deviation (RMSD) of the PFT assimilation versus the reference simulation without assimilation (1st row) and versus total chlorophyll assimilation (2nd row). Negative values indicate that the RMSD of PFT assimilation is lower than the RMSD of the alternative simulation. The dotted lines indicate the 50 m and 200 m isobaths. Figure modified from Ciavatta et al. (2018), Creative Commons Attribution License (CC BY).

Quantitative skill metrics demonstrate that assimilation of ocean-colour markedly improved the model simulation of *in situ* data of pCO$_2$ and phosphate (Figure 6.7), as well as for all the PFTs. However, the reanalysis also clearly increased the bias of nitrate and, slightly, of silicate. The skill for the remaining variables was affected only negligibly (i.e., pH, O$_2$, NH$_4$). Temperature and salinity were not included in the analysis, though their good reference skill is reported for completeness.

6.3.4.4 Lessons learned

In this case study, the assimilation of PFT data not only improved the reference model simulation of phosphate and partial pressure of CO$_2$, but it also outperformed the assimilation of total chlorophyll in estimating both the ocean-colour PFTs and the total chlorophyll itself. In fact, total chlorophyll assimilation performed relatively poorly in this case study, when compared to the application by Ciavatta et al. (2016) in the same region. This follows the use of a new, more realistic PFT parameterization,

Figure 6.7 Skill of the reanalysis (circles) and reference (squares) simulations in estimating *in situ* data of ten physical and biogeochemical variables observed in the shelf waters in the years 1998–2003. The *in situ* observed variables are: NO3 - nitrate; Si - silicate; PO4 - phosphate; Chl - chlorophyll; NH4 - ammonium; O2 - oxygen; pCO2 - partial pressure of CO_2; T - temperature; S - salinity. The robust skill metrics represented in the diagram are: bias, Spearman correlation; IQR: interquartile range; MAE': unbiased median absolute error. The subscript "o" indicates "observations". *In situ* data were extracted from the ICES (www.ices.dk) and SOCAT (http://www.socat.info; Bakker et al. 2014) databases. Figure modified from Ciavatta et al. (2018), Creative Commons Attribution License (CC BY).

which enhanced the non-linearity between the concentrations of total chlorophyll and of the PFTs, weakening the linear approximation and effectiveness of the ensemble Kalman filter in the case of total chlorophyll assimilation. On the other hand, PFT assimilation corrected, directly and selectively, the biomasses of the PFTs through linear links between observed and simulated variables. PFT assimilation also improved the simulation of key emergent properties of the ecosystem, i.e., the net ecosystem respiration and the flux of atmospheric carbon dioxide at the air sea-sea interface (Figure 6.7). This case study suggests that the assimilation of ocean-colour PFTs is a new promising approach to simulate marine ecosystems, with potential benefits in the context of operational oceanography (Skákala et al. 2018) as well as analysis of PFT distributions and impacts on ecosystem functioning at regional (Ciavatta et al. 2019) and global scales (Pradhan et al. 2019).

6.3.5 Assimilation of satellite derived bio-optical properties: impact on short-term model predictions

The objective in this study was to investigate whether the assimilation of satellite-derived bio-optical properties (either as chlorophyll or absorption coefficient) can improve the ecosystem model predictions of chlorophyll and phytoplankton populations in a coastal ocean on time scales of 1–5 days. The specific time scale of 1 – 5 days was chosen because it coincides with the time scale of the atmospheric model forecast needed to force the oceanic model forecast.

Data assimilation experiments were conducted during five days of steady upwelling in the Monterey Bay area. With the assimilation of satellite-derived bio-optical properties (chlorophyll-a or absorption due to phytoplankton), the model was able to reproduce intensity and tendencies in subsurface chlorophyll distributions observed at water samples locations in Monterey Bay, CA. Data assimilation

also improved agreement between the observed and model-predicted ratios between diatoms and small phytoplankton populations (Shulman et al. 2013).

6.3.5.1   Study region and model configuration

Computational experiments were designed to coincide with a large bio-optical field campaign that was conducted in Monterey Bay, California during a sustained wind-driven upwelling event in June, 2008. The field programme captured the dynamic response of the Bay ecosystem to the continuous supply of nutrients from coastal upwelling.

The Monterey Bay model consists of a physical model (primitive equation model), which is coupled to a biochemical model (Shulman et al., 2013). The biochemical model simulates the dynamics of two sizes of phytoplankton, small phytoplankton cells (< 5 μm in diameter) and diatoms, two zooplankton grazers, nitrate, silicate, ammonium, and two detritus pools (Chai et al. 2002). Constituents from the biochemical model are used to estimate chlorophyll and inherent optical properties (IOPs). For example, absorption due to phytoplankton, $a_{ph}(\lambda)$, is modelled as the sum of absorption from small phytoplankton and diatoms. The chlorophyll-specific absorption coefficients for small phytoplankton and diatoms are modelled separately, taking into account their photo-adaptive state (e.g., their specific chlorophyll-to-carbon ratio). This requires specification of high/low light absorption coefficients for each phytoplankton group (small phytoplankton and diatoms). For more details see Shulman et al. (2013). Phytoplankton chlorophyll-to-carbon ratios are not constant and depend on light, nutrients, temperature etc. However, to model the ratio as variable will require the introduction of more state variables, as well as more highly-uncertain model parameters into the bio-chemical model. Because the objective was modelling on short-term time scales (1 to 5 days), constant relations were used, rather than increasing the number of biochemical model state variables and highly-uncertain model parameters.

6.3.5.2   Data assimilation system

Chlorophyll-a (Chl) data for assimilation were derived from MODIS-Aqua imagery using the OC3M algorithm (O'Reilly et al. 2000), while phytoplankton absorption data, $a_{ph}(488)$, were derived using the quasi-analytical algorithm (QAA) of Lee et al. (2002). Resolution for both data sets is 1 km.

To assimilate bio-optical measurements into the ecosystem model, reduced-order Kalman filter was used with a stationary forecast error covariance. The forecast error covariance was specified in the subspace of the multivariate (bio-optical, physical) empirical orthogonal functions (EOFs) estimated from a month-long model run. It was assumed that observational error covariance had diagonal structure (uncorrelated errors), stationary and was equal to 10% of the field variance (see Shulman et al. 2013 for more details). No localization of the forecast error covariance was needed to assimilate MODIS-Aqua Chl data into the model. However, localization was needed for assimilation of phytoplankton absorption to mitigate the presence of spurious correlations in the approximation to the forecast error covariance, and to exclude remote observations from the analysis of the local grid point. For assimilation of MODIS-Aqua derived phytoplankton absorption, the forecast error covariance was localized using the box-car localization function with localization distance set to 10 km.

A 12 h data assimilation cycle was used, and assimilation of physical properties into the physical model of the coupled system was separated from the assimilation of bio-optical properties into the biochemical model. Therefore, for each 12 h of the model run, physical observations were first assimilated into the physical model, creating a new restart file (containing physical and biochemical state variables) with updated (analyzed) temperature and salinity fields. This restart file was used to assimilate MODIS-Aqua Chl or absorption data into the biochemical model, creating a new restart file (nowcast) with updated (analyzed) small phytoplankton and diatom fields. The next segment of the

model run was started from this restart file and was run for 12 h until the next model restart file was created.

6.3.5.3   Skill improvement of the system

The assimilation of MODIS-Aqua derived optical properties (Chl or phytoplankton absorption) improved surface and subsurface agreement between the model and observations (Figure 6.8). Results show that the reduction in RMSE errors between model and independent water samples ranged from 5% to 35%, in contrast to the non-assimilative run (Shulman et al., 2013). Assimilation of MODIS-Aqua bio-optical observations increased (decreased) the concentration of diatoms (small phytoplankton) inside the Bay in comparison to the non-assimilative run. At the same time, the assimilation of $a_{ph}(488)$ also created an artificial tongue of small phytoplankton offshore from the northern part of the domain along the coast (Shulman et al., 2013). This might be a result of difficulties in assimilation of offshore values of absorption, which are significantly lower in comparison to the values in the Bay. Assimilation of bio-optical data improved fractionation of phytoplankton biomass between diatoms and small phytoplankton in the model. Without assimilation, the percentage of large diatoms varied during the experiment between 20% and 80% (Figure 6.8). In contrast, high performance liquid chromatography (HPLC) observations showed that the fraction of diatoms to the total phytoplankton population was in the range of 90%. Assimilation of MODIS-Aqua surface chlorophyll produced much better agreement with the independent, non-assimilated HPLC observations. With the assimilation, the RMSE error between HPLC observed and model-predicted fraction of diatoms is less than the RMSE error for the non-assimilative run. There were also improvements in the fraction of diatoms to total phytoplankton predictions for the run with assimilation of $a_{ph}(488)$ after a couple days of assimilation (Figure 6.8).

6.3.5.4   Lessons learned

While the assimilation improved the model predictions, the model subsurface Chl distributions retained an under-prediction bias as compared to observed profiles from water samples (Figure 6.8). One of the reasons might be that MODIS-Aqua bio-optical data are assimilated as observed surface values, while satellite data provide an estimate of the average, e.g., chlorophyll concentration over the layer between the surface and one attenuation depth.

The assimilation of MODIS-Aqua observations did not improve the model predictions of nitrate. This can be explained by the fact that multivariate data assimilation tends to increase the phytoplankton population in the Bay (due to the underestimated *a priori* Chl and absorption values in the model), and at the same time tends to decrease nutrients. Results showed that an instantaneous update of nitrate based on statistical relations between temperature and nitrate (derived from the AUV observations taken prior to the data assimilation experiments) corrected the model underestimation of the nitrate fields (Shulman et al. 2013).

The experiments conducted in this study, were limited to a five-day period during a steady upwelling event. More complicated bio-optical conditions are usually observed during wind weakening and relaxation, when transitions from diatoms to other phytoplankton groups might occur with corresponding drastic changes in bio-optical properties on time scales of days to a week. Our experiments with the ensemble computed from a month-long model simulation suggest that ensemble methods are very capable at capturing complex multi-variate relationships between optical properties, phytoplankton biomass, and ecosystem structure (as represented by small and large phytoplankton pools in the model).

Finally, in this study, assimilation of physical properties and assimilation of bio-optical properties are separated. The adjustment of updated physical and bio-optical variables is achieved through the coupled, bio-optical physical model run during the data assimilation cycle. At the same time, an instantaneous joint update of physical and bio-optical properties is preferable, to maintain dynamical consistency between the assimilated physical and bio-optical fields.
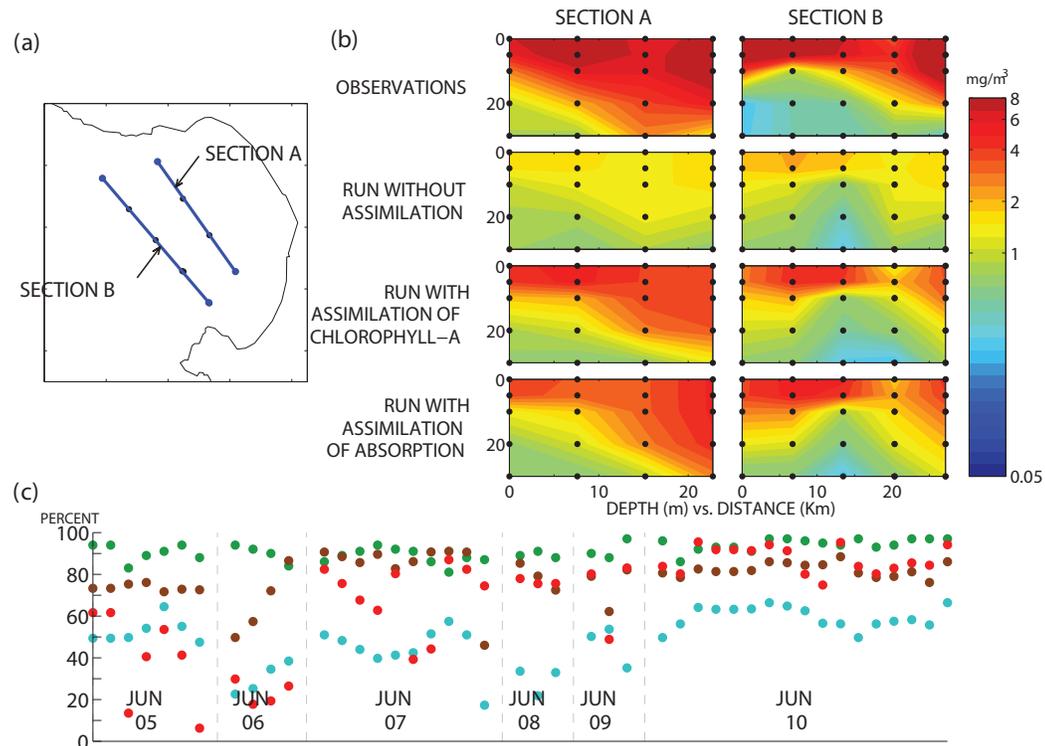
Figure 6.8 (a) Location of water sampling stations (Sections A and B); (b) Comparisons of observed and model-predicted subsurface chlorophyll distributions at water sampling stations in Sections A and B; (c) Observed and model-predicted fractions of diatoms to total phytoplankton population at sampling locations: green = HPLC observed fractions; light blue = run without assimilation of bio-optical properties; brown = run with assimilation of satellite Chl data; red = run with assimilation of phytoplankton absorption.

6.4   State Estimates/Re-analysis

Long-term trends and interannual variability of biogeochemical variables that are relevant to climate studies and marine policy can be evaluated by merging numerical models and ocean colour in an extended "biogeochemical reanalysis" (Lahoz and Schneider 2014). Typically, this is a multi-annual model simulation that assimilates ocean colour time series with a consistent model set-up and analysis algorithm, which corrects the model state variables to provide consistently-processed, gap-free estimates of the "true" state of the ocean. These estimates include non-assimilated variables that cannot be observed from satellite and that are sparse in time and space (e.g., nutrient and oxygen, Figure 6.9), as well as biogeochemical fluxes and emergent ecosystem properties simulated by the model (e.g., air-sea carbon fluxes). Furthermore, reanalyses can be used to provide boundary and initial conditions for other models, e.g., in the context of operational oceanography (Gehlen et al. 2015). They can also offer insides into model biases and processes, which can be used to inform model development and observing network design. In performing reanalysis, special care is needed in choosing the assimilated ocean colour time series, to avoid that spurious ocean colour signals (e.g., trends due to sensor decays or inter-sensor shifts) that can drive unrealistic trends and shifts in the reanalysed ocean biogeochemistry (Ford and Barciela 2017).

Assimilative reanalysis is a well-established approach in environmental disciplines such as atmospheric science (Bengtsson and Shukla 1988; Trenberth and Olson 1988) and ocean physics modelling

Total Chlorophyll Trend (% y⁻¹)

Diatom Trend (% y⁻¹)

Cyanobacteria Trend (% y⁻¹)
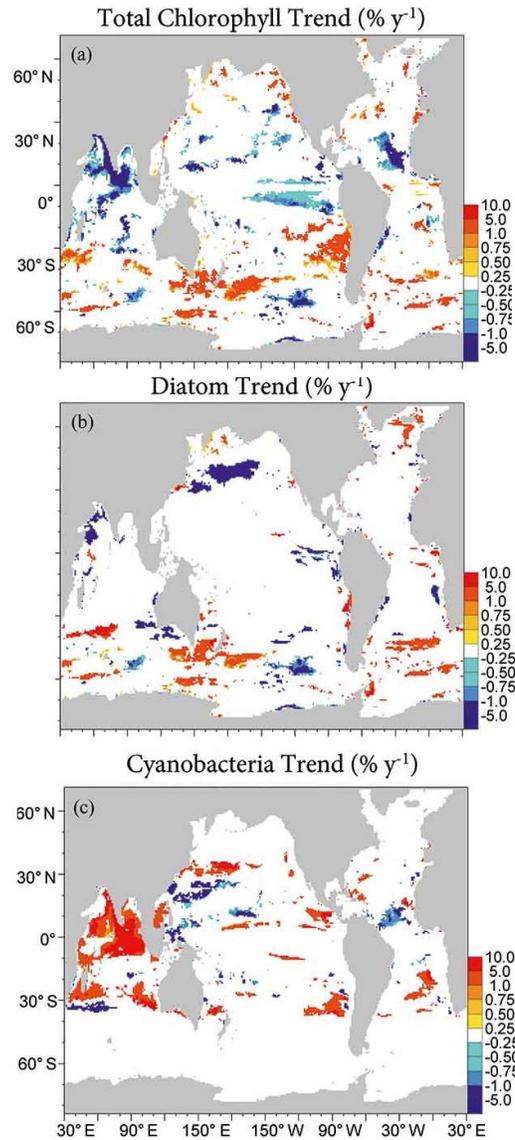
Figure 6.9 Reanalysis of the global ocean ecosystem by assimilating integrated ocean colour time series of chlorophyll concentrations measured by SeaWiFS, MODIS, and VIIRS. Significant trends in a) total chlorophyll, b) diatoms, c) cyanobacteria over the period 1998–2015 (% y^{-1}). From Gregg et al. (2017), reprinted with permission from Taylor & Francis Ltd.

(Stockdale et al. 1998), however it is relatively new in the framework of ocean biogeochemical modelling. The first (quasi) decadal biogeochemical reanalysis estimated the interannual variability of global primary production in years 1998–2004 by assimilating chlorophyll from SeaWiFS into the NASA Ocean Biogeochemical Model (OBM, Nerger and Gregg 2007). A comparable variability of primary production was obtained in the reanalysis by Gregg (2008), who in addition described the spatial patterns of chlorophyll in the global oceans. The reanalysis by Fontana et al. (2013) evaluated spatial-temporal patterns of chlorophyll and nitrate in the North Atlantic Ocean in years 1998–2006, by assimilating

SeaWiFS chlorophyll into a coupled physical-biogeochemical model. Reanalyses for years 1998–2012, using chlorophyll observations from SeaWiFS and MODIS and the NASA OBM, evaluated significant declining trends of chlorophyll in the Northern Hemisphere and Indian oceans (Gregg and Rousseaux 2014), and estimated declining trends of phytoplankton functional groups in part of the global oceans (Rousseaux and Gregg 2015). These trends were confirmed by the reanalyses of Gregg et al. (2017), who integrated and assimilated ocean colour data from SeaWiFS, MODIS, and VIIRS, spanning the years 1998–2015 (Figure 6.9). Ford and Barciela (2017) assimilated two different ocean-colour products in separate global reanalyses (ESA's Ocean Colour–Climate Change Initiative, and GlobColour), and found that, despite some differences, both reanalyses improved not only the simulation of *in situ* chlorophyll profiles, but also the estimates of nutrient concentrations and carbon dioxide fugacity.

The above works all demonstrated the value of reanalysis for open ocean ecosystems. However, only a few examples are currently available for biogeochemical reanalysis in shelf-sea ecosystems, where the influence of riverine inputs and sediments make less obvious the availability of adequate ocean colour time series. Ciavatta et al. (2016) assimilated an ocean colour product developed for both Case I and II waters, with per-pixel error estimates of bias and root-mean-square-deviation (ESA's OC-CCI Version 3; Sathyendranath et al. 2017) in a decadal reanalysis of biogeochemical indicator and fluxes in the North West European shelf (1998–2009). The reanalysis improved the model prediction of the assimilated ocean colour chlorophyll and provided quality-assessed estimates of biogeochemical indicators that are relevant for marine policy (e.g., dissolved oxygen deficiency, Figure 6.10). However, the skilled reference simulation of biogeochemical variables was not improved significantly by assimilating the ocean colour product, possibly due to the relatively low assimilation frequency (monthly) and relatively high errors of the ocean colour chlorophyll product in coastal areas. In shelf-seas, alternative ocean colour products have been suggested as valuable approaches to reanalysis, e.g., plankton functional types and optical data (see Sections 6.3.2 to 6.3.5), because they are more directly related to a large number of biogeochemical model variables, or they are less prone to analytical and/or difference-in-kind errors. The importance of biogeochemical reanalysis for marine ecosystem assessment and management is pushing the development of new operational reanalysis systems in a number of recent research and development projects (e.g., in Europe, the H2020 "Operational Ecology" project and the Copernicus Marine Environment Monitoring Service, Le Traon et al. 2017) prospecting future relevant progress in this area of synergic application of modelling and ocean colour observation (see also Fennel et al. 2019).

6.4.1 Available software for biogeochemical data assimilation

❖ Data Assimilation Research Testbed (DART): www.image.ucar.edu/DAReS/DART/

❖ Ensemble Kalman Filter (EnKF): http://enkf.nersc.no/

❖ Ensemble Kalman Filter coded in C (EnKF-C): https://github.com/sakov/enkf-c

❖ Employing MPI for Researching Ensembles (EMPIRE): http://www.met.reading.ac.uk/~darc/empire/index.php

❖ Parallel Data Assimilation Framework (PDAF): http://pdaf.awi.de

6.5 Recommendations

The assimilation of ocean colour products into marine biogeochemical models is a relatively new and promising field of research. This approach has the potential to improve the understanding of marine ecosystem functioning, prediction of biogeochemical indicators and simulation of climate-relevant biogeochemical fluxes. Ocean colour assimilation methods have been influenced by the more established fields of physical oceanographic and meteorological assimilation; however biogeochemical
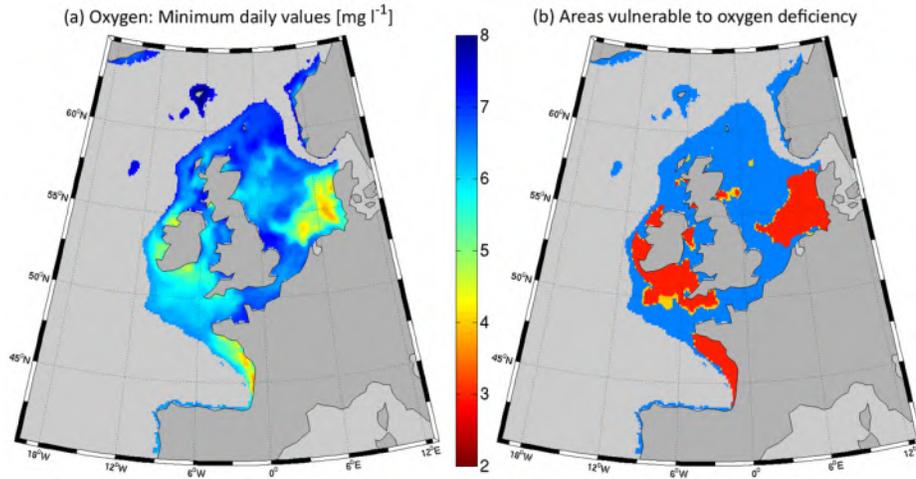
Figure 6.10 Reanalysis of biogeochemical indicators in the North West European shelf (years 1998–2009) by assimilating an ocean colour chlorophyll product with the ensemble Kalman filter (100 ensemble members; Ciavatta et al. 2016): (a) Minimum daily values of dissolved oxygen simulated by the ensemble median at the bottom of the shelf, and (b) map of the areas at risk of oxygen deficiency, i.e., with at least one daily value in 1998–2009 below the threshold of 6 mg l^{-1} (OSPAR 2013). In (b), yellow colour represents deficient areas at the 1% confidence level (i.e., at least one member of the ensemble signals oxygen deficiency), red represents 100% confidence (all one hundred members signal deficiency). Figure from Ciavatta et al. (2016), Creative Commons Attribution License (CC BY 4.0).

data assimilation poses specific challenges, such as high non-linearity, non-Gaussianity and positive-definition of the simulated ecosystem states and parameters, which is being addressed by means of new approaches.

The novel assimilation of ocean-colour products complementary to the traditional total chlorophyll (e.g., remote sensing reflectance, phytoplankton functional types, ocean carbon stocks) demonstrated some advantages and should be explored further. One possible limitation of ocean colour assimilation is that its impact might be limited to the simulated upper layers of the ocean. Though some assimilation methods are capable of propagating the observed information within the water column, the simultaneous assimilation of remotely-sensed and *in situ* biogeochemical profiles (e.g., from gliders and biogeochemical-Argo floats) appears a useful way forward. We expect that the use of ocean colour assimilation will continue to expand in modern operational modelling systems and the increased availability of better resolved, error-characterized, hyperspectral satellite data will further support such applications.

Synergistic Use of Ocean Colour Data and Models to Understand Marine Biogeochemical Processes

Stephanie Henson, Colleen Mouw, Cecile Rousseaux and Jerry Wiggert

---

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

In the early years of model development, satellite ocean colour data was used simply to validate the climatological values, spatial distribution and seasonal variability of chlorophyll or primary production. More recently, however, there has been increasing interest in synthesising the satellite data and model output to provide insight into the driving factors behind the observed seasonal and interannual variability in ocean colour. The increasing number of satellite ocean colour missions and products, combined with the continuing development of complex numerical models, allows for new and exciting multi-disciplinary approaches for tackling ecological and biogeochemical questions. Studies that have combined ocean colour data with models have significantly enhanced our understanding of global-, basin-, and mesoscale time-series and elucidated details of regional processes and phytoplankton physiological states. Furthermore, the scope of these integrated approaches can be extended to encompass, for example, complex food-web interactions such as the recruitment of fish.

7.1   Use of Hindcast Simulations to Explore Processes that Result in Observed Ocean Colour Variability

In hindcast simulations, ocean biogeochemical models are generally forced with observed atmospheric conditions, such as wind speed, heat flux etc. (see also Section 3.4 for description of different model types, and Appendix 3 for terminology). Reanalysis products, which incorporate *in situ* and satellite-derived data, are typically available from 1948 onwards, enabling multi-decadal simulations. Because these models are driven with realistic atmospheric forcing, the simulated seasonal and interannual variability should be directly comparable with the observations. For example, we would expect a hindcast model to simulate a strong El Niño event in 1997/1998, followed by a La Niña in 1998/1999. This permits the use of a hindcast model to investigate processes and mechanisms underlying the seasonal and interannual variability observed in satellite ocean colour data. Often the biogeochemical model provides information on parameters important to chlorophyll variability but difficult to measure *in situ* at the resolution necessary to relate to satellite ocean colour fields, e.g., macronutrients or iron concentration. Provided a model shows the relevant property-property relationships to the observations, and the model's dynamical features are thoroughly understood, meaningful physical and biogeochemical interpretations of the observed variability are possible.

Seasonal variability in chlorophyll concentration is pronounced in many ocean regions and has been well-characterised by satellite ocean colour observations. However, defining the mechanisms

117

driving seasonal and intra-seasonal chlorophyll patterns often requires more information than can be obtained from remote sensing images alone. Synthesis of hindcast simulations and satellite data allows for the investigation of both the processes controlling seasonal variability, and its subsequent impacts. For example, the canonical theory of bloom initiation, Sverdrup's critical depth hypothesis (Sverdrup 1953), has been examined in different biogeochemical provinces of the northern Atlantic (Dutkiewicz et al. 2001; Follows and Dutkiewicz 2002). A model was used to investigate basin-scale patterns of nutrient supply, mixed layer depth and euphotic zone depth, which were then correlated with satellite observations of bloom timing and magnitude. This study demonstrated how the interplay between light, buoyancy provided by heat flux, and wind mixing acted on phytoplankton blooms in both the subpolar and subtropical North Atlantic.

Model output can provide information on potential forcing mechanisms that may be difficult to diagnose on the appropriate time and space scales from satellite data or *in situ* observations alone. For example, the response of chlorophyll to Indian Ocean monsoon variability was investigated using SeaWiFS data and a biogeochemical model which revealed that wind-induced mixing drives the intraseasonal patterns of chlorophyll distribution (Resplandy et al. 2009). The additional advantage of using a model in conjunction with satellite data is that the consequences of variability can also be investigated. In the case of Resplandy et al. (2009), the model allowed investigation of the vertical distribution of chlorophyll, with an increase in surface production in response to wind-induced mixing episodes compensated by a decrease in subsurface production. Consequently, there was no change in carbon export. Similarly, Bennington et al. (2009) explored the influence of seasonal variability in satellite-derived chlorophyll concentration on North Atlantic air-sea CO_2 flux, concluding that phytoplankton bloom timing drives summer variability in pCO_2.

Processes acting on interannual timescales to alter ocean productivity are also readily investigated through synthesis of hindcast models and ocean colour data. For example, Fauchereau et al. (2011) noted that in the Southern Ocean, the correlation between mixed layer depth, measured from Argo floats, and satellite chlorophyll could be either positive or negative and was surprisingly heterogeneous. Use of a biogeochemical model allowed Fauchereau et al. (2011) to investigate chlorophyll responses to transient mixing events and subsequent alteration of iron and light availability, and thus elucidate the mechanisms driving the observed variability. In a similar fashion, Santoleri et al. (2003) applied a coupled physical-biogeochemical model to the Adriatic Sea to investigate whether changes in convective mixing, and thus nutrient supply, could explain the observed interannual variability in chlorophyll. A series of sensitivity analyses with the model allowed them to pinpoint interannual variability in water masses, in particular the East Mediterranean Transient, and subsurface nutrient concentration as the primary control on chlorophyll. A sensitivity analysis was also used by Echevin et al. (2008) to investigate the controls on observed seasonal and interannual variability of chlorophyll in the Peruvian upwelling. In this case, the authors concluded that mixed layer depth was the most important factor, whilst surface irradiance and SST had little impact on chlorophyll variability. The role of nutrient supply was also implicated in a study of the Arabian Sea that found the winter bloom amplitude derived from SeaWiFS data to be driven by interannual variability in the mixed layer depth and heat flux (Keerthi et al. 2016).

In the following case studies, we explore how biogeochemical models have been used in combination with satellite ocean colour data to explore the drivers of variability in phytoplankton populations over interannual to decadal timescales.

7.1.1 Case Study: Elucidating the nutrient supply routes controlling primary production

The eastern boundary upwelling system of the subtropical North Atlantic is the most variable of the four major eastern boundary currents and is associated with high annual primary production (Carr 2002). The size of the upwelling region varies interannually, likely driven by variability in the large-scale

wind stress field (Pelegri et al. 2006), which is hypothesised to alter the supply route of nutrients to the region through changes in stratification and/or Ekman transport. However, testing which nutrient supply pathway dominates the region and consequently the links to decadal-scale oscillations cannot be achieved via satellite data alone. Pastor et al. (2013) were able to tackle this issue through combined use of SeaWiFS chlorophyll concentration and biogeochemical model output. This allowed them to assess the drivers of chlorophyll interannual variability in the region by quantifying all terms in the euphotic zone nutrient budget.

7.1.1.1 Approach

The biogeochemical model used was NOAA-GFDL's TOPAZ (Dunne et al. 2013), integrated with the MOM-4 physical model (Gnanadesikan et al. 2006). The TOPAZ model includes four nutrient elements (N, P, Si, Fe), dissolved organic pools, variable chlorophyll-to-carbon ratio, and light and nutrient co-limitation. Three classes of phytoplankton are represented: 'large' (diatoms and other eukaryotes), 'small' (cyanobacteria and picoeukaryotes) and 'diazotroph' (nitrogen fixers). The phytoplankton are grazed by a single zooplankton class. The model was forced with reanalysis meteorological fields, i.e., a hindcast simulation, so that the model ocean is responding to meteorological conditions with the correct timing (the term 'reanalysis' in this context refers to the assimilation of historical observations over an extended period into a numerical model to generate spatially and temporally varying fields). The model was validated by assessing its ability to reproduce the phenomenon of interest, in this case the observed spatial and temporal patterns of chlorophyll variability.

7.1.1.2 Findings

The authors were able to exploit the model output to investigate the relative importance of the different mechanisms of nutrient supply in the region. Vertical advection was found to dominate over diffusivity in driving interannual variability in chlorophyll concentration (Figure 7.1). On interannual to decadal timescales, the upwelling of nutrients was found to vary in association with changing vertical velocity, rather than the subsurface nutrient reservoir. Finally, the authors found no link between decadal scale variability in the size of the high chlorophyll region and the El Niño-Southern Oscillation (ENSO), as previously suggested, but did find a weak relationship with the Atlantic Meridional Mode, although even the 50 year model run was insufficiently long to capture more than one cycle of the oscillation.
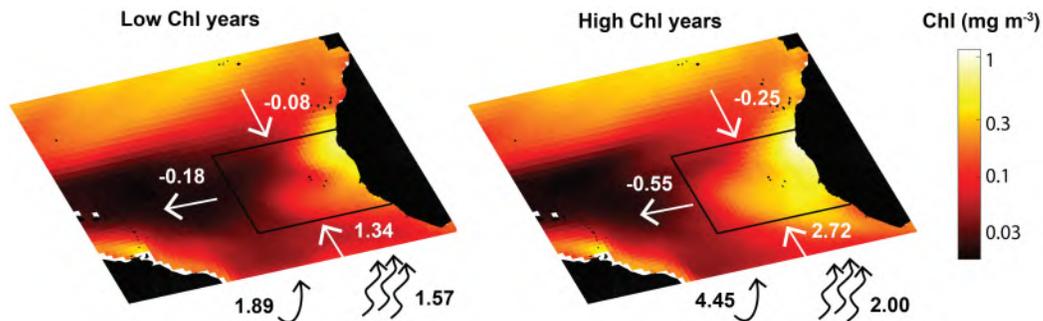
Figure 7.1 Model results of nutrient supply to the upwelling region of the northeast Atlantic in contrasting high and low chlorophyll years. Solid arrows show the horizontal and vertical advective supply and curly arrows show the mixing term (all terms in 10^{-10} mol m^{-2} s^{-1} (redrawn using data from Pastor et al. 2013).

7.1.2   Case Study: Drivers of trends in phytoplankton community composition

Identifying major trends in the biogeochemical composition of the oceans is essential to improve our understanding of biological responses to climate forcing. Previous studies have assessed the existence of trends at the global scale using ocean colour (e.g., Gregg et al. 2005; Henson et al. 2010; Vantrepotte and Mélin 2011; Beaulieu et al. 2013; Siegel et al. 2013). While these efforts have provided a first line of information on the existence of trends in ocean biology, little is known about global and large-scale regional trends, or multi-annual to decadal variability in phytoplankton composition and the processes leading to these trends.

7.1.2.1   Approach

Rousseaux and Gregg (2015) used the NASA Ocean Biogeochemical Model (NOBM) to investigate the decline in chlorophyll concentration in the high northern latitudes over the period 1998–2012. The NOBM includes four nutrients (silicate, nitrate, ammonium and iron), four phytoplankton groups (diatoms, cyanobacteria, chlorophytes and coccolithophores), variable chlorophyll:carbon ratios, particulate and dissolved carbon pools, and is coupled to a global general circulation model (Poseidon) and a radiative transfer model (Ocean-Atmosphere Spectral Irradiance Model, OASIM; Gregg and Carder 1990; Gregg 2002b; Gregg and Casey 2009). The NOBM assimilates satellite ocean chlorophyll from SeaWiFS, MODIS-Aqua and VIIRS and uses the Modern-Era Retrospective analysis for Research and Applications (MERRA) data to force the circulation model.

7.1.2.2   Findings

Using the NOBM, Rousseaux and Gregg (2015) found that the previously observed declining trends in phytoplankton in the northern high latitudes (Gregg and Rousseaux 2014) were the result of a significant decline in diatoms (-1.22% yr^{-1}). This decline coincided with a decline in nitrate (-0.38% yr^{-1}), a shallowing of the mixed layer depth (MLD) of -0.20% yr^{-1} and a significant increase in surface PAR of 0.09% yr^{-1}. These trends illustrate the diversity and complexity of mechanisms that drive phytoplankton community response to variable conditions of nutrients, light, and mixed layer depth (Figure 7.2). This type of study provides a first insight into the existence of trends in phytoplankton composition over the maturing satellite ocean colour era and illustrates how changes in the conditions of the oceans in the last 15 years may have affected phytoplankton populations. Whether the trends are part of a longer-term climate change response, or a response to interannual/decadal variability, is unclear given the relatively short ocean colour time series. This issue is discussed in more detail in Section 8.2.

7.1.3   Case Study: Investigating the potential for iron limitation of primary production

The geomorphology of the northern portion of the Indian Ocean basin leads directly to the monsoon cycle which is the leading order control on the region's dynamics, including its unique seasonally reversing boundary currents (Hood et al. 2017). In addition, these arid terrestrial regions serve as a source of wind blown dust that have led to the Arabian Sea being considered as "Mother Nature's Iron Experiment" (Smith 2001). Within this context, climatological forcing was applied to a coupled physical-biogeochemical model of the Indian Ocean that included iron as a micronutrient. Two results, relating to the spatio-temporal distribution patterns of iron limited phytoplankton growth, are highlighted here to demonstrate the power and utility of combining modelling and remote sensing to advance mechanistic understanding of ocean biogeochemical processes.
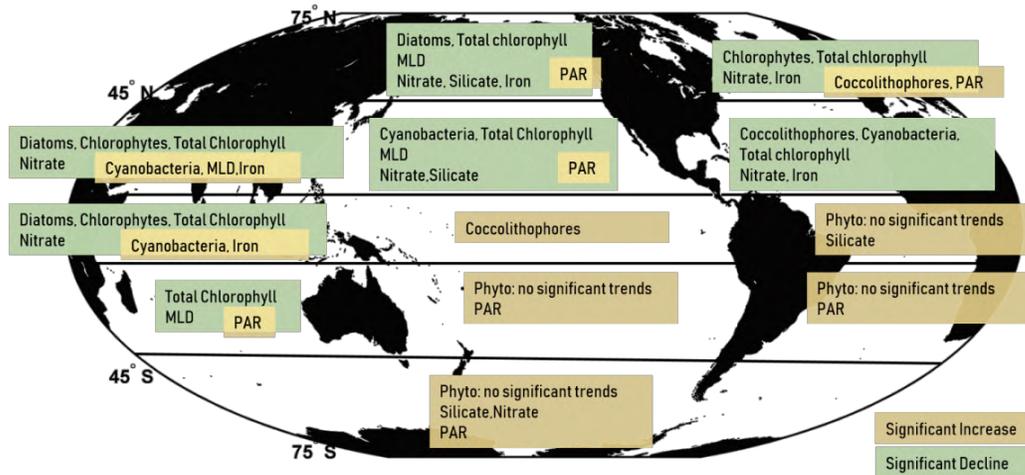
Figure 7.2 Model derived statistically significant trends (1998–2012) in various biological and biogeochemical properties in 12 oceanographic regions. Redrawn from Rousseaux and Gregg (2015), with permission from John Wiley and Sons.

7.1.3.1 Approach

Following the JGOFS Arabian Sea Expedition, a coupled physical-biogeochemical model was used to explore how to integrate the available observations and extrapolate to broader spatial and seasonal scales, with the goal of gaining deeper insight into the basin's biophysical variability and biogeochemical cycling (Wiggert et al. 2006; Wiggert and Murtugudde 2007). The biogeochemical model was originally developed and applied to study iron cycling in the Pacific Ocean (Christian et al. 2002). The model was run with climatological forcing, with the objective of illuminating seasonal variability in this poorly sampled ocean domain.

7.1.3.2 Findings

A surprising result concerning the seasonal variability of the Arabian Sea was that, despite the elevated dust fluxes known to impact the region, the model indicated that the coastal upwelled waters propagating offshore of the Arabian Peninsula during the Southwest Monsoon were prone to iron limitation. Within the model, this reflects the N:Fe ratio of the waters being upwelled into the euphotic zone by the monsoonal forcing. The data obtained at the station nearest to shore during the US JGOFS Southwest Monsoon cruise established that the N:Fe of these waters in the upper 50 m was greater than 15,000, which suggests this area is prone to iron limitation (Measures and Vink 1999). In addition to this unexpected manifestation of iron limitation in the Arabian Sea upwelling region, the model indicated that a broad swathe of the southwest Indian Ocean over the Seychelles-Chagos Thermocline Ridge is consistently iron limited (Figure 7.3), which is to be expected given its distance from terrestrial source regions. Support for iron limitation in both of these regions was provided through a satellite-based ocean colour algorithm that used the fluorescence line height product available from the MODIS-Aqua sensor to obtain distributions of quantum yield of fluorescence (ϕ_{sat}), where elevated ϕ_{sat} is indicative of phytoplankton subject to iron stress (Behrenfeld et al. 2009). The spatial correspondence between iron-limited growth suggested by the model and iron-stressed phytoplankton apparent in satellite-derived ϕ_{sat} is clear, which gives confidence in the iron limitation variability in the model, which is not subject to the cloudiness issues that are a particular challenge for the Arabian Sea during the Southwest Monsoon. Subsequent fieldwork seeking to confirm whether iron limitation

occurs in the Arabian Sea has indeed confirmed this, with the majority of stations exhibiting indications of iron limitation during the latter stages of the 2007 Southwest Monsoon (Moffett et al. 2015).
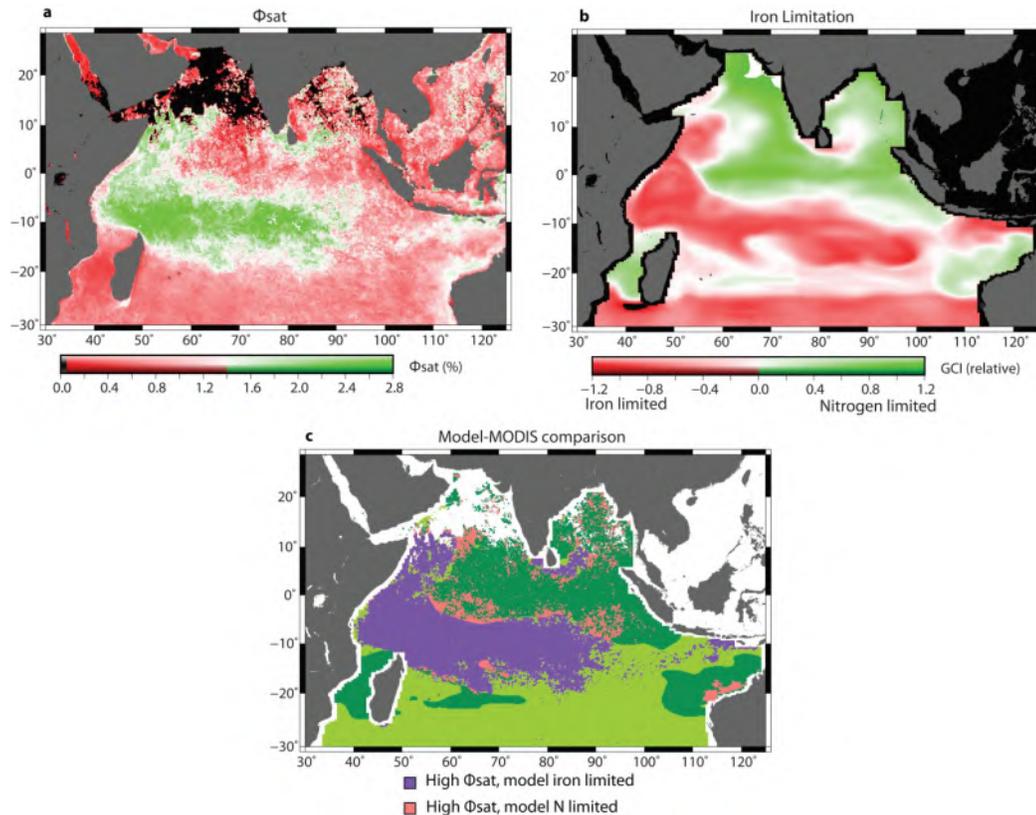
Figure 7.3 Indian Ocean (a)ϕ_{sat} and (b) model-based growth constraint index (GCI) (Wiggert et al., 2006) for boreal summer (June–August). (c) State-space comparison of ϕ_{sat} and GCI. Credit: redrawn from Behrenfeld et al. (2009), Creative Commons Attribution 3.0 Unported (CC BY 3.0).

7.2 Using Hindcast Models to Extend Satellite Data into the Recent Past to Explore Decadal Variability

Biogeochemical models run in hindcast mode have an important additional function, other than examining the forcing controlling observed variability in the satellite chlorophyll record. Because forcing fields from reanalysis efforts (e.g., NCEP/NCAR) stretch back to 1948 in many cases, hindcast models can be used to extend the satellite record back in time to examine decadal variability. For example, the North Atlantic Oscillation (NAO) has generally been in a positive or neutral phase for much of the satellite ocean colour record (the exception was 2009–2011 which saw strong negative conditions). Thus investigating the phytoplankton response to decadal variability associated with the NAO has been limited by the relatively short satellite data time series. Biogeochemical models are a powerful tool in such a situation, provided of course that they can reproduce the phenomenon of interest as observed in the satellite data. Henson et al. (2009), for example, used a biogeochemical model to examine the effect of decadal variability in physical conditions on the timing of the phytoplankton bloom in the

North Atlantic. The model was first shown to reproduce the observed bloom timing and interannual variability, before the differences in historic positive and negative NAO phases were examined. The authors concluded that increased mixing in subpolar regions in positive NAO periods delayed the onset of the spring bloom. Further insights into the processes controlling phytoplankton blooms on a decadal timescale were gained by Patara et al. (2011) who constructed means over NAO positive and negative periods of chlorophyll anomalies and seasonal cycles using a hindcast model. Their detailed study of the instantaneous and lagged responses to shifts in the NAO revealed how advection of subsurface nutrient anomalies in the North Atlantic Current altered the subpolar chlorophyll bloom.

The El Niño-Southern Oscillation (ENSO) is perhaps the most widely studied ocean phenomenon, and here too the combination of biogeochemical models and satellite data provides new insights into the processes at work. By extending their model run back to 1993, Radenac et al. (2001) were able to investigate the chlorophyll response to ENSO prior to the 1997–1999 event observed by SeaWiFS. The studied interplay of advective and vertical supply of nutrients throughout the Equatorial Pacific demonstrated the relevance of conditions prior to the satellite observations in setting the chlorophyll response. A close inspection of the SeaWiFS chlorophyll record undertaken by Ryan et al. (2006) revealed that blooms occurred at the termination of El Niño events. The authors then used a biogeochemical model to investigate the changes in subsurface currents associated with the end of an El Niño and discovered that the iron-rich New Guinea Counter Current was intensified. This resulted in increased iron transport from the New Guinea margin into the equatorial undercurrent which relieved iron limitation.

The changes in nutrient availability associated with ENSO events are hypothesised to alter the phytoplankton community structure. Community structure is extremely challenging to measure *in situ* at spatial and temporal scales large enough to investigate basin-wide phenomena such as ENSO, which made validating results from early modelling studies difficult. Nevertheless, Rousseaux and Gregg (2012) used a data-assimilating model to demonstrate that the ENSO-driven changes in chlorophyll observed in the satellite data arise from a phytoplankton composition shift during the 1997/98 El Niño. The low nutrient conditions during El Niño conditions favoured an increase in cyanobacteria concentration while diatoms plummeted. Recent advances in satellite biological oceanography have resulted in several algorithms to estimate phytoplankton community structure remotely (e.g., IOCCG 2014). This allowed Masotti et al. (2011) to confirm that modelled changes in community composition in the Equatorial Pacific during El Niño/La Niña transitions resemble those estimated from satellite ocean colour data. Again, the use of a biogeochemical model allowed the authors to investigate the environmental variables controlling the interannual shift in phytoplankton composition.

As yet, limited use has been made of hindcast models to investigate natural climate oscillations other than ENSO or the NAO. An exception to this relates to the impact of the Indian Ocean Dipole (IOD) on chlorophyll distribution patterns throughout the Indian Ocean basin. Wiggert et al. (2009) contrasted the impact of two recent IOD events (1997/1998 and 2006/2007) in a detailed satellite and *in situ* data exploration of the spatio-temporal patterns of surface chlorophyll, and the associated physical drivers (surface winds and planetary waves). The principal ocean colour signature of the IOD is extensive phytoplankton blooms extending westward off Java and Sumatra that occur in boreal autumn. A 40-year hindcast of a coupled biophysical model has provided more comprehensive insight into the impact of the IOD on Indian Ocean bloom dynamics (Currie et al. 2013). By extending the ocean colour time series with model output, Currie et al. (2013) demonstrated that the most pronounced anomalous blooms associated with the IOD occur in coastal and equatorial upwelling regions of the eastern Indian Ocean where upward thermocline displacement occurs in response to atypical upwelling favourable winds. Through their modelling effort, Currie et al. (2013) also demonstrated that negative, off-equator chlorophyll anomalies in the western Indian Ocean during boreal winter were the result of atypical planetary wave dynamics.

Extending the satellite ocean colour record back in time using biogeochemical models was also instrumental in determining the driving factors of changes that have been observed since the advent of

ocean colour. Several studies on the expansion of oligotrophic gyres for example concluded, on the basis of 9–10 years of SeaWiFS chlorophyll data, that the size of the gyres was increasing (Polovina et al. 2008; Irwin and Oliver 2009). By examining variability in gyre size from the 1950's onwards using a hindcast model, Henson et al. (2010) showed that the recent apparent expansion of the gyres was more likely due to decadal variability associated with ENSO. Other studies have used ocean colour data in combination with numerical models to assess the driving factors of trends in the oceans. For example, Rousseaux and Gregg (2015) found that the decline in total phytoplankton observed in the northern latitudes could be attributed to a shallowing of the mixed layer and a decrease in mixed layer nutrient concentration that coincided with a decline in diatoms. This example demonstrates the difficulty of detecting long-term trends with a dataset that is substantially shorter than the natural time scales of variability, a theme we will return to in the next chapter, where the potential for synthesising satellite data and model output for detecting biogeochemical response to global climate change is examined.

7.2.1 Case Study: Exploring the response of the North Atlantic bloom to decadal variability

In the North Atlantic Ocean, a major climate oscillation mode, the North Atlantic Oscillation (NAO), controls westerly wind strength and winter storm tracks (Hurrell et al. 2003). A biological response to the NAO has been posited on the basis of long-term plankton recorder datasets (e.g., Barton et al. 2003), however the underlying mechanisms and spatial extent of the response cannot be easily quantified due to data restrictions. Also, the ocean colour satellite record is not sufficiently long to investigate decadal-scale variability. The decadal variability in the physical conditions driving the timing of the phytoplankton bloom in the North Atlantic were investigated in Henson et al. (2009) using a biogeochemical model.

7.2.1.1 Approach

Henson et al. (2009) overcame the temporal limitations of the satellite ocean colour data record by using a biogeochemical model to investigate the satellite-era phytoplankton bloom seasonality back to the late 1950's. The biogeochemical model used was NOAA-GFDL's TOPAZ (Dunne et al. 2013) in the same configuration as described in Case Study 7.1.1. The model was run in hindcast mode so that the phasing of the NAO variability in the model corresponds directly to the real world. The model was first shown to reproduce the phenomenon of interest, i.e., the satellite-derived bloom timing and its interannual variability, which gave confidence that the model could be used to explore the decadal variability in phenology.

7.2.1.2 Findings

By comparing the modelled bloom timing and spatial patterns in periods of historic positive and negative NAO phases, the basin-scale response to the NAO could be explored (Figure 7.4). The onset of the spring bloom was found to be delayed in positive NAO periods with respect to negative NAO conditions. The use of the model further allowed the mechanism underlying this contrast to be investigated. The strong westerly winds across the subpolar gyre associated with positive NAO periods drove deeper mixed layers in the model, which had the effect of delaying the start of the spring bloom by 2–3 weeks in this light limited region. On the other hand, the timing of the subtropical bloom was found not to correlate with the NAO index.
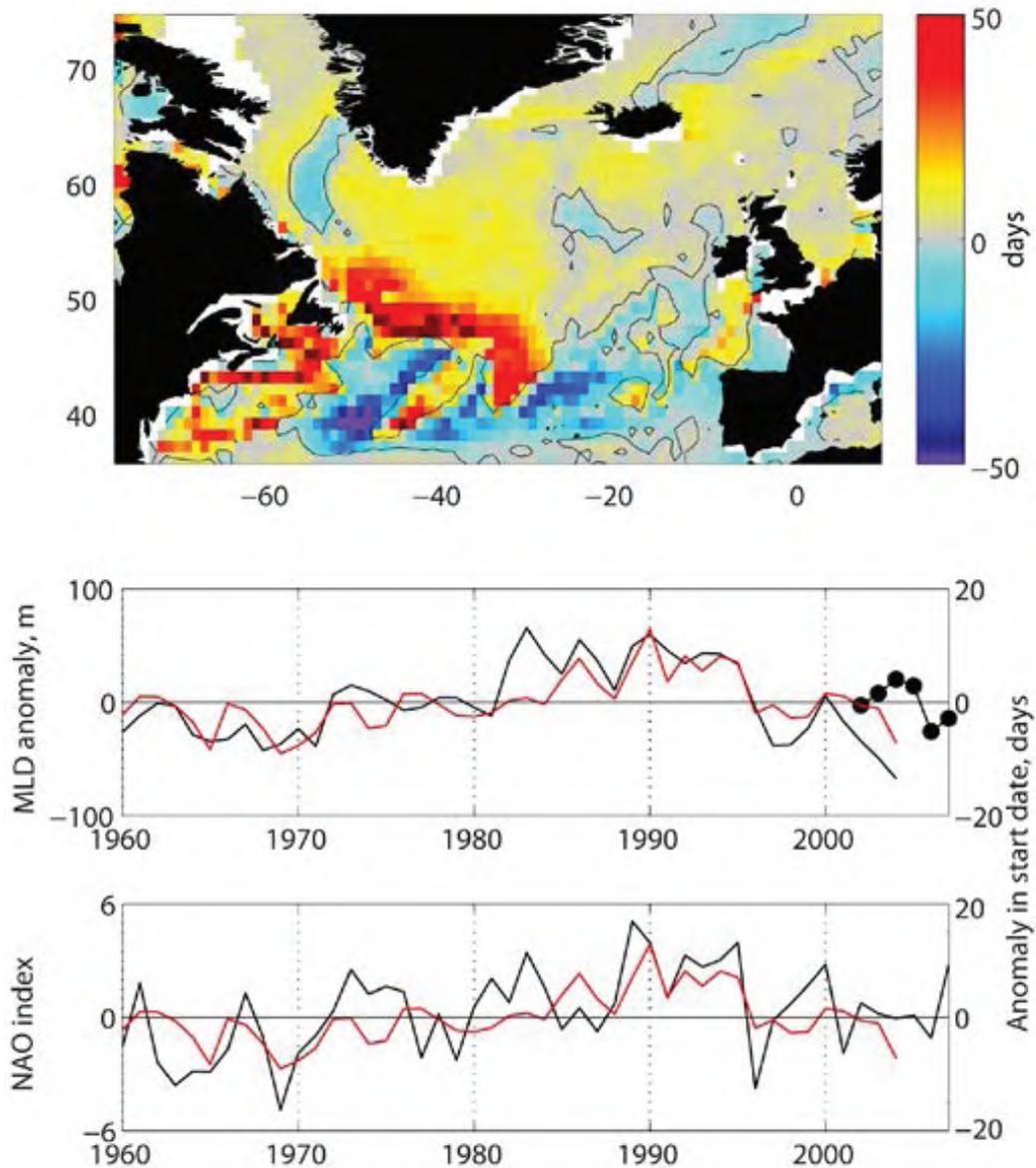
Figure 7.4 Top panel: difference in modelled bloom start date between the long-term mean (1960–2004) and a positive NAO phase (1988–1995), where positive values indicate a later bloom start in positive NAO phases. Middle panel: Time series of modelled subpolar bloom start date (red line), annual mean modelled MLD (black line) and Argo float-derived MLD (black dots). Bottom panel: Time series of modelled subpolar bloom start date (red line) and the NAO index (black line)(redrawn from Henson et al. 2009 with permission from John Wiley and Sons.)

7.2.2 Case Study: Investigating the mechanisms of decadal variability in North Atlantic biomass

The distribution of phytoplankton biomass in the North Atlantic is affected by both vertical processes (modification of nutrient or light limitation) and horizontal processes (Ekman processes). The response

of phytoplankton to horizontal processes is difficult to ascertain without recourse to a biogeochemical model due to the large scale nature of the forcing. The influence of changing ocean circulation, geostrophic advection and induction (large scale movement of nutrients from the permanent thermocline into the mixed layer) on upper ocean nutrient supply is challenging to assess using observational datasets alone. A regional model was used by McKinley et al. (2018) to explore how changes in both vertical and horizontal nutrient supply may have driven the satellite-observed variability in chlorophyll concentration. McKinley et al. (2018) then investigated the response on decadal timescales by analysing the model output over 1949–2009.

7.2.2.1 Approach

The MIT-GCM (Massachusetts Institute of Technology General Circulation Model) in its North Atlantic configuration was used (Marshall et al. 1997a,b). The model was forced with reanalysis meteorological and radiative fields from 1948–2009. The biogeochemical module contains a phosphorus-based system with two phytoplankton classes ("large" diatoms and "small" phytoplankton) and one zooplankton class (Dutkiewicz et al. 2005). The biogeochemical model incorporates phosphorus, silica and iron, as well as carbonate chemistry. In previous work, the model had been demonstrated to reproduce the timing and magnitude of the subpolar bloom as captured by SeaWiFS (Bennington et al. 2009).

7.2.2.2 Findings

Over the period of SeaWiFS observations (1998–2007), phytoplankton biomass (model of Westberry et al. 2008) increased in the northwest subpolar gyre and decreased east of 30°W. Using the 3D fields provided by the biogeochemical model, McKinley et al. (2018) were able to determine that a weakening of the subpolar gyre and associated shoaling of the mixed layer (which relieved light limitation) led to the observed increases in biomass in the northwest of the region. The biomass decline in the east of the basin was attributed to reduced horizontal convergence of phosphate, due to reduced vertical phosphate supply as a consequence of suppressed deep winter mixing west of 35°W. The use of a hindcast model run allowed the authors to extend their analysis to decadal timescale (1949–2009). Variability in biomass in the northeast of the subtropical gyre was ascribed to variability in both horizontal and vertical phosphate supply, whereas in the subpolar gyre horizontal fluxes dominated.

7.3 Caveats and Recommendations

Use of biogeochemical models in combination with satellite-derived ocean colour data has provided a wealth of insight on the underlying mechanisms controlling observed phytoplankton variability in several regions of the global ocean. Hindcast runs of biogeochemical models have also proven useful to evaluate phytoplankton variability on interannual to decadal timescales occurring prior to the start of the ocean colour satellite data record. However, there are many areas of research in which biogeochemical models could be used to inform ocean colour studies which have not yet been explored. One example is to use model output to extend satellite data sub-surface. For example, a biogeochemical model could be used to explore how much of the variability in chlorophyll concentration in the subtropical ocean is due to changes in the depth of the sub-surface chlorophyll maximum.

However, the use of biogeochemical models to study variability in phytoplankton populations does have its challenges. Clearly, the model must be able to reproduce the observed variability. This is actually more challenging than might be expected. In a comparison of three state-of-the-art climate models, Schneider et al. (2008) found that only one of them was able to reproduce the observed variability in subtropical primary production. At a regional scale, the models may fare even worse: a comparison of 14 biogeochemical models to *in situ* primary production measured at BATS and HOTS

showed that almost all the models underestimated the observed mean, variance and trends (Saba et al. 2010). Similarly, a comparison of nine models to a database of *in situ* primary production measured in the tropical Pacific demonstrated consistent underestimation of the observed variability (Friedrichs et al. 2009). Even when a model is able to adequately represent observed interannual variability there may be features of the longer term record, such as regime shifts or trends, that the model does not capture (Friedrichs et al. 2009).

The fact that a biogeochemical model may not necessarily capture the observed variability may arise because the model is less complex than reality. This can be both a boon and a disadvantage. On the positive side, the relative simplicity of the typical biogeochemical model allows the processes underlying the variability in phytoplankton populations to be readily diagnosed. On the negative side, the model may be missing processes which are fundamental to driving the observed variability.

Additionally, the biogeochemical model needs sufficient spatial and temporal resolution to match the real-world phenomenon of interest. As many biogeochemical models (particularly those projecting future climate scenarios) are ~1° spatial resolution and 1 month temporal resolution, this limits the underlying processes or variability that can be diagnosed. Typically, the combination of ocean colour and models is most suited to exploration of basin-scale variability over interannual to decadal timescales, although higher resolution models (both spatially and temporally) have been used to explore, for example, phenology (Henson et al. 2018) and mesoscale eddies (Levy et al. 2014).

Models are not, and can never be, a perfect reproduction of reality. Nevertheless, the adage "all models are wrong, but some models are useful" (Box 1976) applies equally well to the combined use of ocean colour and models. Models can be an extremely useful additional tool in the kit as outlined in this chapter, when applied bearing the above caveats in mind.

Using Models to Inform Ocean Colour Science

Stephanie Dutkiewicz, Stephanie Henson, Colleen Mouw and Cecile Rousseaux

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

Numerical models can be used as laboratories to help understand some of the limitations and uncertainties of ocean colour products, help understand future needs of ocean colour missions, and assist in algorithm development. Numerical models can extend beyond the depth, temporal and spatial scales that satellites can observe. Models can be subsampled similarly to observations to investigate issues with missing data. In this chapter, we review studies that have used models to inform interpretation of ocean colour output and plan for the future. Just as ocean colour products are important for models, we show that models can also be useful for ocean colour science.

8.1 Exploring the Consequences of Missing Data

Satellite ocean colour data are limited by clouds (and associated stray light), thick aerosols, inter-orbit gaps, sun glint, sensor tilt changes and low light in polar regions (Figure 8.1). These data gaps can severely limit the availability of daily data in some regions of the oceans. Furthermore, high latitudes and regions with frequent clouds and aerosols can have persistent data gaps, even at the monthly timescale. Here we provide two case studies to explore the impact of missing data in terms of calculating means (or "climatologies", see Chapter 4) and in determining phytoplankton phenology.

### 8.1.1 Case Study: Effect of missing data on annual and monthly means

Polar orbiting spectroradiometers (e.g., MODIS, SeaWiFS, MERIS, OLCI, and VIIRS) are assumed to provide global representation of ocean colour. However, to achieve true global representation, the data have to be merged over several weeks or even months and there can be considerable gaps in the data (Figure 8.2). By averaging the available data over a month, the spatial coverage is improved, except for regions where persistent clouds and aerosols exist, or high latitude regions in winter. The effects that these sampling irregularities have on the monthly means and whether those irregularities in sampling produce biases in the global and regional chlorophyll means was investigated by Gregg and Casey (2007b).

8.1.1.1 Approach

Gregg and Casey (2007b) quantified the bias due to missing data using the satellite chlorophyll assimilating NASA Ocean Biogeochemical Model (NOBM). This three-dimensional global biogeochemical model includes a circulation model, radiative transfer model and a biological model. The biological model includes multiple nutrients, phytoplankton groups and carbonate chemistry components within a dynamical representation of the global oceans. This configuration is used to create a "truth-field" that
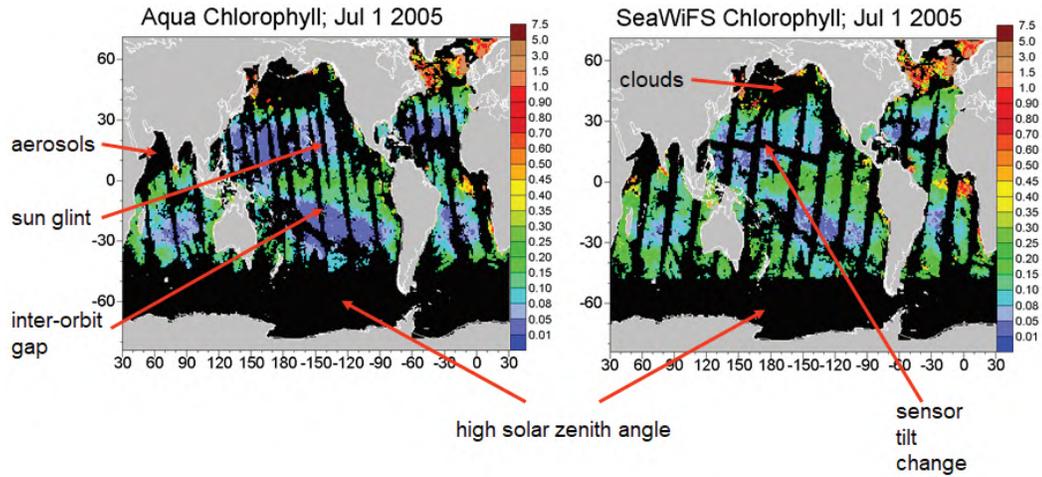
Figure 8.1 Daily ocean coverage by MODIS-Aqua and SeaWiFS, with sources causing data gaps identified. Reproduced from Gregg and Casey (2007b) with permission from Elsevier.
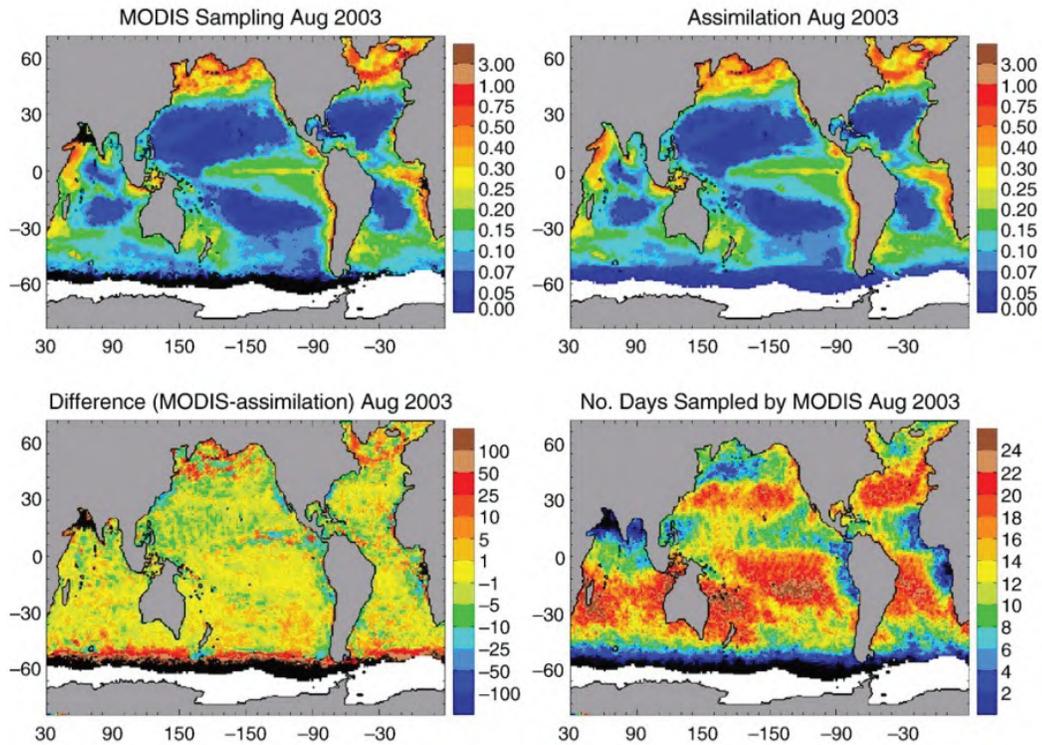
Figure 8.2 Top left: chlorophyll concentration (mg m^{-3}) from assimilation model sampled by MODIS-Aqua (white indicates ice, black indicates missing data). Top right: chlorophyll concentration (mg m^{-3}) from a run assimilating MODIS data for August 2003. Bottom left: percent difference MODIS sampling – assimilation. Bottom right: number of days sampled by MODIS. Reproduced from Gregg and Casey (2007b) with permission from Elsevier.

is sub-sampled as though a satellite were making the observations (i.e., using the orbit and scanning characteristics of the corresponding ocean colour satellite, as well as the lack of data from clouds etc.). Some regions have considerable lack of coverage on a monthly basis (see Figure 8.1). Mean error statistics are then calculated from the model "truth" and the subsampled results.

8.1.1.2 Findings

This study found that global annual mean biases due to missing data were ~8% (MODIS) and >~6% (SeaWiFS). Biases were much larger at high latitudes (20% as basin annual mean and over 80% in some months) as a result of the low sun angle. These biases at high latitudes vary monthly leading to artifacts in the seasonal cycle. Clouds are the second largest source of sampling error. The North Pacific is the region most impacted by clouds, which can lead to biases of between 6 and 13% between May and July. After clouds, aerosols are the next contributor to sampling error but their effects are relatively localized in both time and space. The effect is most pronounced in the North Indian ocean, with satellite ocean colour having a negative bias of ~-30%, driven by data gaps occurring during the Southwest Monsoon when the chlorophyll concentrations are at their highest. The biases in chlorophyll concentration as well as the seasonal and interannual variability resulting from this irregular sampling therefore needs to be taken into consideration when using ocean colour products. The use of satellite ocean colour assimilation in biogeochemical models provides a potential solution to these biases by providing consistent, global data-constrained, estimates of chlorophyll concentration.

8.1.2 Case Study: How do gaps in satellite data affect phenology studies?

Satellite ocean colour data is ideally suited to estimating phytoplankton phenology — the study of the timing of regular seasonally occurring events. The high temporal resolution of the data is essential to capturing transient events such as the onset of the spring bloom, while the spatial coverage allows large-scale patterns to be discerned. However, missing data may impact the estimation of phenology as gaps could occur during key events. For example, the North Atlantic, site of one of the largest phytoplankton blooms on Earth, can have missing data due to cloud or low sun angle (i.e., too little light) for > 50% of the time.

8.1.2.1 Approach

The study of Cole et al. (2012) used the NOBM (described in case study 8.1.1). The assimilation of SeaWiFS data into the NOBM ensures that it reproduces the seasonal cycle of chlorophyll concentration closely, making it useful for investigating the influence of missing data on phenological metrics. Cole et al. (2012) first verified that the modelled chlorophyll time series reproduced the phenomenon of interest — in this case, the timing of the spring phytoplankton bloom and of peak chlorophyll concentration. The bloom start and peak timing were initially calculated from the gap-free model output. Then, gaps were introduced into the model time series on the same date and location as found in the SeaWiFS dataset. Bloom start and peak timing were calculated again using the time series with gaps, and the two estimates of bloom metrics (with and without gaps) were compared. This allowed the uncertainty introduced by gaps in the satellite record to be quantified.

8.1.2.2 Findings

Cole et al. (2012) concluded that gaps in time series from sub-polar regions result in uncertainty on the bloom initiation date. The uncertainty in predicting the bloom initiation was ~30 days, and ~15 days for predicting the bloom peak (Figure 8.3). In subtropical regions, gaps are less frequent and as a result errors are smaller, ~10 days for bloom initiation and ~5 days for bloom peak. In addition,

in subpolar regions the error introduced by gaps in the data was likely to result in later estimates of bloom timing than the true date. Cole et al. (2012) were also able to derive an empirical relationship between the percentage of missing data at a particular location and the error in phenological metrics. This information is valuable in studies of bloom phenology using satellite ocean colour data, as the error introduced into the estimates of bloom timing or peak can be quantified if the proportion of missing data for a location is known.
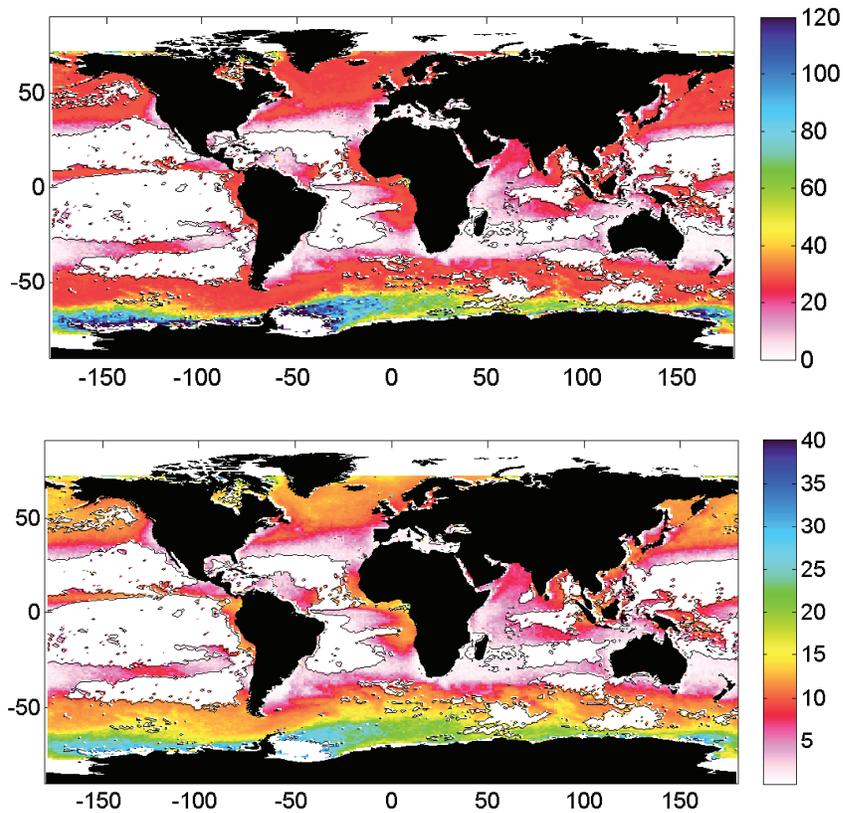
Figure 8.3 Maps of (top) uncertainty in bloom initiation date and (bottom) bloom peak date due to gaps in the satellite ocean colour time series (both in days). White areas do not have a distinct seasonal chlorophyll cycle (and therefore bloom initiation cannot be reliably calculated). Figure redrawn from Cole et al. (2012).

8.2   Use of Models to Inform on Ocean Colour Signals and Products

By resolving variables that are similar to ocean colour measurements (e.g., reflectance), models can also be used to help explore uncertainties in ocean colour products and potentially aid in algorithm development. The early efforts to investigate coupling physical/ecosystem/optical models (Bissett et al. 1999a,b; Fujii et al. 2007; Gregg and Casey 2007a) demonstrated a better fit with observations and more closely reproduced biogeochemical processes than those that did not include optics. The first study to use numerical output to investigate uncertainties in ocean colour products was Mouw et al. (2012), which used offline diagnostics (i.e., after the model has been run) to calculate ocean colour-like products such as reflectance. The study isolated the effects of chlorophyll concentration, phytoplankton cell size, and size-varying absorption on remotely sensed reflectance (see Case Study 8.2.1).

However, it is only recently that models have included treatment of light at a level of detail which allows for online diagnostics (i.e., that are output while the model is running), such as remotely sensed reflectance (e.g., Dutkiewicz et al. 2015a; Baird et al. 2016; Gregg and Rousseaux 2017; Dutkiewicz et al. 2018). In such fully 3-D physical, optical and ecosystem coupled models, optical properties vary with optically active state variables (such as Chl-a, CDOM, detrital matter). This allows for the in-water light field to be accurately represented and permits the model to capture critical feedbacks between the light field, phytoplankton and biogeochemical parameters. Further, the coupled model allows estimation of remote sensing reflectance, enabling a direct comparison between model outputs and satellite products. Such models can be used to explore uncertainties in ocean colour products and algorithms. Dutkiewicz et al. (2018) provided a proof-of-concept study on using model output to explore an often used Chl-a algorithm (Case Study 8.2.2). It is recommended that the more sophisticated models become in their treatment of satellite-like products, the more they should be used to quantify ocean colour uncertainty.

8.2.1 Case Study: Contribution of phytoplankton functional types to ocean colour product uncertainty

The determination of phytoplankton community structure using satellite remote sensing has evolved to a highly active area of research with varied discrimination approaches and output products (see Chapter 2, IOCCG 2014; Mouw et al. 2017). Mouw et al. (2012) used model output to determine the relative contribution of phytoplankton cell size and chlorophyll to overall spectral remote sensing reflectance. Further, the study looked at the impact of cell size on remotely sensed reflectance and how that affected the uncertainty in chlorophyll estimated by band ratio algorithms (i.e., OC4, O'Reilly et al. 1998).

8.2.1.1 Approach

Mouw et al. (2012) used optical and radiative transfer models linked offline to a model that couples the Community Climate System Model (CCSM-3) with the ocean Biogeochemical Elemental Cycling (BEC) model (Doney et al. 2009). The CCSM-3 BEC model output includes three phytoplankton groups (a small fraction consisting of pico/nanoplankton (<20 μm), diatoms (considered to be all microplankton, > 20 μm) and diazotrophs) and dissolved organic carbon (DOC). The ecosystem state variables were converted to inherent optical properties (IOPs; absorption and scattering). The IOPs were then converted to apparent optical properties (AOPs), specifically $R_{rs}(\lambda)$, through the radiative transfer software Hydrolight (Mobley 1994; Mobley and Sundman 2008a,b). Two sensitivity studies were carried out, each with several scenarios to isolate the effects of chlorophyll concentration, phytoplankton cell size, and size-varying phytoplankton absorption on $R_{rs}(\lambda)$. The goal was to determine the relative contribution of phytoplankton cell size and chlorophyll to overall $R_{rs}(\lambda)$ and to understand where a standard band-ratio algorithm (OC4) may under/overestimate chlorophyll due to $R_{rs}(\lambda)$ being significantly affected by phytoplankton size.

8.2.1.2 Findings

In the first sensitivity study, an annual average of CCSM-3 BEC model output was explored for the following scenarios: 1) full cell size and chlorophyll concentration variability, 2) chlorophyll concentration held constant at the global mean (0.17 mg m^{-3}) and, 3) phytoplankton size composition held constant at the global mean (35% microplankton). Mouw et al. (2012) found that chlorophyll concentration has the primary impact on remotely sensed reflectance. Phytoplankton cell size was found to contribute secondarily to $R_{rs}(\lambda)$ variability and to amplify or dampen the seasonal cycle in $R_{rs}(\lambda)$, driven by chlorophyll. Globally, the percent fractional error of phytoplankton cell size on $R_{rs}(443)$ was lower than the percent fractional error of Chl-a, and generally opposite in sign. High latitude regions

expressed the greatest impact on $R_{rs}(443)$ due to phytoplankton size (Figure 8.4a,b). High latitudes are particularly susceptible to under- and overestimation of remote sensing reflectance. When not considering phytoplankton size effects, $R_{rs}(443)$ will be underestimated in the subtropical oceans in both hemispheres during local spring through early fall months and overestimated during local winter months.
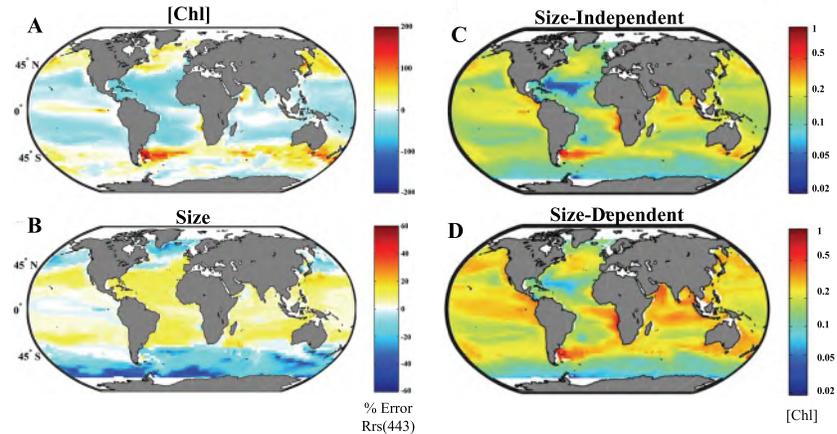
Figure 8.4 Examples of the sensitivity studies investigating the impact of phytoplankton cell size and chlorophyll on overall spectral remote sensing reflectance and chlorophyll concentration. Sensitivity Study 1: annual percent fractional error of normalized remote sensing reflectance at 443 nm for A) constant chlorophyll concentration, B) constant phytoplankton community size distribution. Sensitivity Study 2: SeaWiFS standard band-ratio chlorophyll concentration (OC4, mg m^{-3}) resulting from the spectral remote sensing reflectance values for C) size-independent chlorophyll-specific absorption and D) size-dependent chlorophyll-specific absorption. This resulted in an annual average of -21.5% and 1.4% difference in OC4 Chl-a estimates for the size-dependent and size-independent cases respectively. Modified from Mouw et al. (2012), with permission from Elsevier.

Satellite algorithms and optical models often assume constant average chlorophyll-specific phytoplankton absorption spectra. The second sensitivity study aimed to demonstrate when and where this assumption does not hold true; identifying the seasonal and spatial variability of where $R_{rs}(\lambda)$ was significantly impacted by phytoplankton size composition. Two scenarios were performed in this sensitivity study: 1) full cell size and chlorophyll concentration variability, where phytoplankton absorption varies due to phytoplankton cell size (size-dependent) and, 2) chlorophyll concentration variability only; phytoplankton chlorophyll-specific absorption held constant, thus absorption did not change with variation in cell size (size-independent absorption). In each of the scenarios, only the parameters of interest were changed; all other optical parameters were treated identically.

When applying band-ratio algorithms, remote sensing reflectance and chlorophyll concentration are empirically related; thus an impact on $R_{rs}(\lambda)$ by phytoplankton size will also impact Chl-a estimates. As Chl-a increases, phytoplankton absorb more light, resulting in decreased $R_{rs}(443)$. In the second sensitivity study, Mouw et al. (2012) showed that when phytoplankton size effects are not considered, Chl-a will be underestimated by standard satellite algorithms (e.g., OC4). Mouw et al. (2012) found that considering size-varying phytoplankton absorption in an optical model resulted in the global average $R_{rs}(443)$ and OC4 Chl-a being, respectively, 14% lower and 22% higher than when holding phytoplankton absorption constant (Figure 8.4c,d). The difference between scenarios reflects the impact that phytoplankton size has on remote sensing reflectance and chlorophyll estimates. These results have important implications as to when and where the satellite standard algorithms will either overestimate or underestimate Chl-a due to $R_{rs}(443)$ being significantly affected by phytoplankton size.

8.2.2 Case Study: Exploring uncertainty in the remotely sensed Chl product derived from limited *in situ* observations

Ocean colour products such as Chl-a are determined via algorithms that are based on concurrent measurements of remotely sensed reflectance and *in situ* observations of radiometry and of the variable of interest (e.g., Chl-a from HPLC). However, there are a limited number of *in situ* measurements with which to calibrate the algorithms. Moreover, recognizing that different waters can have distinct optical properties (Szeto et al. 2011; Moore et al. 2014), there have been several projects to produce regionally distinct algorithms (e.g., Szeto et al. 2011; Johnson et al. 2013, release V3 of the OC-CCI project, https://www.oceancolour.org). How much of the uncertainty in Chl-a arising from the algorithm itself could be reduced with more *in situ* data? Models with their "full" knowledge (at least of the model system) can be used to explore these uncertainties. Dutkiewicz et al. (2018) used a numerical model that fully coupled radiative transfer with a 3-D physical-biogeochemical model to explore uncertainty in a widely applied ocean colour Chl-a algorithm.

8.2.2.1 Approach

The model simulates the spectral absorption and scattering properties of water molecules, nine phytoplankton types (and their Chl-a content), detritus, and coloured dissolved organic matter (CDOM). Irradiance just below the surface of the ocean is provided by the Ocean-Atmosphere Spectral Irradiance Model (OASIM, Gregg and Casey 2009). The model includes explicit radiative transfer of spectral irradiance in 25 nm bands between 400 and 700 nm, which is used to calculate upwelling irradiances at the surface of the ocean. Together with the downwelling irradiance, the model outputs remotely sensed reflectance. This model provides both the "actual Chl-a", that is dynamically changing within the nine phytoplankton functional types, and the ocean colour-like "remotely sensed Chl-a" product derived using the same approach as the commonly used blue/green ratio algorithm (O'Reilly et al. 2000). Specifically, the equivalent empirical relationship used in the real world algorithm is found within the model, but using model results to determine the necessary coefficients. Comparison to the model "actual" Chl-a and this derived Chl-a provides a means to ascertain uncertainties in the derived product, which is very challenging to achieve in the real ocean due to very limited concurrent ocean colour and *in situ* measurements.

8.2.2.2 Findings

The study found that the mean absolute bias between derived and actual Chl-a was 22% (Figure 8.5 a,d) when using a simple blue/green reflectance ratio algorithm defined using coefficients determined by limited sampling of concurrent model "actual" measurements of Chl-a and radiometry. Using the maximum number of "*in situ*" measurements (i.e., all the model daily output for 13 years) to train the algorithm, produced mixed results. There was an overall global improvement, but at the expense of some regions, especially in lower latitudes where the biases increase (Figure 8.5b,e). Regional specific algorithms (i.e., where the coefficients were determined using the daily output in each grid location independently of other locations) provide a significant improvement (Figure 8.5c,f), at least in the annual mean (bias reduced to 17%). However, regardless of the data density available for calculation of the algorithm coefficients, there was often a temporal mismatch between the derived and actual Chl-a. These mismatches stemmed from the temporal decoupling between Chl-a and other optically important water constituents (such as coloured dissolved organic matter and detrital matter). The degree of decoupling differs regionally and over time. For example, in many highly seasonal regions, the timing of initiation and peak of the spring bloom in the derived Chl-a lags the actual Chl-a by days, and sometimes weeks.
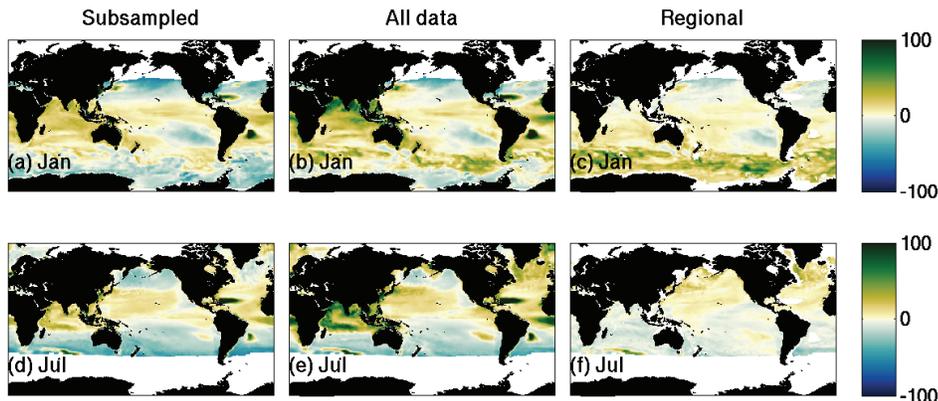
Figure 8.5 Percentage bias between monthly mean model "actual" Chl-a and model "derived" Chl-a using algorithm coefficients estimated for: (a, d) subset of output similar to that used in real world algorithm; (b, e) full model output daily over 13 years (about 11 million data points); and (c, f) each grid cell only using daily data over 13 years. Top row is for January, Bottom row is for July. White areas indicate unresolved high latitude regions and/or where PAR is less than 15 μEin m^{-2} s^{-1}. Credit: Modified from Dutkiewicz et al. (2018), Creative Commons Attribution 4.0 International License (CC BY 4.0).

8.2.3 Caveats

Including optics and radiative transfer sub-routines in large scale biogeochemical models is a relatively new development. Continued development is likely to improve the representation of ocean colour-like diagnostics. A continued dialogue with ocean colour experts will help in this regard. In the meantime, the limitations of these models must be kept in mind when using them in studies such as those described here. How well are the models capturing the individual optically important water constituents? For instance, different models make different assumptions on the parameterization of coloured dissolved organic matter and also the spectral nature of phytoplankton absorption (see for instance the sensitivity studies and discussion in Dutkiewicz et al. 2015a). Such choices could alter their ability to aid in algorithm development.

8.3 Climate Models for Trend Detection and Attribution

Biogeochemical models can also be used to project the future response of phytoplankton to climate change. Globally integrated primary production is, in general, predicted to decrease with continued climate change as a result of reduced supply of macro-nutrients to the sunlit layer (Bopp et al. 2013). However, the responses vary regionally, with some areas projected to increase in productivity due to reduction in light limitation (due to increased stratification) and higher growth rates due to increased temperatures (Taucher and Oschlies 2011; Dutkiewicz et al. 2013). Among the many model projections of phytoplankton response to climate change, there is large uncertainty in these regional responses. Models also do not necessarily account for all the changes that are likely to affect productivity, such as ocean acidification (though see e.g., Dutkiewicz et al. 2015b). In general, models suggest a shift to a greater relative abundance of small cells (Bopp et al. 2005; Dutkiewicz et al. 2013; Marinov et al. 2013) since smaller cells tend to require fewer nutrients. Studies have also investigated additional responses to climate change, such as the timing of the phytoplankton spring bloom, which is predicted to advance by up to 13 days in the North Pacific (Jang et al. 2011), and the area of the subtropical biome which is predicted to increase by 30% as the boundary moves northward by 1000 km by 2100 (Polovina et al. 2011).

All of these indicators of the phytoplankton response to climate change may be detectable in satellite ocean colour data. Models can be used to help us understand when effects of climate change will be observable relative to natural variability. They can also inform on locations for enhanced *in situ* observational efforts by, for instance, projecting which regions will undergo the fastest or most significant changes, and also areas that could benefit from enhanced satellite validation and/or regional satellite products?
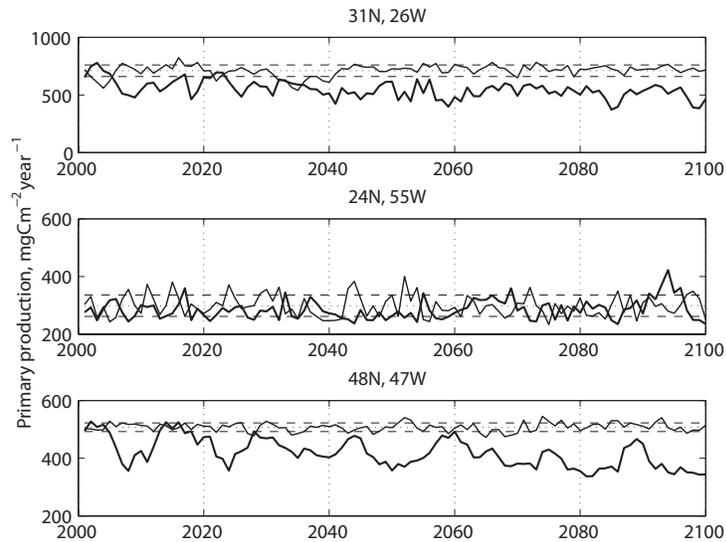
Figure 8.6 Examples of control and global warming simulations showing how both natural variability and long-term trends occur simultaneously in long time series. Output from the GFDL MOM4-TOPAZ model (Dunne et al. 2013) control and warming simulation, run for the period 2001–2100 under the IPCC A2 scenario (Nakicenovic and Swart 2000), is plotted for three locations in the North Atlantic. Thick lines are the annual mean primary production from the warming simulation; thin solid lines are the primary production from the control simulation; thin dashed lines are the mean of the control run ± one standard deviation (calculated over the 100 year time series of the control run).

Coupled biogeochemical models are particularly useful for discriminating between patterns of natural variability and anthropogenic climate change trends, as the same model configuration can be run in two modes: control and warming (see e.g., Figure 8.6). In a control run, no external forcing is applied, so that the modelled system reflects only the natural variability. In a warming run, anthropogenic forcing, typically in the form of a CO_2 emission scenario or representative concentration pathway developed by the IPCC (Nakicenovic and Swart 2000), is imposed. This results in a modelled system with both inherent variability and the climate change response. By comparing the control and warming runs, the response to natural variability can be distinguished from the response to climate change (see Figure 8.6). As coupled climate models generate their own internal variability, results can only be compared to observations in a statistical sense. For example, the model should simulate an ENSO cycle with a 2–7 year periodicity, but is not expected to generate specific El Niño events, such as the one that occurred in 1997/1998.

8.3.1 Case Study: Detection of Climate Change Trends in Satellite Ocean Colour Records

Since 1998, ocean colour satellites have provided a continuous record of data on surface ocean chlorophyll concentration. With longer time series, investigations of long-term, potentially climate

change-driven trends become a prospect. A key factor in determining long-term trends is the need to separate the natural variability (which in this context is "noise") from the trend (the desired "signal"). Chlorophyll concentrations have substantial natural variability, particularly at high latitudes, at a range of scales from daily through seasonal to interannual and decadal. Notwithstanding the pronounced seasonal cycle, chlorophyll concentration also responds to interannual variability in light and nutrient availability that may be driven by climatic oscillations, such as the El Niño-Southern Oscillation or the North Atlantic Oscillation. In order to distinguish a trend from the background natural variability, the signal must be substantially greater than the noise.

8.3.1.1 Approach

Henson et al. (2010) explored this issue using a suite of three climate models, run in both hindcast (1958–2006) and future projection (2001–2100; IPCC A2 scenario) models (see Chapter 3 and Appendix 4 for definition of terms). Each of the models (GFDL TOPAZ, Dunne et al. 2013; IPSL PISCES, Aumont and Bopp 2006; and NCAR CCSM-3, Moore et al. 2004; Doney et al. 2006) represents two or three phytoplankton groups (diatoms, small phytoplankton and, in the case of GFDL and NCAR, diazotrophs) and one or two zooplankton groups (PISCES includes both micro and mesozooplankton). All models include limitation of phytoplankton growth by nitrate, phosphate, silicate and iron. Hindcast simulations were forced with reanalysis data of atmospheric conditions. For the future projection runs, the fully coupled climate-biogeochemistry versions of the models were used. Historical forcing (greenhouse gases, aerosol emissions) from 1860–2000, and the IPCC A2 scenario (Nakicenovic and Swart 2000) from 2001–2100 were used.

First, hindcast runs were used to determine if trends in the existing satellite record are representative of long-term change, or are capturing natural decadal variability. The study showed that trends in Chl-a for 14 representative ocean biomes and the oligotrophic gyre size over 10 years of SeaWiFS data (1998–2007) were within the range of previous decadal variability simulated by the model for the period 1958–2006, implying that the satellite-derived "trends" were artifacts arising from the pronounced interannual variability in Chl-a and short satellite time series. In the second part of the analysis, Henson et al. (2010) used the future model projections to investigate how many years of data would be needed to separate a climate change trend from the natural variability. The formulation of Weatherhead et al. (1998), which uses the ratio of noise (natural variability) to signal (trend), including accounting for autocorrelation in the time series, allows the length of data record needed to distinguish a trend to be calculated. The trend was calculated for each model grid point using linear regression, and the noise was defined as the standard deviation of the residuals after the trend was removed from the time series.

8.3.1.2 Findings

This study concluded that ~40 years of continuous data are needed to detect a globally-averaged climate change trend (Figure 8.7). Some regions had shorter detection times (e.g., equatorial regions, 20 to 30 years), while the Southern Ocean had very long detection times (~60 years). In some regions, notably the oligotrophic gyres, a climate change-driven trend was not distinguishable from natural variability before the end of the simulation in 2100. A similar framework was used by Henson et al. (2013) and Henson et al. (2018) to explore whether trends in phytoplankton phenology may be more rapidly detectable than trends in chlorophyll concentration. However, use of bloom initiation or peak timing did not reduce the 30 to 40 year timescale for trend detection. Finally, Henson et al. (2016) found that derived properties such as primary production, export production or diatom-driven production, also require ~30 years of continuous data for trend detection.
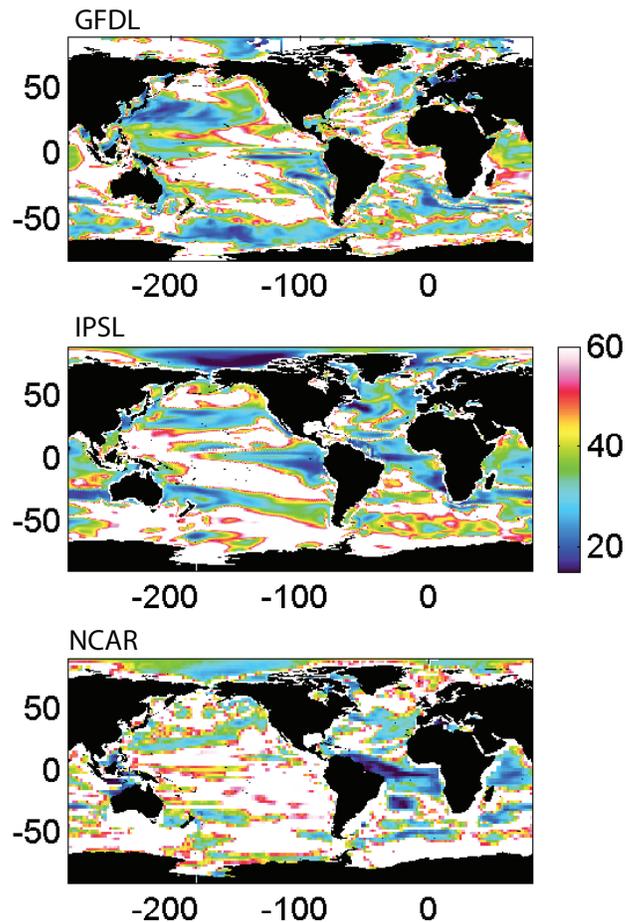
Figure 8.7 Number of years of continuous data required to distinguish a climate change-driven trend in chlorophyll concentration from natural variability, calculated for the GFDL, IPSL and NCAR models (IPCC A2 scenario, 2001–2100). White areas indicate where the trend is not statistically significant. Figure modified from Henson et al. (2010), Creative Commons Attribution 3.0 Unported License (CC BY 3.0).

8.3.2   Case Study: Exploring the effect of data record gaps on trend detection

Detecting long-term trends requires a stable, consistent and (ideally) gap-free dataset. How do discontinuities in a time series, such as those that might be introduced by a break in the satellite ocean colour record or a change of sensor without sufficient cross-calibration, affect our ability to detect climate change-driven trends? Beaulieu et al. (2013) explored this question by using output from three coupled models forced with the IPCC A2 future climate change scenario from 2001–2100.

8.3.2.1   Approach

Beaulieu et al. (2013) used the same models as Henson et al. (2010) (Section 8.3.1) and extended their approach to include a discontinuity in the time series. The modelled time series were "interrupted" at different points in the time series and the effect on trend detection was assessed. The interruption was included in the linear trend calculation as an indicator function representing the effect of the change, where the indicator had a value of 0 before the discontinuity and 1 afterwards.

8.3.2.2 Findings

The number of additional years of data needed to detect a trend depends on when the discontinuity occurs. Beaulieu et al. (2013) concluded that if a discontinuity occurred half-way through the satellite chlorophyll time series, the number of years of data needed to detect a trend would increase from 27 to 43 years (global average). Model-derived trends were also combined with the observed range in autocorrelation and standard deviation of the SeaWiFS time series to explore how these factors influence trend detection. In a time series with substantial variability and autocorrelation (i.e., subsequent months are strongly dependent on each other), the number of years required to detect a trend would increase to 65 years (compared to 25 years with low variability and autocorrelation). For trend detection in ocean colour records, this presents a serious challenge: natural variability tends to be pronounced, particularly at high latitudes, and the decorrelation time scale is on the order of three months. At the time of publication of Beaulieu et al. (2013), the Earth observing community had serious concerns that a gap in the ocean colour record could occur due to the failure of MODIS-Aqua before the launch of VIIRS or OLCI on Sentinel-3. Of equal concern for trend detection was the possibility of an insufficient period of overlap between successive sensors which would prohibit a thorough cross-calibration, and therefore the data record may not be consistent over time. Although VIIRS and OLCI are now operational, combined satellite data and model studies of this kind can provide the impetus to maintain a long-term, consistent ocean colour record.

Breaks in the satellite record are just one of many possible sources of discontinuities in the record that can affect trend detection. For example, reprocessing of datasets (as is done regularly for ocean colour data as algorithms improve or sensors deteriorate) or sensor drifts due to aging, can introduce a discontinuity. Datasets that merge ocean colour records from multiple satellites, e.g., GlobColour or ESA OC-CCI, attempt to overcome issues associated with discontinuities by providing a long-term, consistent time series.

8.3.3 Case Study: Comparing trends in multiple components of the system

The inclusion of additional complexity in the parameterization of the light field in models can provide a mechanism to explore how additional ocean colour products (such as remotely sensed reflectance) may change over time.

8.3.3.1 Approach

The study of Dutkiewicz et al. (2019) used the ecosystem/biogeochemical optics model of Dutkiewicz et al. (2015a), described in Case Study 8.2.2, and applied a future climate change scenario. The model explored the changes that occur to the system, including the optics, under an unchecked emissions scenario (Sokolov et al. 2009), similar to RCP 8.5 (IPCC Report 2014) over the course of the 21st century. Using generalized least squares fits, the study calculated the trends of remotely sensed reflectance in 25 nm bands from 400 to 700 nm, as well as the optically important water constituents (e.g., Chl-a, detrital matter and CDOM), and asked when and where trends driven by the forcing were unambiguous relative to the natural interannual variability.

8.3.3.2 Findings

Remotely sensed reflectance (at least at wavebands shorter than 600 nm) had an unambiguous signal from the forced trends over more of the ocean than did the optically important constituents (Figure 8.8). This includes Chl-a which is usually used in studies to determine changes to the marine ecosystem. This finding can be explained by the reflectances individually having smaller relative interannual variability and also integrating changes over all in-water constituents. This study identifies 467–512 nm as promising wavelengths for capturing the trends in remotely sensed reflectance over more of the ocean

than other wavebands. In the model, the trends in remotely sensed reflectance at 475 nm reached over 1% per decade in some locations. Thus, sensor drift will need to be less than 1% to observe trends. Although the model bands were too coarse (at 25 nm) to provide more details of the best wavelength and bandwidth most suitable for trend analysis, current and historic sensors (e.g., SeaWiFS, MODIS, VIIRS, MERIS) have all included wavebands around 490 nm. This modelling study suggests that future missions should maintain a compatible band in order to detect the earliest signatures of marine ecosystem changes.
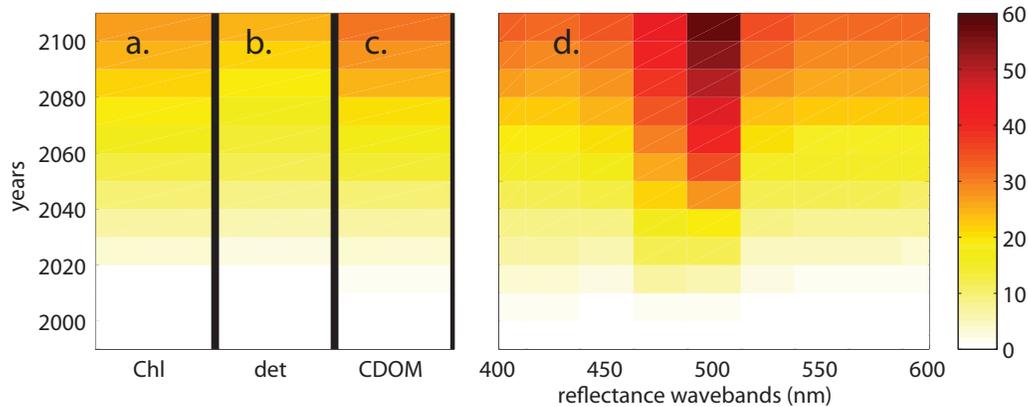
Figure 8.8 The amount of open ocean area that has a statistically significant trend over the course of the 21st century. (a) Chl-a, (b) "detritus" is the non-living particulate organic pool, (c) "CDOM" is the coloured dissolved organic matter, (d) the remotely sensed reflectance (R_{rs}) in 25 nm wavebands from 400 to 600 nm. R_{rs} results for higher wavebands are not shown as the model underestimates the natural variability in those wavebands. Figure modified from Dutkiewicz et al. (2019), Creative Commons Attribution 4.0 International License.

8.3.4 Caveats

A caveat of using biogeochemical models for studies of variability or climate change is that these responses may depend strongly on the chosen parameterisations for biological processes (Taucher and Oschlies 2011; Laufkötter et al. 2015). A corollary is that if a model has been tuned to reproduce the response to contemporary ocean, future responses to climate change can be expected to mirror the response to natural variability. For example, suppose we observe a decrease in primary production and trace it to an increase in stratification. We might assume that in the future, stronger stratification will consistently result in decreased primary production. However, additional processes that may not be included in the model, such as changes to dust deposition, ecosystem adaptation or evolution, may occur, resulting in differences in how primary production will change relative to what the model projects. Provided these caveats are borne in mind when assessing biogeochemical model output, the combination of satellite ocean colour data and models is a powerful tool for investigating natural variability and climate change.

8.4 Models Informing Future Ocean Colour Products and Missions

Biogeochemical models provide a spatially explicit indication of the magnitude and rate of change at interannual (see Case Studies in Chapter 7) and century (see Case Studies 8.3.1, 8.3.2 and 8.3.3) time-scales. Analyses such as Henson et al. (2010) and Dutkiewicz et al. (2019) also provide spatially explicit

estimates of the length of time series needed to separate the climate change-driven trends from natural variability. Taken together, this information can be used to plan sites for future time series stations to ensure they are located in regions of large or rapid response to interannual variability or climate change. The model studies also indicate regions where climate change response in ecological/optical properties is likely to be weak, or where the natural variability is so large that a trend is undetectable. Clearly, if the goal is to study the phytoplankton response to climate change, these areas should be avoided. Existing time series of observations can be analysed in the framework of these projections to distinguish the climate change signal, provided the data record is long enough. However, formal detection/attribution studies (e.g., Hasselmann 1993) are currently limited by the short length of ocean colour time series. Modelling studies, as outlined here, provide impetus for the sustained observations of ocean biology that are essential for detecting long-term trends. These priority sites will also facilitate improved satellite validation in those regions and/or development of regional products.

In planning for future missions, one factor to be considered is the degree of acceptable instrument drift. This is particularly important if the instruments are designed with the intention of capturing trends. There is also considerable effort expended to determine the best spectral bands for satellite sensors (Lee et al. 2007; Wolanin et al. 2016). However, estimates of trends and the choice of the bands best suited to capturing long term trends have been hindered by lack of knowledge of how the ocean's optics will change. Climate change models can help in these regards. The study of Dutkiewicz et al. (2019) (Case Study 8.3.3) found that the trends in remotely sensed reflectance at 475 nm reached over 1% per decade in some locations, i.e., sensor drift will need to be less than this to capture the predicted trend.

Observing System Experiments (OSEs for existing observations, see Appendix 4 for definition) and Observing System Simulation Experiments (OSSEs for future observations) are modelling studies that can be used to help design observing systems. A typical OSSE consists of a state-of-the-art "nature run" that realistically represents the phenomena of interest and is assumed to be the "true" ocean, and a second non-assimilative, operational ocean model. The nature run output is sub-sampled using the space/time sampling characteristics of the observing system under evaluation, and these "observations" are then assimilated into the operational ocean model. The impact of the observing system is quantified by comparing the reduction in errors from the operational model with respect to the nature run. OSSEs have been in use longer, and are more advanced for the atmosphere, compared to the ocean.

OSEs and OSSEs can be used in a variety of ways: (1) to provide a quantitative assessment of the potential impact of existing or future observing systems on the Earth system and its prediction, (2) to evaluate new methodologies for the processing and assimilation of remotely sensed data, (3) to evaluate tradeoffs in the design and configuration of observing systems (e.g., coverage, resolution, accuracy and data redundancy), and (4) to determine the ability of existing observing systems to detect climatic trends and to optimize the observing system for climate monitoring and other applications. OSSEs are useful tools for determining the usefulness of a proposed observing system and how this system may improve ocean analyses and forecasts. Such experiments are relatively cheap when compared to the investment required to establish new observing systems.

Observing System Experiments have been successfully performed to evaluate ocean observing systems such as satellite altimetry, satellite sea surface temperature, and Argo floats (e.g., Raicich and Rampazzo 2003; Oke and Schiller 2007; Oke et al. 2015; Halliwell et al. 2015). The use of OSSEs to evaluate the effects of existing or future observations on ocean biogeochemical models is even less advanced. Some studies have used OSSEs to demonstrate the impact of an existing *in situ* dataset on simulated conditions for ocean biogeochemistry (e.g., Lenton et al. 2006; Majkut et al. 2014; Oke et al. 2015; Basu et al. 2018) but to our knowledge no OSSEs have been performed to assess existing or future satellite ocean colour missions.

Models can support the development of future satellite missions in various ways. For example, models can be used to generate simulated datasets to test algorithms, atmospheric corrections and instrument design (band location, width). Models can also be used for Observing System Simulation

Experiments to assess the effects that existing or future observations may have. While this approach has not been used widely, upcoming field and satellite missions should include a modelling component from the early stage onwards.

Currently, measurements of ocean water constituents and phytoplankton community composition are hindered because of the limited number of spectral bands available on contemporary satellite ocean colour sensors. Future satellite missions such as the Plankton, Aerosols, Clouds and Ecosystem (PACE) mission will provide hyperspectral water leaving radiances (350–800 nm at 5 nm resolution plus several additional short-wave infrared bands). Moving from multispectral to hyperspectral resolution will allow us to distinguish an increasing number of water constituents and to potentially identify the composition and distributions of phytoplankton communities. This information will improve our understanding and quantification of ocean biogeochemical cycling and ecosystem function in response to anthropogenic and natural environmental variability and change. To support pre-launch mission development and assess on-orbit capabilities, Gregg and Rousseaux (2017) conducted a model study to simulate upwelling radiance between 250 and 800 nm at 1-nm resolution. There was significant correlation between the simulated dataset and data from an existing satellite ocean colour platform (MODIS-Aqua, Figure 8.9).
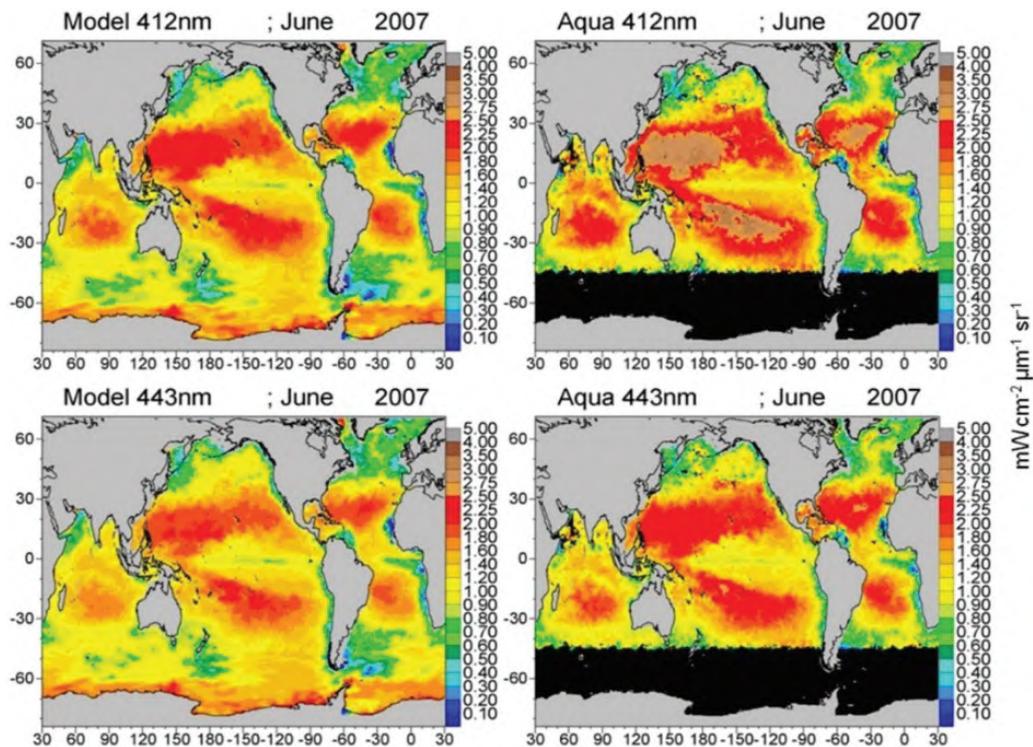
Figure 8.9 Model normalized water-leaving radiances for 412 and 443 nm compared to MODIS-Aqua radiances. Reproduced from Gregg and Rousseaux (2017), Creative Commons CC-BY licence.

Gregg and Rousseaux (2017) found that the unassimilated radiances were within -0.074 mW cm^{-2} μm^{-1} sr^{-1} of MODIS-Aqua radiances at 412, 443, 488, 531, 547, and 667 nm. This difference represented a bias of -10.4% (model low). A mean correlation of 0.706 (P < 0.05) was found with global distributions of MODIS radiances. The skill in the model to generate hyperspectral water leaving radiances suggests that it can be used for mission design and analysis. While the authors caution the limitations of this dataset, they also highlight a range of applications from band selection (location, width, number and

center location), to testing the variability of various orbital tracks over a range of solar and satellite angles, to the inclusion of optics, such as the spectral response function. These model applications can be used as a part of trade-off studies to provide answers on the best possible instrument to tackle the goals of the mission. The model can also be used for algorithm development effort and design of data collection strategies.

8.5   Summary: Using Models to Inform Ocean Colour Science

In this chapter we have reviewed modelling studies that contribute to ocean colour science. Models are imperfect simplified systems representing the real ocean. However, models are internally consistent and outputs provide full representation in 4 dimensions (space and time). As such, models can be more thoroughly interrogated than the real world (where there are limited *in situ* data, missing satellite imagery, and observations sometimes have unquantifiable uncertainties). For example, we can sub-sample the model in space and time to mimic typical satellite coverage, which allows exploration of biases in ocean colour products (e.g., monthly means) and in phenological studies. Future model projections provide estimates of climate change impacts on marine ecosystems, and suggest the spatial patterns, magnitudes and rates of change that may occur. Future projections also provide a basis for development of monitoring strategies for detection of climate change impacts on ocean ecosystems, including where, when and for how long we should observe.

There is tremendous potential for models to aid even more directly in ocean colour studies. Recent advances in models to include optics, and output such as reflectance, means that models can assist in algorithm development and mission planning, e.g., choice of wavebands. There is also important potential for use of biogeochemical/ecosystem models in OSSEs for future observational network planning. Further, models could be used to extend ocean colour output to depth, to provide information on the vertical distribution of Chl-a, information which is not currently observable from space.

8.6   Recommendations to Facilitate Modelling Applications

In this chapter we have highlighted how models can be an additional tool that can be of great use to the ocean colour community. They provide a laboratory for trend and detection analyses, for quantifying consequences of data gaps in space and time, and for testing and developing ocean colour algorithms. Models can provide valuable context for mission development. The following steps can facilitate the full potential of these applications:

❖ Communication and collaboration: Increased dialogue between the modelling and ocean colour communities is recommended through workshops, dedicated collaborations, and inclusion of models and modellers in ocean colour algorithm development, future satellite mission and observational network design.

❖ Inclusion of optics into models: A continuation of model development is recommended that includes optically important components and other output that relates directly to ocean colour.

❖ Use of models for algorithm development: The use of optical numerical model output is recommended as a laboratory to explore hyperspectral algorithm methods and frameworks.

Chapter 9

Summary and Recommendations

Stephanie Dutkiewicz, Mark Baird, Stefano Caviatti, Stephanie Henson, Anna Hickman, Colleen Mouw, Cecile Rousseaux and Charles Stock

NOTE: In this report, the word "model" refers to process-based three-dimensional biogeochemical/ecosystem computer models at large regional or global scales.

Through this report, we hope to promote an open discussion between the ocean colour and the biogeochemical/ecosystem/climate modelling communities. We have provided easily accessible information about ocean colour (Chapter 2), biogeochemical and ecosystem modelling (Chapter 3), and the inherent problems of linking the two (Chapter 4). We have demonstrated that ocean colour is uniquely important for model evaluation and data assimilation (Chapter 5 and 6), that models can also be useful for the ocean colour community by providing context beyond what is measured from space (Chapter 7), and can be a laboratory or tool to help ocean colour science (Chapter 8). In essence, we have emphasized the strength that can come from a more synergistic use of ocean colour and model products.

Models, *in situ* observations, and ocean colour products are different tools that can each be used to understand ecological and biogeochemical processes in the ocean. However, each provides a different "measurement" (see Chapter 4, Figure 4.2) inhibiting straightforward inter-comparison (Chapter 4). Ocean colour scientists and modellers are distinct communities that have evolved separately, with different priorities and terminology. However, a synthesis between ocean colour and models can help us understand the ocean state and the underlying dynamics better. Integrated "results" are therefore important, but will need a discussion between the different communities. Improved alignment between model output and satellite products is likely to be an important feature of future work. New developments in numerical models over the last few years to include components that directly link to ocean colour products (e.g., remotely sensed reflectance) can help with stronger links between modelling and ocean colour communities.

9.1 Summary

9.1.1 Using ocean colour products to evaluate models

There is a need for increased skill assessment in ecosystem/biogeochemical modelling studies. Increasing calls for ecosystem predictions to inform decision-making demands a closer focus on the metrics most suitable — most "fit to purpose" — for the intended applications. Ocean colour, with its combination of higher spatial and temporal coverage than *in situ* measurements, enables us to match the space and time-scales of model predictions. The multi-decadal duration of the ocean colour time series supports retrospective forecast experiments on seasonal to inter-annual time-scales. These become more robust as each year of new data is added. However, limitations of the use of ocean colour products come from:

❖ Regional (and temporal) biases;

❖ Large, usually unquantified and uncharacterized, uncertainties;

❖ Complications with merged products and satellite sensor drifts.

Recommendations

We recommend that modellers endeavour to be better informed about using satellite data, as well as the development of new satellite approaches. We stress the importance of learning more about the other discipline, and taking the responsibility to understand the tools being used, and how and when to use them. For instance, point-wise model comparison to ocean colour products is often not appropriate. Moreover, it is important to choose the appropriate and "most similar" ocean colour product to compare model output to, and understand the uncertainties (Chapter 4 and Section 9.2.2). However, the task of learning will be greatly enhanced by better and more easily available documentation, as well as more opportunities for cross-disciplinary education (e.g., through high level, but not community-specific workshops, town halls, breakout sessions at large international meetings etc.). Agencies and groups providing ocean colour products should examine and explore how additional information can be presented alongside satellite products to help modellers make informed choices and interpretations. This includes limitations from merging products and details of potential instrument drifts. On the other hand, the needs of modellers should be borne in mind with new ocean colour product development, e.g., through "User requirement" questionnaires and reports. Though difficult to implement, there might be an attempt to standardize the way ocean colour products are used in model evaluation methods. The ocean colour community should also have larger appreciation of models, their needs and limitations.

9.1.2 Using models and ocean colour products together: data assimilation

Data assimilation provides a formal mechanism to synthesize observations and models into a single product. Biogeochemical/ecosystem data assimilation is still a new field, with teams usually aligned with groups undertaking physical data assimilation. As a result, new techniques are regularly becoming available, but the work to adapt these techniques to biogeochemical phenomena is moving more slowly.

Recommendations

Given the relatively small number of marine biogeochemical data assimilation studies that have been undertaken so far, it is recommended that groups continue to explore present approaches more fully, and try new approaches. It is too early to recommend a single, or small number, of data assimilation approaches. Nonetheless, it does appear that the assimilation of satellite-derived chlorophyll has problems. It is likely that assimilation of less uncertain products, such as vertical attenuation at 490 nm, or remote-sensing reflectance, is a direction forward. These represent an effort to use the model to produce an output that is closer to the raw observation of the satellite. It is also clear that, for biogeochemical/ecosystem data assimilation to reach its full potential, a better understanding of the uncertainty in ocean colour data products is required.

9.1.3 Using models and ocean colour products together: process studies

Formal data assimilation is not the only way to use models and ocean colour products together. Use of data assimilated or "unconstrained" (i.e., those that do not use data assimilation) biogeochemical models in combination with satellite-derived ocean colour data has provided a wealth of information on the processes controlling observed phytoplankton variability, leading to a greater understanding of how this variability arises. Using a model to help understand the phenomena (such as interannual variability) observed in the ocean colour records is a promising direction (see e.g., studies of Resplandy et al. 2009 and McKinley et al. 2018, Chapter 7). Provided a model captures the processes of interest and

variability at the appropriate scales, it can also provide context not available from satellite observations, such as nutrient distribution, biogeochemical and trophic fluxes (e.g., air-sea carbon exchange, net ecosystem metabolism, zooplankton grazing), physical phenomena and depth dependent information (e.g., horizontal transport, deep chlorophyll maximum and carbon export).

Recommendations

Model and ocean colour products should be chosen based on the question to be addressed, and more discussion between the ocean colour community and modellers is advised. Points to keep in mind are the spatial and temporal resolution required, as well as the level of the biogeochemical, ecological and optical complexity to be included. Higher levels of complexity require increased computational expense, and in most models, there is a trade-off between these different components. As with ocean colour products, model documentation/articles should be clear on the limitation of the model and its outputs. Studies that use model and satellite products as independent and complimentary types of information will be improved by knowing the limitations and assumptions of each type of data. For applications where model and satellite products are assumed to represent the same property, it is important for models to keep developing to be closer to ocean colour products, so they compare more closely and at better spatial and temporal resolutions, to ocean colour products.

9.1.4   Using models to inform ocean colour science

Models offer an additional tool to the ocean colour community, by providing information that cannot be currently derived from satellite data, such as depth resolution and coverage, even in the presence of (for instance) clouds. Models can provide a laboratory to aid in trend and detection analyses, quantifying consequences of data gaps in space and time, testing and developing ocean colour algorithms, providing depth and temporal context, and supporting mission development (Chapter 8).

Recommendations

This is a particularly promising, but as yet under-utilized use of models. Continued dialogue and collaboration between modellers and ocean colour communities will be required to facilitate the full potential of this model application. Collaborations could provide demonstrations of models' usefulness in, for instance, algorithm development. We recommend continued development of models that include ocean optics and output that relate directly to ocean colour, as such models will be able to more directly connect with the ocean colour community. The potential usefulness of such models, for instance in algorithm development (see Chapter 8), should be borne in mind with new ocean colour product development. There should be an emphasis on including models in ocean colour studies, and particularly in mission and ocean monitoring design and planning. In particular, additional development of biogeochemical/ecosystem models for Observing System Simulation Experiments (OSSEs) should be a priority to help with observational network design. As yet OSSEs have not been used much in biogeochemical or ecological contexts (though they have been used in physical oceanography). For instance, OSSEs could be valuable in generating more robust global Chl-a products by indicating where additional *in situ* measurements are needed to reduce biases.

9.2   Final Recommendations

9.2.1   Recommendations for continued and new developments of ocean colour products

For the synergistic uses of models and ocean colour products, there needs to be a continued development and maintenance of ocean colour products. We recommend the maintenance of projects that continue

to improve algorithms (for instance including fixes for instrument drift) and that maintain a long-term continuity in ocean colour products (e.g., OC-CCI and GlobColour). Including a modeller in such projects would be helpful to provide a model context for the use of ocean colour products. We suggest prioritizing improving global products (e.g., Chl-a, POC) that reduce regional biases, rather than developing a diversity of unconstrained products (especially if these products are not consistent between themselves e.g., pools of carbon, primary production, phytoplankton functional types, see Chapter 4).

Rate measurements (e.g., primary production, grazing) from ocean colour are particularly useful to the modelling community, but need significant improvement in their overall certainty and documentation. For example, there are now a plethora of primary production products, with little clarity on which is most appropriate for a modeller to validate the output against. We recommend that the ocean colour community provide more information on the uncertainties in these products (e.g., review papers, inter-comparisons) especially when they are included as standard products. New geostationary satellites offer a promising avenue for better, and more types of, rate measurements.

As with primary production, there are a confusing number of phytoplankton functional type products. These would be more useful with greater information supplied by the ocean colour community. Reviews and intercomparisons of phytoplankton functional types exist (IOCCG 2014; Kostadinov et al. 2017; Mouw et al. 2017), but new methods are continually being developed.

We recommend improving the ease of access of the ocean colour products through portals e.g., Unidata's Thematic Real-time Environmental Distributed Data Services (THREDDS, `unidata.ucar.edu/software/thredds/current/tds/`), NASA Giovanni (`giovanni.gsfc.nasa.gov/giovanni/`), ESA CCI-OC (`www.oceancolour.org/portal/`) and NOAA ERDDAP (`upwell.pfeg.noaa.gov/erddap/index.html`). We also suggest greater transparency on what the products represent by providing more thorough documentation.

9.2.2 Recommendations for choosing ocean colour products to use in model studies

It is not necessarily obvious which ocean colour products to use in model/ocean colour synthesis studies, data assimilation, or model evaluation. Some considerations for deciding which of several ocean colour products to use are:

1. Which model output compares most directly to the ocean satellite colour product?

2. Are the uncertainties in the ocean colour product quantified, and if not, what are the levels of "derivedness" and uncertainty of the product (see Chapter 4, Figure 4.1)?

3. Are there underlying biases in model or satellite products due to assumptions, optical properties or biases in *in situ* data?

4. Which products are the priorities to "get right" (e.g., Chl-a, phytoplankton carbon or primary production)? What is the tradeoff in uncertainty with other products?

5. What are the biases from missing satellite measurements (i.e., cloud cover)?

6. For long time periods, merged products (e.g., OC-CCI) will be needed, and the issues with instrument drift must be taken into account.

It is important to consider separately the biases associated with a) the satellite measurement that provides the spatial and temporal variability, and b) the (normally empirical) relationships that scale the variability to the property of interest. It is also important to remember that mechanistically derived ocean colour products often do not have less uncertainty than empirically derived products. For some products (e.g., primary production) it is probably best to consider several products, in particular those that have been determined to fit into different categories (e.g., see Carr et al. 2006). It is also very

important to keep in mind that such comparisons are model-to-model: an empirically-based satellite model and a (typically more) mechanistic simulation model. Careful consideration should be given to what information is actually being gained from the satellite product and how to interpret comparisons. There is also value in exploring whether the model's emergent properties match the empirically found relationships in the ocean colour products (and the *in situ* data used for training). In fact, getting such relationships, and in particular gradients, in the properties right, may be more important than getting the absolute values of a certain field correct.

9.2.3 Recommendations for choosing model and model output for ocean colour/ process studies

Model and model output choice should be based on the type of question to be addressed, and the spatial and temporal scale required. The model should adequately capture the processes and timescales of interest (see Section 3.5). For instance, a coarse resolution climate model will not adequately capture processes for a harmful bloom algal study in a specific small region. Model output and how to compare to ocean colour (or other) products also requires careful consideration. For instance, model Chl-a may not always be exactly comparable to satellite derived Chl-a (see Chapter 4). The level of complexity of the ecosystem model should also be considered in the choice of model. For questions concerning changes in phytoplankton communities, one will require a model that captures at least some level of complexity of functional types or size classes. If the question is more optical in nature, then there are a few models that have incorporated various elements of the optics and radiative transfer that might be more appropriate (see Section 3.3). Since no model is perfect, we suggest that, if possible, it would be good to use several models. In this way model structure uncertainty will at least be partially addressed. We recommend contacting the model developers to discuss the use of model output, and the appropriateness to the question of interest, as well as becoming relatively conversant with the model to be used and the meaning of the output.

9.3 Bridging Across Scientific Communities

Many of the recommendations above emphasize the importance of facilitating more collaboration between the modelling and ocean colour communities. It is our belief that synergistic use of model and ocean colour products will provide a significant avenue for understanding our oceans and the life within it. This report is part of an important dialogue that needs to happen between the communities to facilitate these synergies. We suggest that such an open discussion should continue through mechanisms such as:

1. Breakout sessions or working groups at both large ocean colour and modelling meetings — with modellers presenting at the ocean colour meetings and ocean colour scientists at modelling meetings. The goal is to make each community more aware of the other. At the modelling meeting it would be useful to have talks based on which ocean colour products are available, how they are produced, and discussion of their uncertainties. At the ocean colour science meeting, talks could focus on the simple explanation of models, discussion of different types of model, and how ocean colour products are used. In both cases, it would be important to emphasise the synergistic use of models and ocean colour. Some targeted meetings include the International Ocean Colour Science (IOCS) and the Ocean Optics meetings, both held every second year, and the Advances in Marine Ecosystem Modelling Research (AMEMR) meetings held every third year, as well as connecting with the GODAE Ocean View community meetings. Some funding opportunities will likely need to be established to promote researchers attending meetings that they may not normally consider.

2. Facilitate more early-career cross-discipline collaboration through summer schools designed to attract scientists from both communities.

3. Integrate ocean colour and models at the project level and include both communities in large projects (e.g., EXPORTS, the Climate Modelling User Group (CMUG) in the ESA's Climate Change Initiative, www.esa-cmug-cci.org/, Ford and Barciela 2017).

4. In the process of developing ocean colour products there should be more evaluation across disciplines: modellers should be brought into the discussion earlier. This will also help to facilitate further collaborations.

5. Include modellers early on in future ocean colour satellite mission planning, and make more effort to include Observing System Simulation Experiments (OSSEs) in such plans for both satellite mission details, and in designing calibration and validation systems.

9.4 Looking Forward

9.4.1 Model development

Models continue to be developed to include more detail at the biogeochemical level, for example, including additional nutrients (Tagliabue et al. 2018), and carbon pools (Polimene et al. 2018), and at the ecological level, such as including more diversity and other trophic levels (Ward et al. 2012; Ward and Follows 2016; Coles et al. 2017; Stock et al. 2017; Leles et al. 2018; Zakem et al. 2018). The needs of the modelling community will change as ecosystem models become more complex (such as resolving mixotrophs, bacteria, viruses, other carbon pools). Simply developing satellite algorithms for deriving each of these is not recommended. Consideration is needed of users' requirements, *in situ* and satellite observational constraints, and of the assumptions, errors and uncertainties of different approaches, to assess the most robust comparisons between satellite products and model output. On the other hand, model development could be directed towards addressing novel ocean colour products. For example, as ocean colour products develop, it may emerge that (for instance) size distribution of particles can be more robustly determined than function. If that is the case, more emphasis on simulating size distribution in models may be a worthwhile direction.

There is some debate as to whether models should produce output that is more similar to what satellites measure (e.g., reflectance) or whether there should be satellite-derived products that are more similar to what models require (e.g., phytoplankton carbon). This discussion should take care to consider discrepancies between the measurements and the model output (see Chapter 4). Comparing the most similar quantities has distinct advantages. However, there are computational costs and additional complexities associated with models providing direct reflectance and optical output. Additionally, users of models (e.g., ecosystem managers) are often more familiar with quantities such as Chl-a rather than reflectance. We anticipate that different models and different uses will dictate how closely models will link to ocean colour products in the future. However, the development of models with greater connection to optics and ocean colour is likely to spur more synergy between the two communities, and should be strongly encouraged.

Improved numerical models and satellite products depend on *in situ* data. Continued field sampling and development of new techniques is clearly a requirement for both communities. Dialogue and interaction between observationalists, remote sensing scientists, and modellers is clearly desirable to maximise the opportunities for, and use of, field observations for modelling and remote sensing, and the synergy between all three (see for instance schematic Figure 4.2 in Chapter 4).

9.4.2 New ocean colour missions

With the anticipated increase of spectral, spatial and possibly temporal resolution, an expansion and refinement of satellite products is anticipated. Geostationary satellites allowing for multiple images of the same location, multiple times per day, will be important for capturing the short-term dynamics found in coastal and inland environments, and for improving estimates of rate processes. Improved spectral resolution (e.g., that anticipated on the PACE mission) will allow for improved/expanded phytoplankton discrimination through functional types and pigment discrimination, as well as lower uncertainty in colored dissolved organic matter absorption and its spectral slope. Improved discrimination of all other IOPs and products further downstream will be possible with expanded spectral observations at shorter wavelengths. The higher spectral resolution will expand carbon pool products and, potentially, carbon export, and improve parameters related to phytoplankton physiology. Increased spectral resolution is also anticipated to improve uncertainty of all of these products. Greater focus on product continuity over algorithm continuity will aid in the improvement. With greater sensing capability, there will be the possibility of more ocean colour products that are independent of each other. Lidar observations will allow depth resolution, that will be particularly important for modellers. Including OSSEs in the future missions is highly encouraged, as is the continual recruitment of modellers on science advisory teams.

Appendix 1: Mathematical Notation

Symbol	Description	Typical Units
θ_0	Solar zenith angle	deg
$a(\lambda)$	Spectral total absorption coefficient	m^{-1}
$a_{CDOM}(\lambda)$	Spectral coloured dissolved organic matter absorption coefficient	m^{-1}
$a_{dg}(\lambda)$	Spectral absorption due to coloured dissolved and detrital matter	m^{-1}
$a_{NAP}(\lambda)$	Spectral non-algal particle absorption coefficient	m^{-1}
$a_p(\lambda)$	Particulate absorption coefficient	m^{-1}
$a_{ph}(\lambda)$	Spectral phytoplankton absorption coefficient	m^{-1}
$a_w(\lambda)$	Spectral water absorption coefficient	m^{-1}
$b_b(\lambda)$	Spectral total backscattering coefficient	m^{-1}
$b_{NAP}(\lambda)$	Spectral particulate backscattering coefficient due to non-algal particles	m^{-1}
$b_{b-ph}(\lambda)$	Spectral particulate backscattering coefficient due to phytoplankton	m^{-1}
$b_{bp}(\lambda)$	Spectral particulate backscattering coefficient	m^{-1}
$b_{bw}(\lambda)$	Spectral water backscattering coefficient	m^{-1}
[Chl]	Chlorophyll-a concentration	$mg\ m^{-3}$ or $\mu g\ l^{-1}$
$E_d(\lambda)$	Spectral downwelling plane irradiance	$\mu W\ cm^{-2}\ nm^{-1}$
$F_0(\lambda)$	Spectral extraterrestrial solar irradiance	$mW\ cm^{-2}\ \mu m^{-1}$
$K_d(\lambda)$	Diffuse attenuation coefficient	m^{-1}
$L_a(\lambda)$	Spectral radiance due to scattering by aerosols	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
$L_g(\lambda)$	Spectral radiance resulting from sun glint	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
$L_r(\lambda)$	Spectral radiance due to (Rayleigh) scattering by air molecules	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
$L_{ra}(\lambda)$	Spectral radiance from multiple scattering interactions between molecules and aerosols	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
$L_t(\lambda)$	Spectral top of atmosphere radiance measured by the satellite sensor	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
L_w	Spectral upwelling water-leaving radiance	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
$L_{wc}(\lambda)$	Spectral radiance resulting from white caps	$\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$
$R_{rs}(\lambda)$	Spectral remote-sensing reflectance	sr^{-1}
$t(\lambda)$	Spectral atmospheric diffuse transmittance from the ocean surface to a satellite sensor	unitless
$T(\lambda)$	Spectral direct transmittance from the surface to the sensor	unitless
$t_0(\lambda)$	Spectral atmospheric diffuse transmittance from the sun to the ocean surface	unitless
Z_{eu} or $Z_{1\%}$	Euphotic zone depth	m

Appendix 2: Ocean Colour Acronyms and Sensors

Ocean Colour Acronyms

AOPs	Apparent Optical Property
CDOM	Chromophoric Dissolved Organic Matter
CI	Colour Index
DIC	Dissolved Inorganic Carbon
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter
ENSO	El Niño-Southern Oscillation
EO	Earth Observation
ESA	European Space Agency
ESRID	Empirical Satellite Radiance-In Situ Data
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
IOD	Indian Ocean Dipole
IOPs	Inherent Optical Properties
IQR	Interquartile Range
LME	Large Marine Ecosystem
LUT	Look-Up-Tables
MAD	Median Absolute Deviation
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MLD	Mixed Layer Depth
NAO	North Atlantic Oscillation
NAP	Non-algal Particles
NIR	Near Infra-Red
NPP	Net Primary Production
NSE	Nash-Sutcliffe Model Efficiency
OASIM	Ocean-Atmosphere Spectral Irradiance Model
OSC	Optically Significant Constituents
PAR	Photosynthetically Available Radiation
PIC	Particulate Inorganic Carbon
PFT	Plankton Functional Type
PIC	Particulate Inorganic Carbon
POC	Particulate Organic Carbon
PSC	Plankton Size Class
PSD	Particle Size Distribution
PTC	Plankton Taxonomic Composition
QAA	Quasi-Analytical Algorithm

RMS   Root-Mean-Square
RMSD   Root Mean Squared Difference
SNAP   Sentinel Application Platform
SST   Sea Surface Temperature
TOA   Top of Atmosphere
TSM   Concentration of total suspended material (mg l^{-1})
VGPM   Vertically Generalized Production Model

Ocean Colour Sensors

CZCS   Coastal Zone Color Scanner
EnMAP   Environmental Monitoring and Analysis Program
GEO-CAPE   GEOstationary Coastal and Air Pollution Events
GOCI   Geostationary Ocean Color Imager
HICO   Hyperspectral Imager for the Coastal Ocean
HyspIRI   Hyperspectral Infrared Imager
MODIS   Moderate Resolution Imaging Spectrometer
MERIS   Medium Resolution Imaging Spectroradiometer
MSI   Multispectral Instrument (Sentinel-2)
OCM   Ocean Colour Monitor
OLCI   Ocean and Land Colour Instrument
OLI   Operational Land Imager (Landsat-8)
PACE   Plankton, Aerosol, Cloud, ocean Ecosystem
SeaWiFS   Sea-viewing Wide-Field-of-view Sensor
SGLI   Second Generation Global Imager
VIIRS   Visible Infrared Imaging Radiometer Suite

Appendix 3: Satellite Imagery Terminology

Term/Acronym	Definition
Absorption	Any process whereby radiant energy is converted to non-radiant energy and results in the disappearance of photons.
Apparent optical property (AOP)	Any optical quantity that depends on the properties of the water and on the ambient light field.
Algorithm	A sequence of mathematical steps to convert radiometric observations to geophysical parameters.
Atmospheric correction	The process of removing all other radiances beyond the water-leaving radiance from the total radiance measured at the top of the atmosphere.
Attenuation	The loss of radiant energy due to absorption or scattering.
Backscattering	Scattering through angles greater than 90 degrees relative to the incident direction.
Band ratio	The division of radiances measured at two difference wavelength bands often empirically related to a geophysical parameter.
Climatology	Averaging satellite pixel values over a specified period of time.
Coloured dissolved organic matter (CDOM)	High-molecular-weight organic compounds (humic and fulvic acids) formed from the decomposition of plant tissue; they strongly absorb light at the blue end of the spectrum and can give water a yellowish colour at high concentrations. Also called chromophoric dissolved organic matter, yellow matter, gilvin, or Gelbstoff.
Conus	The area of the Earth able to be imaged by a satellite in geostationary orbit.
Empirical	Relationships derived from *in situ* observations.
Euphotic depth	The maximum depth in the water column to which significant phytoplankton photosynthesis can take place.
Euphotic zone	The upper region of a water body in which significant phytoplankton photosynthesis can take place; often taken to be the layer down to which photosynthetically available radiation at noon falls to 1% its value just below the sea surface.
Fluorescence	A scattering process in which a photon is absorbed by a molecule and shortly thereafter another photon of greater wavelength is emitted.
Fluorescence line height	Fluorescence magnitude determined from three bands centered at 667, 678 and 748 nm. The magnitude of the 678 nm band is used to determine fluorescence and the 667 and 748 nm bands are used for the removal of the background contribution.
Geophysical parameter	Fundamental biological or physical measurement able to be made with remote sensing techniques that can be mapped on the Earth.
Imagery	A representation of measured energy emitted from a mapped location on Earth that may be converted to a geophysical parameter. Numerous pixels mapped together.
Inherent optical property (IOP)	Any optical quantity that depends only on the properties of the water and is independent of the ambient light field.
Irradiance	The radiant power per unit area, per unit wavelength, interval.
Nadir	A point directly below an observation point in the downward vertical direction.
Non-algal particle	All particles except living phytoplankton cells.

Normalized water leaving radiance	The radiance that could be measured by a nadir-viewing instrument, if the Sun were at the zenith, in the absence of any atmospheric loss, and when the Earth is at its mean distance from the Sun.
Ocean colour	A generic term referring to the spectral dependence of the radiance leaving a water body.
Optical depth	The optical distance in the vertically downward direction. The first optical depth is regarded as the depth for which light exiting the ocean is able to be measured remotely (1/attention coefficient).
Volume scattering function	Describes both the angular pattern of the light scattered from an incident direction and the magnitude (strength) of the scattering.
Photic depth	The depth at which an ocean colour radiometer is able to sense radiant energy.
Photosynthetically available radiation (PAR)	The integral over visible wavelengths (400–700 nm) of the number of photons available for photosynthesis.
Phytoplankton functional type	An aggregation of phytoplankton that share similar biogeochemical or ecological roles.
Pixel	The smallest non-divisible element of a two-dimensional digital image grid (an abbreviation for picture element).
Radiance	The radiant power in a beam per unit solid angle, per unit area perpendicular to the beam, per unit wavelength interval.
Radiative transfer	Conservation of energy in terms of the radiance that describes the rate of change with distance of the radiance in a collimated beam with a specified location, direction, and wavelength; the equation accounts for all losses (e.g., due to absorption and scattering out of the beam) and gains (e.g., by emission or scattering into the beam).
Radiometer	An instrument used to measure radiant energy.
Rayleigh scattering	The scattering (identical in the forward and backward directions) of light off of molecules that are smaller than the wavelength of light; inversely proportional to the fourth power of wavelength.
Remote sensing reflectance	The ratio of the "water-leaving" radiance in air to the downward plane irradiance incident onto the sea surface, with both measured just above the sea surface.
Satellite (ocean colour) product	The output of an algorithm applied to satellite imagery.
Scattering	Redirection of photons from the incident path.
Semi-analytical	An algorithm approach that accounts for theoretical understanding of optical relationships and reduces the reliance on empirical coefficients as much as possible.
Solar zenith angle	The angle between zenith (an imaginary point directly above a particular location) and the sun's location.
Spatial resolution	Ground size of an image pixel.
Spectral resolution	Number, spacing, and width of the different wavelength bands recorded.
Stray light	Light that is detected by an optical system and that belongs to a wavelength other than the one initially selected.
Temporal resolution	Frequency of flyovers by the sensor.
Top of atmosphere (TOA) radiance	Solar radiance scattered by atmospheric molecules and aerosols, Sun and sky radiance reflected by the sea surface, and from water-leaving radiance.
Validation	Checking satellite algorithm output against *in situ* measured reference values.
Zenith	An imaginary point directly above a particular location.

A more comprehensive list of optical and ocean colour remote sensing terms can be found at: `http://www.oceanopticsbook.info/view/references/brief_definitions`.

Appendix 4: Model Terminology and Acronyms

Term/Acronym	Definition
1D/3D	1-Dimensional model usually includes a single water column. 3-Dimensional model includes many water columns that have advection/diffusion between them.
Adjoint	In linear algebra, the transpose (interchange of rows and columns), of a linear map between two fields. An adjoint model is a component of a variational data assimilation system.
Climate Model	A coupled atmosphere/terrestrial/ocean model developed to study climate from sub-seasonal to centennial (sometimes even millennial) timescales.
Climatology	Modeller: Multiyear mean of a state variable. Ocean Colour Community: Composite of all available ocean colour data over a certain time period (e.g., months, year, multi-year).
CMIP	Coupled Model Intercomparison Project.
Data Assimilation	Formal integration of observations into model simulations.
Detrital Matter	Dead organisms at a number of stages of breakdown, sometimes referred to as POC in models.
Ensemble Kalman Filter	A sequential data assimilation algorithm that estimates the uncertainty in the model from the statistical distribution of an ensemble of model simulations.
Earth System Model (ESM)	Integrated model of the earth system components, including cryosphere, ocean, atmosphere, land, and may include ocean biogeochemistry. Generally, for centennial and global scale simulations.
Forcing Fields	Fields (e.g., solar radiation, wind, iron dust) that are used to drive a model (see Chapter 3).
Future projections	Model derived estimates of future climate (usually century timescales) under alternative scenarios.
Future predictions	Model derived estimate of climate in the nearer term (e.g., seasonal to decadal).
Grid	Model is spatially broken down in a collection of boxes, the equations (see Box 3.1) are applied to each box.
Hindcast Model (retrospective simulation)	Model developed to study past ocean conditions (often over last several decades); typically forced by atmospheric re-analyses.
Initial Conditions	Initial values of the model state variables (e.g., nitrate concentration, see Chapter 3.2.3).
Inverse Model	A model that uses the outcome of a suite of processes to calculate process rates. (In marine optics this specifically means the calculation of IOPs from observed AOPs).
IPCC	Intergovernmental Panel on Climate Change.
Nowcast/Forecast	Nowcast is estimation of the present state of the marine ecosystem. Forecast is prediction of the future state of the marine ecosystem.
NPZD	Nutrient-Phytoplankton-Zooplankton-Detritus (a typical ecosystem model configuration, see Section 3.1 and Fig. 3.2).

Online/offline	Online: When physical and biogeochemical equations are solved in the same model simulation, and therefore can (but not necessarily) include feedbacks between biogeochemistry and physics. Offline: When previous estimated physical fields are used to drive a biogeochemical model (see Section 3.4).
Operational model	Provides routinely nowcast and forecast (normally for a short term, from a few days to a few months) of ocean conditions, possibly including biogeochemical properties, delivered on a regular basis to inform time critical decisions (see Section 3.4).
OSE	Observing system simulations: Simulation to evaluate the effect of adding or removing individual components of an existing observing system.
OSSE	Observing System Simulation Experiment: Simulation where model generated data are assimilated into a model (often the same model) to assess and compare the performance of alterative (usually future) monitoring networks (see Section 8.4).
Phytoplankton Functional Type (PFT)	Aggregation of phytoplankton that share similar biogeochemical or ecological roles (e.g., silica users, nitrogen fixers etc., see Section 3.1).
RCP	Representative Concentration Pathways: Emissions Scenarios used by the IPCC.
Spinup	Initial part of simulation that allows physical, biogeochemical and ecosystem to come into consistency with each other (see Section 3.2.3).
State estimate	Estimate of ocean state often derived from synthesizing models and multiple observation.
State variable (prognostic variables)	Model quantities that change over time (e.g., temperature, nutrient pool, phytoplankton biomass, see Section 3.2.1).
Re-analysis	A long term model simulation that uses consistent data assimilation algorithms (possibly with multiple data sets), forcing, and boundary conditions.
Resolution – spatial	Distance between centers of model grid cells. A coarse resolution model is typically about 100 km between grid centers laterally and 10 m vertically (see Section 3.2.2).
Tendency term	Change in state variable with time — used for the model integration (see Section 3.2.1, Box 3.1).
Timestep	Time between changes in model state variables. Note: Resolution of save model output will likely be at a longer time average than the timestep (see Section 3.2.3).

Bibliography

Aas, E. (1987). Two-stream irradiance model for deep waters. *Appl. Opt.* 26: 2095–2101.

Abbott, M. R. and P. M. Zion (1985). Satellite observations of phytoplankton variability during an upwelling event. *Cont. Shelf Res.* 4:6, 661–680.

Ackleson, S. G., W. M. Balch, and P. M. Holligan (1994). Response of water-leaving radiance to particulate calcite and chlorophyll a concentrations: A model for Gulf of Maine coccolithophore blooms. *J. Geophys. Res.* 99: DOI: 10.1029/93JC02150.

Ahmad, Z. et al. (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. *Appl. Opt.* 49: 5545–5560.

Aiken, J., J. Fishwick, S. Lavender, R. Barlow, G. Moore, and H. Sessions (2007). Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October, 2002: preliminary validation of new products for phytoplankton function types and photosynthetic parameters. *Int. J. Remote Sens.* 28:3–4, 497–516.

Allen, A. L., C. W. Brown, A. J. Lewitus, and P. A. Sandifer (2015). The roles of emerging technology and modeling techniques in operational ecological forecasting at NOAA. *Mar. Technol. Soc. J.* 49:2, 193–203.

Allen, J. I., J. T. Holt, J. Blackford, and R. Proctor (2007). Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM. *J. Mar. Syst.* 68:3–4, 381–404.

Allen, J. I., T. J. Smyth, J. R. Siddorn, and M. Holt (2008). How well can we forecast high biomass algal bloom events in a eutrophic coastal sea? *Harmful Algae* 8:1, 701176.

Alvain, S., C. Moulin, Y. Dandonneau, and F. Breon (2005). Remote sensing of phytoplankton groups in case 1 waters for global SeaWiFS imagery. *Deep-Sea Res. I* 52: 1989–2004.

Alvain, S., C. Moulin, Y. Dandonneau, and H. Loisel (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochem. Cycles* 22: GB3001. DOI: 10.1029/2007GB003154.

Anderson, A. R. (2005). Plankton functional type modelling: running before we can walk? *J. Plankton Res.* 27: 1073–1081. DOI: 10.1093/plankt/fbi076.

Antoine, D. and A. Morel (1996). Oceanic primary production: 1. Adaptation of a spectral light-photosynthesis model in view of application to satellite chlorophyll observations. *Global Biogeochem. Cycles* 10: 43–55.

Antoine, D., A. Morel, H. Gordon, V. Banzon, and R. Evans (2005). Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. *J. Geophys. Res.* 110: DOI: 1029/2004JC002620.

Arakawa, A. and V. Lamb (1977). Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods Comput. Phys.* 17: 173–265. DOI: 10.1016/B978-0-12-460817-7.50009-4.

Arhonditsis, G. B. and M. T. Brett (2004). Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Mar. Ecol. Prog. Ser.* 271: 13–26.

Armstrong, R. A., J. L. Sarmiento, and R. D. Slater (1995). Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models. In: *Ecological Time Series*. Ed. by T. Powell and J. Steele. Chapman and Hall, New York, 371–390.

Artioli, Y. et al. (2012). The carbonate system in the North Sea: Sensitivity and model validation. *J. Mar. Syst.* 102: 1–13.

Artioli, Y. et al. (2014). Heterogeneity of impacts of high CO_2 on the North Western European Shelf. *Biogeosciences* 11:3, 601–612.

Aumont, O. and L. Bopp (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochem. Cycles* 20: GB2017. DOI: 10.1029/2005GB002591.

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen (2015). PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev. Discuss.* 8:2.

Aumont, O., E. Maier-Reimer, S. Blain, and P. Monfray (2003). An ecosystem model of the global ocean including Fe, Si, P colimitations. *Glob. Biogeochem. Cycles* 17:2, 1060.

Aumont, O., O. Maury, S. Lefort, and L. Bopp (2018). Evaluating the potential impacts of the diurnal vertical migration by marine organisms on marine biogeochemistry. *Glob. Biogeochem. Cycles* 32:11, 1622–43.

Austin, R. W. and T. J. Petzold (1981). The determination of the diffuse attenuation coefficient of sea water using the Coastal Zone Color Scanner. In: *Oceanography from Space*. Ed. by J. F. R. Gower. Springer, New York, 239–256.

Babin, M., J. C. Therriault, L. Legendre, and A. Condal (1993). Variations in the specific absorption coefficient for natural phytoplankton assemblages: Impact on estimates of primary production. *Limnol. Oceanogr.* 38: 154–177.

Bagniewski, W., K. Fennel, M. Perry, and E. D'Asaro (2011). Optimizing models of the North Atlantic spring bloom using physical, chemical and bio-optical observations from a Lagrangian float. *Biogeosciences* 8: 1291–1307.

Bailey, S. W., B. A. Franz, and P. J. Werdell (2010). Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. *Opt. Express* 18: 7521–7527.

Baird, M. E. (2010). Limits to prediction in a size-resolved pelagic ecosystem model. *J. Plankton Res.* 32: 1131–1146.

Baird, M. E., P. J. Ralph, K. Wild-Allen, F. Rizwi, and A. D. L. Steven (2013). A dynamic model of the cellular carbon to chlorophyll ratio applied to a batch culture and a continental shelf ecosystem. *Limnol. Oceanogr* 58: 1215–1226.

Baird, M. et al. (2016). Remote-sensing reflectance and true colour produced by a coupled hydrodynamic, optical, sediment, biogeochemical model of the Great Barrier Reef, Australia: Comparison with satellite data. *Environ. Modell. Software* 78: 79–96. DOI: 10.1016/j.envsoft.2015.11.025.

Bakker, D. C. E. et al. (2014). An update to the Surface Ocean CO_2 Atlas (SOCAT version 2). *Earth Syst. Sci. Data* 6:1, 69–90.

Balch, W. M., H. R. Gordon, B. C. Bowler, D. T. Drapeau, and E. S. Booth (2005). Calcium carbonate measurements in the surface global ocean based on Moderate-Resolution Imaging Spectroradiometer data. *J. Geophys. Res. Oceans* 110:7, 11121. DOI: 10.1029/2004JC002560.

Barber, R. T. and A. K. Hilting (2002). *History of the Study of Plankton Productivity*. Ed. by P. J. B. Williams, D. N. Thomas, and C. Reynolds. Wiley, 16–43.

Baretta, J. W., W. Ebenh, and P. Ruardij (1995). The European regional seas ecosystem model, a complex marine ecosystem model. *Neth. J. Sea Res.* 33:3–4, 233–246.

Baretta-Bekker, J. G., J. W. Baretta, and W. Ebenh (1997). Microbial dynamics in the marine ecosystem model ERSEM II with decoupled carbon assimilation and nutrient uptake. *J. Sea Res.* 38:3–4, 195–211.

Barton, A. D., C. H. Greene, B. C. Monger, and A. J. Pershing (2003). The Continuous Plankton Recorder survey and the North Atlantic Oscillation: Interannual- to multidecadal-scale patterns of phytoplankton variability in the North Atlantic Ocean. *Prog. Oceanogr.* 58:2–4, 337–358.

Barton, A. D. et al. (2016). *Report on the Trait-based approaches to ocean life scoping workshop, October 5-8, 2015, U.S.* Ocean Carbon and Biogeochemistry Program. DOI: 10.1575/1912/8017.

Basu, S., D. F. Baker, F. Chevallier, P. K. Patra, J. Liu, and J. B. Miller (2018). The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2, *Atmos. Chem. Phys.* 18:10, 7189.

Beaulieu, C. et al. (2013). Factors challenging our ability to detect long-term trends in ocean chlorophyll. *Biogeosciences* 10(4): 2711–2724. DOI: 10.5194/bg-10-2711-2013.

Becker, E., H. V. den Dool, and Q. Zhang (2014). Predictability and forecast skill in NMME. *J. Climate* 27:15, 5891–5906.

Behrenfeld, M. J. et al. (2009). Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences* 6: 779–794.

Behrenfeld, M., E. Boss, D. Siegel, and D. Shae (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochem. Cycles* 19: GB1006. DOI: 10.1029/2004GB002299.

Behrenfeld, M. and P. Falkowski (1997a). A consumer's guide to phytoplankton primary production models. *Limnol. Oceanogr.* 42: 1479–1491.

— (1997b). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.* 42:1, 1–20.

Ben Mustapha, Z., S. Alvain, C. Jamet, H. Loisel, and D. Dessailly (2014). Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters. *Remote Sens. Environ.* 146: 97–112. DOI: 10.1016/j.rse.2013.08.046.

Bengtsson, L. and J. Shukla (1988). Integration of space and *in situ* observations to study global climate change. *Bull. Am. Meteorol. Soc.* 69:10, 1130–1143.

Bennington, V., G. A. McKinley, S. Dutkiewicz, and D. Ulman (2009). What does chlorophyll variability tell us about export and air-sea CO_2 flux variability in the North Atlantic? *Global Biogeochem. Cycles* 23: GB3002. DOI: 10.1029/2008GB003241.

Berthon, J.-F., F. Mélin, and G. Zibordi (2008). Ocean colour remote sensing of the optically complex European seas. In: *Remote Sensing of the European Seas*. Ed. by V. Barale and M. Gade. Springer, 35–52.

Bidegare, R. R., L. V. Heukelem, and C. C. Trees (2002). HPLC phytoplankton pigments: sampling, laboratory methods, and quality assurance procedures. In: *Ocean Optics Protocols for Satellite Ocean Colour Sensor Validation, Revision 3,* ed. by M. J.L. and G. Fargion. Vol. NASA TM 2002-210004. NASA Goddard Space Flight Center, Greenbelt, Maryland: NASA. Chap. 17, 1258–268.

Binding, C. E., T. A. Greenberg, and R. P. Bukata (2012). An analysis of MODIS-derived algal and mineral turbidity in Lake Erie. *J. Great Lakes Res.* 38: 107–116.

Bissett, W. P., K. L. Carder, J. J. Walsh, and D. A. Dieterle (1999a). Carbon cycling in the upper waters of the Sargasso Sea: II. Numerical simulation of apparent and inherent optical properties. *Deep-Sea Res.I* 46: 271–317.

Bissett, W. P., S. DeBra, and D. Dye (2004). *Ecological Simulation (EcoSim) 2.0 Technical Description.* Technical Document FERI-2004-0002-U.D. Tampa, Fl: Florida Environmental Research Center.

Bissett, W. P., J. J. Walsh, and K. L. Dieterle D. A. Carder (1999b). Carbon cycling in the upper waters of the Sargasso Sea: I. Numerical simulation of differential carbon and nitrogen fluxes. *Deep Sea Res. I* 46:2, 205–269.

Bissett, W. P. et al. (2005). Predicting the optical properties of the West Florida Shelf: resolving the potential impacts of a terrestrial boundary condition on the distribution of coloured dissolved and particulate matter. *Mar. Chem.* 95:3–4, 199–233.

Bissinger, J. E., D. J. Montagnes, J. Harples, and D. Atkinson (2008). Predicting marine phytoplankton maximum growth rates from temperature: Improving on the Eppley curve using quantile regression. *Limnol. Oceanogr.* 53:2, 487–493.

Blackford, J. C., J. Allen, and F. J. Gilbert (2004). Ecosystem dynamics at six contrasting sites: a generic modelling study. *J. Mar. Syst.* 52: 191–215.

Blondeau-Patissier, D., V. E. Brando, K. Oubelkheir, A. G. Dekker, L. A. Clementson, and P. Daniel (2009). Bio-optical variability of the absorption and scattering properties of the Queensland inshore and reef waters, Australia. *J. Geophys. Res. : Oceans* 114: C05003.

Blondeau-Patissier, D., J. F. Gower, A. G. Dekker, S. R. Phinn, and V. E. Brando (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Prog. Oceanogr.* 123: 123–144. DOI: 10.1016/j.pocean.2013.12.008.

Blough, N. V. and R. Del Vecchio (2002). Chromophoric DOM in the coastal environment. In: *Biogeochemistry of Marine Dissolved Organic Matter*. Ed. by D. Hansell and C. Carlson. Academic Press, 509–546.

Blower, J. D. et al. (2009). Serving GODAE data and products to the ocean community. *Oceanography* 22: 70–79.

Bopp, L., O. Aumont, P. Cadule, S. Alvain, and M. Gehlen (2005). Response of diatoms distribution to global warming and potential implications: A global model study. *Geophys. Res. Lett.* 32: L19606. DOI: 10.1029/2005GL023653.

Bopp, L. et al. (2013). Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosciences* 10: 6225–6245.

Box, G. E. P. (1976). Science and statistics. *J. Am. Stat. Assoc.* 71: 791–799. DOI: 10.1080/01621459.1976.10480949.

Bracher, A., M. Vountas, T. Dinter, J. P. Burrows, R. Roettgers, and I. Peeken (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences* 6:5, 751–764.

Bracher, A. et al. (2017). Obtaining phytoplankton diversity from ocean colour: A scientific roadmap for future development. *Front. Mar. Sci.* 4: 195.

Brewin, R. J. W., E. Devred, S. Sathyendranath, S. Lavender, and N. Hardman-Mountford (2011). Model of phytoplankton absorption based on three size classes. *Appl. Opt.* 50:22, 4535–4549.

Brewin, R. J. W., D. E. Raitsos, Y. Pradhan, and I. Hoteit (2013). Comparison of chlorophyll in the Red Sea derived from MODIS-Aqua and *in vivo* fluorescence. *Remote Sens. Environ.* 136: 218–224. DOI: 10.1016/j.rse.2013.04.018.

Brewin, R. J. W., S. Sathyendranath, T. Hirata, S. J. Lavender, R. M. Barciela, and N. J. Hardman-Mountford (2010). A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecol. Modell.* 221:11, 1472–1483. DOI: 10.1016/j.ecolmodel.2010.02.014.

Brewin, R. J. W., S. Sathyendranath, P. Lange, and G. Tilstone (2014). Comparison of two methods to derive the size-structure of natural populations of phytoplankton. *Deep Sea Res. I* 85: 72–79.

Brewin, R. J. W. et al. (2015). The Ocean Colour Climate Change Initiative. III. A round-robin comparison on in-water bio-optical algorithms. *Remote Sens. Environ.* 162: 271–294.

Brewin, R. J. et al. (2017a). Uncertainty in ocean-color estimates of chlorophyll for phytoplankton groups. *Front. Mar. Sci.* 4: 104.

Brewin, R. et al. (2017b). Modelling size-fractionated primary production in the Atlantic Ocean from remote sensing. *Prog. Oceanogr.* 158: 130–149.

Bricaud, A., A. M. Ciotti, and B. Gentili (2012). Spatial-temporal variations in phytoplankton size and coloured detrital matter absorption at global and regional scales, as derived from twelve years of SeaWiFS data (1998–2009). *Global Biogeochem. Cycles* 26: 1.

Bricaud, A. and A. Morel (1986). Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. *Appl. Opt.* 25: 571–580.

Brown, C. D. and H. T. Davis (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemom. Intell. Lab. Syst.* 80:1, 24–38.

Brown, C. and G. Podesta (1997). Remote sensing of coccolithophore blooms in the western South Atlantic Ocean. *Remote Sens. Environ.* 60: 83–91.

Brown, C. and J. Yoder (1994). Coccolithophorid blooms in the global ocean. *J. Geophys. Res.* 99:C4, 7467–7482.

Buitenhuis, E. T., T. Hashioka, and C. L. Quere (2013). Combined constraints on global ocean primary production using observations and models. *Glob. Biogeochem. Cycles* 27:3, 847–858.

Buitenhuis, E. T., P. van der Wal, and H. J. de Baar (2001). Blooms of *Emiliania huxleyi* are sinks of atmospheric carbon dioxide: A field and mesocosm study derived simulation. *Global Biogeochem. Cycles* 15:3, 577–587.

Bukata, R. P., J. H. Jerome, K. Y. Kondratyev, and D. V. Pozdnyakov (1995). *Optical Properties and Remote Sensing of Inland and Coastal Waters.* CRC Press, Boca Raton, FL, USA.

Bulgarelli, B., V. Kisselev, and G. Zibordi (2014). Simulation and analysis of adjacency effects in coastal waters: A case study. *Appl. Opt.* 53: 1523–1545.

Burnham, K. P. and D. Anderson (2003). *Model Selection and Multimodel Inference. A Pratical Information-Theoric Approach.* Springer.

Butenschön, M. et al. (2016). ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. *Geosci. Model Dev.* 9:4, 1293–1339.

Cabré, A., D. Shields, I. Marinov, and T. Kostadinov (2016). Phenology of size-partitioned phytoplankton carbon-biomass from ocean color remote sensing and CMIP5 models. *Frontiers Mar. Sci.* 3: 39. DOI: `10.3389/fmars.2016.00039`.

Cael, B. B., K. Bisson, and C. L. Follett (2018). Can rates of ocean primary production and biological carbon export be related through their probability distributions? *Glob. Biogeochem. Cycles* 32:6, 954–970.

Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *J. Geophys. Res. Oceans* 100:C7, 13237–13254. DOI: `10.1029/95JC00458`.

Campbell, J. et al. (2002). Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance. *Glob. Biogeochem. Cycles* 16:3. DOI: `10.1029/2001GB001444|`.

Carder, K. L., S. Hawes, K. Baker, R. Smith, R. Steward, and B. Mitchell (1991). Reflectance model for quantifying chlorophyll a in the presence of productivity degradation products. *J. Geophys. Res.* 96: 20599–20611.

Carmillet, V., J. Brankart, P. Brasseur, H. Drange, G. Evensen, and J. Verron (2001). A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical-biochemical model of the North Atlantic Ocean. *Ocean Modell.* 3:3, 167–192.

Carr, M. E. (2002). Estimation of potential productivity in Eastern Boundary Currents using remote sensing. *Deep-Sea Res. II* 49:1–3, 59–80.

Carr, M. E., M. A. M. Friedrichs, M. Schmeltz, M. Noguchi Aita, D. Antoine, and K. R. Arrigo (2006). A comparison of global estimates of marine primary production from ocean color. *Deep Sea Res. II* 53:5–7, 741–770. DOI: `10.1016/j.dsr2.2006.01.028`.

Cetinić, I., M. J. Perry, N. T. Briggs, E. Kallin, E. A. D'Asaro, and C. M. Lee (2012). Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *J. Geophys. Res.: Oceans* 117:C6. DOI: `10.1029/2011JC007771`.

Chai, F., R. C. Dugdale, T. H. Peng, F. P. Wilkerson, and R. T. Barber (2002). One dimensional ecosystem model of the equatorial Pacific upwelling system. Part I: Model development and silicon and nitrogen cycle. *Deep Sea Res. II* 49: 2713–2745.

Chavez, F. P., M. Messié, and J. T. Pennington (2011). Marine primary production in relation to climate variability and change. *Annu. Rev. Mar. Sci.* 3: 227–60. DOI: `10.1146/annurev.marine.010908.163917`.

Chelton, D. B., P. Gaube, M. G. Schlax, J. J. Early, and R. M. Samelson (2011). The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science* 334:6054, 328–332.

Cheung, W. W. L. et al. (2010). Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. *Glob. Chang. Biol.* 16:1, 24–35. DOI: 10.1111/j.1365-2486.2009.01995.x.

Cheung, W. W. et al. (2016). Building confidence in projections of the responses of living marine resources to climate change. *ICES J Mar Sci* 73:5, 1283–1296.

Chisholm, S. W. (1992). Phytoplankton size. In: *Primary Poductivity and Biogeochemical Cycles in the Sea*. Ed. by P. G. Falkowski and A. D. Woodhead. Plenum Press. Chap. 213-237.

Chomko, R. M. and H. R. Gordon (2001). Atmospheric correction of ocean color imagery: test of the spectral optimization algorithm with the sea-viewing wide field-of-view Sensor. *Appl. Opt.* 40: 2973–2984.

Christian, J. R., M. A. Verschell, R. Murtugudde, A. J. Busalacchi, and C. R. McClain (2002). Biogeochemical modelling of the tropical Pacific Ocean. II: Iron biogeochemistry. *Deep-Sea Res. II* 49:1-3, 545–565.

Christian, J., M. Verschell, R. Murtugudde, A. Busalacchi, and C. McClain (2001). Biogeochemical modelling of the tropical Pacific Ocean. I: Seasonal and interannual variability. *Deep Sea Res. II* 49:1-3, 509–543.

Ciavatta, S., S. Kay, S. Saux-Picart, M. Butenschon, and J. I. Allen (2016). Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem. *J. Geophys. Res. Oceans* 121: 1824–1845. DOI: 10.1002/2015JC011496.

Ciavatta, S., R. Torres, S. Saux-Picart, and J. I. Allen (2011). Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *J. Geophys. Res. Oceans* 116:C12.

Ciavatta, S. et al. (2014). Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Prog. Oceanogr.* 127: 74–95.

Ciavatta, S. et al. (2018). Assimilation of ocean color plankton functional types to improve marine ecosystem simulations. *J. Geophys. Res. Oceans* 123:2, 834–854.

Ciavatta, S. et al. (2019). Ecoregions in the Mediterranean Sea through the reanalysis of phytoplankton functional types and carbon fluxes. *J. Geophys. Res.: Oceans* 124: DOI: 10.1029/2019JC015128.

Ciotti, A. and A. Bricaud (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by coloured detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil. *Limnol. Oceanogr.: Methods* 4: 237–253.

Clark, J. S. et al. (2001). Ecological forecasts: an emerging imperative. *Science* 293:5530, 657–660.

Cole, H., S. Henson, A. Martin, and A. Yool (2012). Mind the gaps: The impact of missing data on the calculation of phytoplankton phenology metrics. *J. Geophys. Res. Oceans* 117: C08030. DOI: 10.1029/2012JC008249.

Coles, V. J. et al. (2017). Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* 358:6367, 1149–1154. DOI: 10.1126/science.aan5712.

Cory, R. M. and G. W. Kling (2018). Interactions between sunlight and microorganisms influence dissolved organic matter degradation along the aquatic continuum. *Limnol. Oceanogr. Lett.* 3: 102–116.

Courant, R., K. Friedrichs, and H. Lewy (1928). On the partial difference equations of mathematical physics. *IBM J. Res.Develop.* 11:2, 215–234. DOI: 10.1147/rd.112.0215,MR0213764,Zbl0145.40402.

Craig, S. E. et al. (2012). Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sens. Environ.* 119: 72–83.

Currie, J. C. et al. (2013). Indian Ocean Dipole and El Niño/Southern Oscillation impacts on regional chlorophyll anomalies in the Indian Ocean. *Biogeosciences* 10:10, 6677–6698.

Da, F., M. A. M. Friedrichs, and P. St-Laurent (2018). Impacts of atmospheric nitrogen deposition and coastal nitrogen fluxes on oxygen concentrations in Chesapeake Bay. *J. Geophys. Res. Oceans* 123: 5004–5025. DOI: 10.1029/2018JC014009.

Daines, S. J., J. R. Clark, and T. M. Lenton (2014). Multiple environmental controls on phytoplankton growth strategies determine adaptive responses of the N : P ratio. *Ecol. Lett.* 174: 414–425.

Dall'Olmo, G., A. A. Gitelson, and D. C. Rundquist (2003). Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophys. Res. Lett.* 30: 1938.

Daszykowski, M., K. Kaczmarek, Y. V. Heyden, and B. Walczak (2007). Robust statistics in data analysis-a review: basic concepts. *Chemom. Intell. Lab. Syst.* 85:2, 203–219.

Dee, D. P. et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.* 137:656, 553–597. DOI: 10.1002/qj.828.

Dekker, A. G. et al. (2011). Intercomparison of shallow water bathymetry, hydro-optics, and benthos mapping techniques in Australian and Caribbean coastal environments. *Limnol. Oceanogr.: Methods* 9: 396–425.

Denman, K. L. and M. R. Abbott (1988). Time evolution of surface chlorophyll patterns from cross spectrum analysis of satellite color images. *J. Geophys. Res. Oceans* 93:C6, 6789–6798.

— (1994). Time scales of pattern evolution from cross spectrum analysis of advanced very high resolution radiometer and coastal zone color scanner imagery. *J. Geophys. Res. Oceans* 99:C4, 7433–7442.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. Roy. Meteor. Soc.* 131: 3385–3396. DOI: 10.1256/qj.05.108.

Devred, E., S. Sathyendranath, V. Stuart, and T. Platt (2011). A three component classification of phytoplankton absorption spectra: Application to ocean-colour data. *Remote Sens. Environ.* 115: 2255–2266. DOI: 10.1016/j.rse.2011.04.025.

Devred, E. et al. (2013). Future retrievals of water column bio-optical properties using the Hyperspectral Infrared Imager (HyspIRI). *Remote Sens.* 5:12, 6812–6837.

deYoung, B., M. Heath, F. Werner, F. Chai, B. Megrey, and P. Monfray (2004). Challenges of modeling ocean basin ecosystems. *Science* 304:5676, 1463–1466. DOI: 10.1126/science.1094858.

Dierssen, H. M. (2010). Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate, *Proc. Natl. Acad. Sci. U.S.A.* 107:40, 17073–17078.

Doerffer, R. and J. Fisher (1994). Concentrations of chlorophyll, suspended matter, and gelbstoff in case II waters derived from satellite coastal zone color scanner data with inverse modeling methods. *J. Geophys. Res.* 99: 7475–7466.

Doerffer, R. and H. Schiller (2007). The MERIS Case 2 water algorithm. *Int. J. Remote Sens.* 28:3–4, 517–535. DOI: 10.1080/01431160600821127.

Doney, S. C., K. Lindsay, I. Fung, and J. John (2006). Natural variability in a stable, 1000-Yr global coupled climate-carbon cycle simulation. *J. Climate* 19: 3033–3054.

Doney, S. C. et al. (2009). Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *J. Mar. Syst.* 76:1–2, 95–112.

Dowd, M., E. Jones, and J. Parslow (2014). A statistical overview and perspectives on data assimilation for marine biogeochemical models. *Environmetrics* 25:4, 203–213.

Doxaran, D., J. M. Froidefond, S. Lavender, and P. Castaing (2002). Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations. *Remote Sens. Environ.* 81: 149–161.

Ducklow, H. (1999). The bacterial component of the oceanic euphotic zone. *FEMS Microb. Ecol.* 30: 1–10.

Dunne, J. P. et al. (2013). GFDL's ESM2 global coupled climate-Carbon Earth System Models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Climate* 26:7, 2247–2267. DOI: 10.5194/bg-10-2711-2013.

Dutkiewicz, S., M. Follows, J. Marshall, and W. W. Gregg (2001). Interannual variability of phytoplankton abundances in the North Atlantic. *Deep-Sea Res* 48:10, 2323–2344.

Dutkiewicz, S., M. Follows, and P. Parekh (2005). Interactions of the iron and phosphate cycles: A three-dimensional model study. *Global Biogeochem. Cycles* 19: GB1021.

Dutkiewicz, S., A. E. Hickman, and O. Jahn (2018). Modelling ocean colour derived Chlorophyll-a. *Biogeosciences* 15: 613–630. DOI: 10.5194/bg-15-613-2018.

Dutkiewicz, S., A. E. Hickman, O. Jahn, W. W. Gregg, C. B. Mouw, and M. J. Follows (2015a). Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences* 12: 4447–4481. DOI: 10.5194/bg-12-4447-2015.

Dutkiewicz, S., A. E. Hickman, O. Jahn, S. Henson, C. Beaulieu, and E. Moneir (2019). Ocean colour signature of climate change. *Nat. Commun.* 10: DOI: 10.1038/s41467-019-08457-x.

Dutkiewicz, S., J. Scott, and M. J. Follows (2013). Winners and losers: Phytoplankton biogeochemical and ecological changes in a warmer world. *Glob. Biogeochem. Cycles* 27: 463–477. DOI: 10.1002/gbc.20042.

Dutkiewicz, S. et al. (2015b). Impact of ocean acidification on the structure of future phytoplankton communities. *Nat. Clim. Chang.* 5: 1002–1006. DOI: 10.1038/nclimate2722.

Echevin, V., O. Aumont, J. Ledesma, and G. Flores (2008). The seasonal cycle of surface chlorophyll in the Peruvian upwelling system: A modelling study. *Prog. Oceanogr.* 79:2–4, 167–176.

Edwards, C. A., A. M. Moore, I. Hoteit, and B. D. Cornuelle (2015). Regional Ocean Data Assimilation. *Ann. Rev. Marine Sci.* 7:1, 21–42.

Eppley, R. W. (1972). Temperature and phytoplankton growth in the sea. *Fish. Bull.* 70:4, 1063–1085.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* 99: 10143–10162.

— (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynam.* 53:4, 343–367. DOI: 10.1007/s10236-003-0036-9.

Evers-King, H., S. Bernard, L. R. Lain, and T. Probyn (2014). Sensitivity in reflectance attributed to phytoplankton cell size: forward and inverse modelling approaches. *Opt. Express* 22:10, 11536–11551.

Evers-King, H. et al. (2017). Validation and intercomparison of ocean color algorithms for estimating particulate organic carbon in the oceans. *Front. Mar. Sci.* 4: 251.

Fasham, M. J. R., H. W. Ducklow, and S. M. McKelvie (1990). A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *J. Mar. Res.* 48: 591–639.

Fauchereau, N., A. Tagliabue, L. Bopp, and P. M. S. Monteiro (2011). The response of phytoplankton biomass to transient mixing events in the Southern Ocean. *Geophys. Res. Lett.* 38:17. DOI: 10.1029/2011GL048498.

Fennel, K., J. Wilkin, J. Levin, J. Moisan, J. O'Reilly, and D. Haidvogel (2006). Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget, *Glob. Biogeochem. Cycles* 20: GB3007. DOI: 10.1029/2005GB002456.

Fennel, K. et al. (2019). Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. *Front. Mar. Sci.* 6: 89. DOI: 10.3389/fmars.2019.00089.

Fichot, C. G. and R. Benner (2011). A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. *Geophys. Res. Lett.* 38:

Fishman, J., L. T. Iraci, J. Al-Saadi, K. Chance, F. Chavez, and M. Chin (2012). The United States next generation of atmospheric composition and coastal ecosystem measurements: NASA's Geostationary Coastal and Air Pollution Events (GEO-CAPE) mission. *Bull. Am. Meteorol. Soc.* 93:10, 1547–1566.

Fletcher, S. J. (2010). Mixed Gaussian-lognormal four-dimensional data assimilation. *Tellus A* 62:3, 266–287. DOI: 10.1111/j.1600-0870.2009.00439.x.

Fletcher, S. J. and M. Zupanski (2006a). A data assimilation method for log-normally distributed observational errors. *Q. J. Roy. Meteor. Soc.* 132: 2505–2519.

— (2006b). A hybrid normal and lognormal distribution for data assimilation. *Atmos. Sci. Lett.* 7: 43–46.

Flynn, K. (2005). Castles built on sand: Dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. *J. Plankton Res.* 27:12. DOI: 10.1093/plankt/fbi099.

Follows, M. J. and S. Dutkiewicz (2002). Meteorological modulation of the North Atlantic spring bloom. *Deep-Sea Res. II* 49:1–3, 321–344.

— (2011). Modelling diverse communities of phytoplankton. *Annu. Rev. Mar. Sci.* 3: 427–51.

Follows, M. J., S. Dutkiewicz, S. Grant, and S. Chisholm (2007). Emergent biogeography of microbial communities in a model ocean. *Science* 315: 1843–1846. DOI: 10.1126/science.1138544.

Fontana, C., P. Brasseur, and J. M. Brankart (2013). Toward a multivariate reanalysis of the North Atlantic ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data. *Ocean Sci.* 9: 37–56.

Ford, D. A., K. P. Edwards, D. Lea, R. M. Barciela, M. J. Martin, and J. Demaria (2012). Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Sci.* 8: 751–771.

Ford, D. and R. Barciela (2017). Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products. *Remote Sens. Environ.* 203: 40–54.

Ford, D., S. Kay, R. McEwan, I. Totterdell, and M. Gehlen (2018). Marine biogeochemical modelling and data assimilation for operational forecasting, reanalysis, and climate research. In: *New Frontiers in Operational Oceanography*. Ed. by E. Chassignet, A. Pascual, J. Tintoré, and J. Verron. GODAE OceanView, 625–652. DOI: 10.17125/gov2018.ch22.

Franks, P. J. S. (2002). NPZ models of plankton dynamics: Their construction, coupling to physics, and application. *J. Oceanogr.* 58: 379–387.

Friedland, K. D. et al. (2012). Pathways between primary production and fisheries yields of large marine ecosystems. *PLoS One* 7:1. DOI: 10.1371/journal.pone.0028945.

Friedrichs, M. A. M., R. R. Hood, and J. D. Wiggert (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. *Deep Sea Res. II* 53: 576–600.

Friedrichs, M. A. M. et al. (2009). Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *J. Mar. Syst.* 76:1–2, 113–133.

Friedrichs, M. A. et al. (2007). Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. *J. Geophys. Res. Oceans* 112:C8.

Friedrichs, M. (2002). The assimilation of SeaWiFS and JGOFS EqPac data into a marine ecosystem model of the central equatorial Pacific. *Deep Sea Res. II* 49: 289–319.

Friedrichs, M. A. M. and D. E. Kaufman (2019). Marine biogeochemical data assimilation. In: *Encyclopedia of Ocean Sciences (Third Edition)*. Ed. by J. K. Cochran, H. J. Bokuniewicz, and P. L. Yager. Third Edition. Oxford: Academic Press, 520–526. DOI: 10.1016/B978-0-12-409548-9.11261-8.

Frouin, R. and B. Pelletier (2015). Bayesian methodology for inverting satellite ocean-color data. *Remote Sens. Environ.* 159: 332–360.

Fujii, M., E. Boss, and F. Chai (2007). The value of adding optics to ecosystem models: a case study. *Biogeosciences* 4: 817–835. DOI: 10.5194/bg-4-817-2007.

Fujiwara, A., T. Hirawake, K. Suzuki, and S. I. Saitoh (2011). Remote sensing of size structure of phytoplankton communities using optical properties of the Chukchi and Bering Sea shelf region. *Biogeosciences* 8: 3567–3580. DOI: 10.5194/bg-8-3567-2011.

Garcia, R. A., Z. Lee, and E. J. Hochberg (2018). Hyperspectral shallow-water remote sensing with an enhanced benthic classifier. *Remote Sens.* 10: 147.

Gardner, W. D., A. V. Mishonov, and M. J. Richardson (2006). Global POC concentrations from in-situ and satellite data. *Deep Sea Res. II* 53: 718–740. DOI: 10.1016/j.dsr2.2006.01.029.

Garver, S. and D. Siegel (1997). Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation 1. Time series from the Sargasso Sea. *J. Geophys. Res.* (C8): 18607–18625.

Gehlen, M. R. et al. (2015). Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. *J. Oper. Oceanogr.* 8: 168–187. DOI: 10.1080/1755876X.2015.1022350.

Geider, R., H. L. MacIntyre, and T. M. Kana (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnol. Oceanogr.* 43: 679–694.

Geider, R., H. Macintyre, and T. Kana (1997). Dynamic model of phytoplankton growth and acclimation: Responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature. *Mar. Ecol. Prog. Ser.* 148: 187–200. DOI: 10.3354/meps148187.

Gelaro, R. et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* 30:14, 5419–5454.

Giering, R. and T. Kaminski (1998). Recipes for adjoint code construction. *ACM Trans. Math. Softw.* 24:4, 437–474. DOI: 10.1145/293686.293695.

Gitelson, A. (1992). The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *Int. J. Remote Sens.* 13: 3367–3373.

Glibert, P. M. et al. (2014). Vulnerability of coastal ecosystems to changes in harmful algal bloom distribution in response to climate change: projections based on model analysis. *Glob. Chang. Biol.* 20:12, 3845–3858.

GlobColour (2017). *GlobColour Product User Guide*. Tech. rep. GC-UM-ACR-PUG-01. European Space Agency.

Gnanadesikan, A. et al. (2006). GFDL's CM2 Global Coupled Climate Models. Part II: The Baseline Ocean Simulation. *J. Climate* 19:5, 675–697. DOI: 10.1175/JCLI3630.1. eprint: https://doi.org/10.1175/JCLI3630.1.

Gons, H. J. (2002). A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *J. Plankton Res.* 24:9, 947–951. DOI: 10.1093/plankt/24.9.947.

Gordon, H. R. (1997). Atmospheric correction of ocean color imagery in the Earth Observing System era. *J. Geophys. Res.* 102: 17081–17106.

Gordon, H. R. and D. K. Clark (1980). Remote sensing optical properties of a stratified ocean: an improved interpretation. *Appl. Opt.* 19:20, 3428–3430.

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 of ship determinations and CZCS estimates. *Appl. Opt.* 22: 20–36.

Gordon, H. R. and A. Morel (1983). *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*. Springer-Verlag.

Gordon, H., O. Brown, and M. Jacobs (1975). Computed relationship between the inherent and apparent optical properties of a flat homogeneous ocean. *Appl. Opt.* 14: 417–427.

Gordon, H. and M. Wang (1994). Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. *Appl. Opt.* 33: 443–452.

Gordon, H. et al. (1988). A semianalytic radiance model of ocean color. *J. Geophys. Res.* 93: 10909–10924.

Gower, J. F. R., R. Doerffer, and G. A. Borstad (1999). Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. *Int. J. Remote Sens.* 20:9, 1771–1786.

Graff, J. R., A. J. Milligan, and M. J. Behrenfeld (2012). The measurement of phytoplankton biomass using flow-cytometric sorting and elemental analysis of carbon. *Limnol. Oceanogr.: Methods* 10: 910–920.

Graff, J. R. et al. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Res. I* 102: 16–25. DOI: 10.1016/j.dsr.2015.04.006.

Gregg, W. W. (2002a). *A Coupled Ocean-Atmosphere Radiative Model for Global Ocean Biogeochemical Models*. NASA GSFC Technical Report Series on Global Modeling and Data Assimilation, 104606, 22. 104606. NASA GSFC Technical Report Series on Global Modeling and Data Assimilation, 104606, 22. NASA.

— (2002b). Tracking the SeaWiFS record with a coupled physical/biogeochemical/radiative model of the global oceans. *Deep-Sea Res. II* 49:1–3, 81–105.

— (2008). Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. *J. Mar. Syst.* 69:3, 205–225.

Gregg, W. W. and K. L. Carder (1990). A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnol. Oceanogr.* 35:8, 1657–1675.

Gregg, W. W. and N. W. Casey (2004). Global and regional evaluation of the SeaWiFS chlorophyll data set. *Remote Sens. Environ.* 93: 463–479.

— (2007a). Modeling coccolithophores in the global ocean. *Deep-Sea Res. II* 54: 447–477.

— (2007b). Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. *Remote Sens. Environ.* 111:1, 25–35.

— (2009). Skill assessment of a spectral ocean-atmosphere radiative model. *J. Mar. Syst.* 76: 49–63.

— (2010). Improving the consistency of ocean color data: A step toward climate data records. *Geophys. Res. Lett.* 37:4, 257–5. DOI: 10.1029/2009GL041893.

Gregg, W. W., N. W. Casey, and C. R. McClain (2005). Recent trends in global ocean chlorophyll. *Geophys. Res. Lett.* 32:3, L03606. DOI: 10.1029/2004GL021808.

Gregg, W. W. and M. E. Conkright (2002). Decadal changes in global ocean chlorophyll. *Geophys. Res. Lett.* 29:15, 1–4. DOI: 10.1029/2002GL014689.

Gregg, W. W. and C. S. Rousseaux (2014). Decadal trends in global pelagic ocean chlorophyll: A new assessment integrating multiple satellites, in situ data, and models. *J. Geophys. Res. Oceans* 119:9, 5921–5933. DOI: 10.1002/2014JC010158.

— (2016). Directional and spectral irradiance in ocean models: Effects on simulated global phytoplankton, nutrients, and primary production. *Front. Mar. Sci.* DOI: 10.3389/fmars.2016.00240.

— (2017). Simulating PACE global ocean radiances. *Front. Mar. Sci.* 4: DOI: 10.3389/fmars.2017.00060.

Gregg, W. W., C. S. Rousseaux, and A. F. Bryan (2017). Global trends in ocean phytoplankton: a new assessment using revised ocean colour data. *Remote Sens. Lett.* 8: 1102–1111. DOI: 10.1080/2150704X.2017.1354263.

Gregg, W. W. and J. J. Walsh (1992). Simulation of the 1979 spring bloom in the Mid-Atlantic Bight: A coupled physical/biological/optical model , *J. Geophys. Res.* 97:C4, 5723–5743. DOI: 10.1029/91JC03057.

Gregg, W., P. Ginoux, P. Schopf, and N. Casey (2003). Phytoplankton and iron: Validation of a global three-dimensional ocean biogeochemical model. *Deep Sea Res. II* 50: 3143–3169.

Griffies, S. M. et al. (2016). OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project, Geosci. *Model Dev.* 9: 3231–3296.

Groom, S. et al. (2019). Satellite Ocean Colour: Current Status and Future Perspective. *Front. Mar. Sci.* 6: 485. DOI: 10.3389/fmars.2019.00485.

Groom, S. et al. (2009). The Western English Channel observatory: Optical characteristics of station L4. *J. Mar. Syst.* 77:3, 278–295.

Gruber, N. et al. (2006). Eddy-resolving simulation of plankton ecosystem dynamics in the California Current System. *Deep Sea Res. I* 53:9, 1483–1516.

Grunert, B. K., C. B. Mouw, and A. B. Ciochetto (2019). Deriving inherent optical properties from decomposition of hyperspectral non-water absorption. *Remote Sens. Environ.* 225: 193–206. DOI: 10.1016/j.rse.2019.03.004.

GUM (2008). Evaluation of measurement data — Guide to the expression of uncertainty in measurements. *Joint Committee for Guides in Metrology, JCGM* 100: 134.

Halliwell, G. R., V. Kourafalou, M. L. Henaff, L. K. Shay, and R. Atlas (2015). OSSE impact analysis of airborne ocean surveys for improving upper-ocean dynamical and thermodynamical forecasts in the Gulf of Mexico. *Prog. Oceanogr.* 130: 32–46.

Hansell, D. A. (2013). Recalcitrant Dissolved Organic Carbon Fractions. *Ann. Rev. Mar. Sci.* 5:1, 421–445. DOI: 10.1146/annurev-marine-120710-100757.

Hashioka, T. et al. (2013). Phytoplankton competition during the spring bloom in four plankton functional type models. *Biogeosci.* 10: 6833–6850. DOI: 10.5194/bg-10-6833-2013.

Hasselmann, K. (1993). Optimal fingerprints for the detection of time-dependent climate change. *J. Climate* 6: 1957–1971.

Hazen, E. L. et al. (2018). A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. *Sci. Adv.* 4:5, 3001.

Heavens, N. G., D. S. Ward, and M. M. Natalie (2013). Studying and projecting climate change with Earth System Models. *Nature Education Knowledge* 4:5, 4.

Helms, J. R., A. Stubbins, E. M. Perdue, N. W. Green, H. Chen, and K. Mopper (2013). Photochemical bleaching of oceanic dissolved organic matter and its effect on absorption spectral slope and fluorescence. *Mar. Chem.* 155: 81–91.

Hemmings, J. C. P., M. Srokosz, P. Challenor, and M. J. R. Fasham (2002). Assimilating satellite ocean-colour observations into oceanic ecosystem models. *Philos. Trans. A. Math. Phys. Eng. Sci.* 361:1802, 33–9. DOI: 10.1098/rsta.2002.1104.

Henson, S. A., H. S. Cole, J. Hopkins, A. P. Martin, and A. Yool (2018). Detection of climate change driven trends in phytoplankton phenology. *Global Change Biol.* 24: e101–e111. DOI: 10.1111/gcb.13886.

Henson, S. A., J. P. Dunne, and J. L. Sarmiento (2009). Decadal variability in North Atlantic phytoplankton blooms. *J. Geophys. Res. Oceans* 114:C4.

Henson, S. A. et al. (2010). Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. *Biogeosciences* 7:2, 621–640. DOI: 10.5194/bg-7-621-2010.

Henson, S. A. et al. (2017). Rapid emergence of climate change in environmental drivers of marine ecosystems. *Nat. Commun.* 8: 14682.

Henson, S., C. Beaulieu, and R. Lampitt (2016). Observing climate change trends in ocean biogeochemistry: when and where. *Glob. Chang. Biol.* 22: 1561–1571. DOI: 10.1111/gcb.13152.

Henson, S., H. Cole, C. Beaulieu, and A. Yool (2013). The impact of global warming on seasonality of ocean primary production. *Biogeosciences* 10: 4357–4369. DOI: 10.5194/bg-10-4357-2013.

Hirata, T., J. Aiken, N. Hardman-Mountford, T. J. Smyth, and R. G. Barlow (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sens. Environ.* 112: 3153–3159. DOI: 10.1016/j.rse.2008.03.011.

Hirata, T., N. J. Hardman-Mountford, R. J. W. Brewin, J. Aiken, R. Barlow, and K. Suzuki (2011). Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences* 8:2, 311–327. DOI: 10.5194/bg-8-311-2011.

Hobday, A. J., C. M. Spillman, J. P. Eveson, and J. R. Hartog (2016). Seasonal forecasting for decision support in marine fisheries and aquaculture. *Fish. Oceanogr.* 25: 45–56.

Hofmann, E. E. and M. A. M. Friedrichs (2001). Biogeochemical data assimilation. In: *Encyclopedia of Ocean Sciences.* Ed. by J. Steele, S. Thorpe, and K. Turekian. Vol. 1. Academic Press, 302–308.

Hoge, F. E. and P. Lyon (1996). Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: an analysis of model and radiance measurement errors. *J. Geophys. Res.* 101: 16631–16648.

Holling, C. S. (1959). Some characteristics of simple types of predation and parasitism. *Can. Entomol.* 91:7, 385–398.

Holt, J. T. and I. D. James (2001). An *s* coordinate density evolving model of the northwest European continental shelf - 1, Model description and density structure. *J. Geophys. Res.* 106:C7, 14015–14034.

Hood, R. R., L. E. Beckley, and J. D. Wiggert (2017). Biogeochemical and ecological impacts of boundary currents in the Indian Ocean. *Prog. Oceanogr.* 156: 290–325. DOI: 10.1016/j.pocean.2017.04.011.

Hood, R. R. et al. (2006). Pelagic functional group modeling: progress, challenges and prospects. *Deep-Sea Res. II* 53: 459–512.

Hooker, S. B., W. E. Esaias, G. C. Feldman, W. W. Gregg, and C. R. McClain (1992). An overview of SeaWiFS and ocean color. In: *NASA Technical Memorandum 104566. SeaWiFS Technical report Series, vol. 1.* Ed. by S. Hooker and E. Firestone. NASA, Goddard Space Flight Center, Greenbelt, Maryland.

Hu, C., K. L. Carder, and F. E. Muller-Karger (2000). How precise are SeaWiFS ocean color estimates? Implications of digitizationnoise errors. *Remote Sens. Environ.* 76: 239–249.

Hu, C., Z. Lee, and B. Franz (2012). Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *J. Geophys. Res.* 117: C01011. DOI: 01010.01029/02011JC007395.

Hu, C., R. Luerssen, F. E. Muller-Karger, K. L. Carder, and C. A. Heil (2008). On the remote monitoring of *Karenia brevis* blooms on the west Florida shelf. *Cont. Shelf Res.* 28: 159–176.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck (2003). An overview of the North Atlantic Oscillation. In: *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*. American Geophysical Union, 1–35.

IOCCG (2000). *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters*. Ed. by S. Sathyendranath. Vol. No. 3. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG. DOI: 10.25607/OBP-95.

— (2006). *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*. Ed. by Z.-P. Lee. Vol. No. 5. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG. DOI: 10.25607/OBP-96.

— (2010). *Atmospheric Correction for Remotely-Sensed Ocean-Colour Products*. Ed. by M. Wang. Vol. No. 10. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG. DOI: 10.25607/OBP-101.

— (2012). *Mission Requirements for Future Ocean-Colour Sensors*. Ed. by C. R. McClain and G. Meister. Vol. No. 13. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG. DOI: 10.25607/OBP-104.

— (2014). *Phytoplankton Functional Types from Space*. Ed. by S. Sathyendranath. Vol. No. 15. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG. DOI: 10.25607/OBP-106.

— (2019). *Uncertainties in Ocean Colour Remote Sensing*. Ed. by F. Mélin. Vol. No. 18. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG. DOI: 10.25607/OBP-696.

IPCC Report (2014). *Climate Change 2013: The physical science basis*. Ed. by T. Stoker et al. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press,

Irwin, A. J. and M. J. Oliver (2009). Are ocean deserts getting larger? *Geophys. Res. Lett.* 36: L18609. DOI: 10.1029/2009GL039883.

Jackson, T., H. A. Bouman, S. Sathyendranath, and E. Devred (2011). Regional-scale change in diatom distribution in the Humbolt Current as revealed by remote sensing: implications for fisheries. *ICES J. Mar. Sci.* 68: 729–736. DOI: 10.1093/icesjms/fsq18.

Jackson, T., S. Sathyendranath, and F. Mélin (2017). An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. *Remote Sens. Environ.* 203: 152–161.

Jamet, C., T. Loisel, and D. Dessailly (2012). Retrieval of the spectral diffuse attenuation coefficient $K_d(\lambda)$ in open and coastal ocean waters using a neural network inversion. *J. Geophys. Res.* 117: C10023. DOI: 10010.11029/12012JC008076.

Jang, C. J., J. Park, T. Park, and S. Yoo (2011). Response of the ocean mixed layer depth to global warming and its impact on primary production: a case for the North Pacific Ocean. *ICES J. Mar. Sci.* 68: 996–1007.

Jiang, M.-S., F. Chai, R. Dugdale, F. Wilkerson, T.-H. Peng, and R. Barber (2003). A nitrate and silicate budget in the equatorial Pacific Ocean: a coupled physical–biological model study. *Deep Sea Res. II* 50:22–26, 2971–2996.

Johnson, R., P. G. Strutton, S. W. Wright, A. McMinn, and K. M. Meiners (2013). Three improved satellite chlorophyll algorithms for the Southern Ocean. *J. Geophys. Res. Oceans* 118:7, 3694–3703. DOI: 10.1002/jgrc.20270.

Jolliff, J. K. et al. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *J. Mar. Syst.* 76:1–2, 64–82.

Jolliffe, I. T. and D. B. Stephenson (2003). *Forecast Verification: A practitioner's Guide in Atmospheric Science*. John Wiley Sons.

Jones, E. M. et al. (2016). Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef. *Biogeosciences* 13:23, 6441–6469.

Jones, E., J. Parslow, and L. Murray (2010). A Bayesian approach to state and parameter estimation in a Phytoplankton-Zooplankton model. *Aust. Meteorol. Oceanogr. J.* 59:(SP), 7–16.

Kalnay, E. et al. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* 77:3, 437–472. DOI: 10.1175/1520-0477.

Kaufman, D. E., M. Friedrichs, J. Hemmings, and W. S. Jr. (2018). Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea. *Biogeosciences* 15: 73–90. DOI: 10.5194/bg-15-73-2018.

Kearney, K. A., C. Stock, K. Aydin, and J. L. Sarmiento (2012). Coupling planktonic ecosystem and fisheries food web models for a pelagic ecosystem: Description and validation for the subarctic Pacific. *Ecol. Modell.* 237: 43–62.

Keerthi, M. G. et al. (2016). Intraseasonal variability of mixed layer depth in the tropical Indian Ocean. *Climate Dyn.* 46:7–8, 2633–2655.

Kettle, H. and C. J. Merchant (2008). Modeling ocean primary production: sensitivity to spectral resolution of attenuation and absorption of light. *Progr. Oceanogr.* 78:2, 135. DOI: 10.1016/j.pocean.2008.04.002.

Kim, G. E., M.-A. Pradal, and A. Gnanadesikan (2015). Quantifying the biological impact of surface ocean light attenuation by colored detrital matter in an ESM using a new optical parameterization. *Biogeosciences* 12: 5119–5132. DOI: 10.5194/bg-12-5119-2015.

Kirtman, B. P. et al. (2014). The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Am. Meteorol. Soc.* 95:4, 585–601.

Kitidis, V., A. P. Stubbins, G. Uher, R. C. U. Goddard, C. S. Law, and E. M. S. Woodward (2006). Variability of chromophoric organic matter in surface waters of the Atlantic Ocean. *Deep-Sea Res. II* 53: 1666–1684. DOI: 10.1016/j.dsr2.2006.05.009.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2009). Retrieval of the particle size distribution from satellite ocean colour observations. *J. Geophys. Res. Oceans* 114:C09015. DOI: 10.1029/2009JC005303.

Kostadinov, T. S. et al. (2017). Inter-comparison of phytoplankton functional type phenology metrics derived from ocean colour algorithms and Earth system models. *Remote Sens. Environ.* 190: 162–177. DOI: 10.1016/j.rse.2016.11.014.

Kostadinov, T., D. Siegel, and S. Maritorena (2010). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. *Biogeosciences* 7: 3239–3257. DOI: 10.5194/bg-7-3239-2010.

Kostadinov, T., S. Milutinović, I. Marinov, and A. Cabré (2016). Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution. *Ocean Sci.* 12: 561–575. DOI: 10.5194/os-12-561-2016.

Kwiatkowski, L. et al. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nat. Clim. Chang.* 7:5, 355.

Lahoz, W. A. and P. Schneider (2014). Data assimilation: making sense of Earth observation. *Front. Environ. Sci.* 2: DOI: 10.3389/fenvs.2014.00016.

Lain, L. R., S. Bernard, and M. W. Matthews (2017). Understanding the contribution of phytoplankton phase functions to uncertainties in the water colour signal. *Opt. Express* 25:4, A151–A165.

Laufkötter, C. et al. (2015). Drivers and uncertainties of future global marine primary production in marine ecosystem models. *Biogeosciences* 12: 6955–6984.

Laurent, A., K. Fennel, D. S. Ko, and J. Lehrter (2018). Climate change projected to exacerbate impacts of coastal eutrophication in the northern Gulf of Mexico. *J. Geophys. Res. Oceans* 123: 3408–3426. DOI: 10.1002/2017JC013583.

St-Laurent, P. et al. (2017). Impacts of atmospheric nitrogen deposition on surface waters of the western North Atlantic mitigated by multiple feedbacks. *J. Geophys. Res. Oceans* 122: 8406–8426. DOI: 10.1002/2017JC013072.

Lawson, L., Y. Spitz, E. Hofmann, and R. Long (1995). A data assimilation technique applied to a predator-prey model. *Bull. Math. Biol.* 57:4, 593–616.

Le Quéré, C. et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob. Chang. Biol.* 11: 2016–2040. DOI: 10.1111/j.1365-2486.2005.1004.x.

Le Traon, P.-Y. et al. (2017). The Copernicus Marine Environmental Monitoring Service: Main Scientific Achievements and Future Prospects. *Mercator Ocean J.* 56:

Lee, Y. J. et al. (2015a). An assessment of phytoplankton primary productivity in the Arctic Ocean from satellite ocean color / *in situ* chlorophyll-a based models. *J. Geophys. Res. Oceans* 120:9, 6508–6541.

Lee, Y. J. et al. (2016). Net primary productivity estimates and environmental variables in the Arctic Ocean: An assessment of coupled physical biogeochemical models. *J. Geophys. Res. Oceans* 121:12, 8635–8669. DOI: 10.1007/BF02460785.

Lee, Z. P., R. Arnone, C. Hu, P. J. Werdell, and B. Lubac (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Appl. Opt.* 49: 369–381.

Lee, Z. P., K. L. Carder, and R. Arnone (2002). Deriving inherent optical properties from water color: A multi-band quasi-analytical algorithm for optically deep waters. *Appl. Opt.* 41: 5755–5772.

Lee, Z. P., K. L. Carder, C. D. Mobley, R. G. Steward, and J. S. Patch (1999). Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization. *Appl. Opt.* 38: 3831–3843.

Lee, Z. P., K. Carter, R. Arnon, and M. He (2007). Determination of primary spectral bands for remote sensing of aquatic environments. *Sensors* 7: 3428–2441.

Lee, Z. P., M. Darecki, K. L. Carder, C. O. Davis, D. Stramski, and W. J. Rhea (2005a). Diffuse attenuation coefficient of downwelling irradiance: An evaluation of remote sensing methods. *J. Geophys. Res. Oceans* 110:2, 1–9. DOI: 10.1029/2004JC002573.

Lee, Z. P., K.-P. Du, and R. Arnone (2005b). A model for the diffuse attenuation coefficient of downwelling irradiance. *J. Geophys. Res.* C02016: DOI: `10.1029/2004JC002275`.

Lee, Z. P. et al. (2015b). Secchi disk depth. A new theory and mechanistic model for underwater visibility. *Remote Sens. Environ.* 169: 139–149.

Lee, Z., J. Marra, M. J. Perry, and M. Kahru (2014). Estimating oceanic primary productivity from ocean colour remote sensing: A strategic assessment. *J. Mar. Syst.* 149: 50–59.

Lee, Z., S. Shang, and R. Stavn (2018). AOPs are not additive: On the biogeo-optical modeling of the diffuse attenuation coefficient. *Front. Mar. Sci.* 5: 8. DOI: `10.3389/fmars.2018.00008`.

Lefort, S., O. Aumont, L. Bopp, T. Arsouze, M. Gehlen, and O. Maury (2015). Spatial and body-size dependent response of marine pelagic communities to projected global climate change. *Glob. Chang. Biol.* 21:1, 154–164. DOI: `10.1111/gcb.12679`.

Legendre, L. and J. Michaud (1999). Chlorophyll-a to estimate the particulate organic carbon available as food to large zooplankton in the euphotic zone of oceans. *J. Plankton Res.* 21:11, 2067–2083. DOI: `10.1093/plankt/21.11.2067`.

Legendre, L. and F. Rassoulzadegan (1995). Plankton and nutrient dynamics in marine waters. *Ophelia* 41:1, 153–172. DOI: `10.1080/00785236.1995.10422042`.

Lehodey, P. et al. (2006). Climate variability, fish, and fisheries. *J. Clim.* 19:20, 5009–5030.

Leles, S. G. et al. (2018). Modelling mixotrophic functional diversity and implications for ecosystem function. *J. Plankton Res.* 40:6, 627–642. DOI: `10.1093/plankt/fby044`.

Lenton, A., R. J. Matear, and B. Tilbrook (2006). Design of an observational strategy for quantifying the Southern Ocean uptake of CO_2. *Glob. Biogeochem. Cycles* 20:4, GB4010. DOI: `10.1029/2005GB002620`.

Levy, M., O. Jahn, S. Dutkiewicz, and M. J. Follows (2014). Phytoplankton diversity and community structure affected by oceanic disperal and mesoscale turbulence. *L&O:F&E* 4: 67–84. DOI: `10.1215/21573869-2768549`.

Lewis, K. and J. I. Allen (2009). Validation of a hydrodynamic-ecosystem model simulation with time-series data collected in the western English Channel. *J. Mar. Syst.* 77:3, 296–311.

Loisel, H. and D. Stramski (2000). Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. *Appl. Opt.* 39: 3001–3011.

Longhurst, A. R. (1998). Ecological Geography of the Sea. *Academic Press, San Diego.* 397.

Lynch, D. R., D. J. McGillicuddy, and F. E. Werner (2009). Skill assessment for coupled biological/physical models of marine systems. *J. Mar. Syst.* 1:76, 1–3.

Maier-Reimer, E., I. Kriest, J. Segschneider, and P. Wetzel (2005). *The Hamburg Ocean Carbon Cycle Model HA-MOCC5.1 - Technical Description Release 1.1.* Report on Earth System Science, 14, Max Planck Institute for Meteorology, Hamburg, Germany, available from `http://www.mpimet.mpg.de`.

Majkut, J. D., B. R. Carter, T. L. Frölicher, C. O. Dufour, K. B. Rodgers, and J. L. Sarmiento (2014). An observing system simulation for Southern Ocean carbon dioxide uptake. *Phil.Trans. Math. Phys. Eng. Sci.* 372: 20130046.

Manizza, M., C. L. Quéré, A. Watson, and E. Buitenhuis (2005). Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model. *Geophys. Res. Lett.* 32:L05603.

Mannino, A., M. G. Novak, S. B. Hooker, K. Hyde, and D. Aurin (2014). Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast. *Remote Sens. Environ.* 152: 576–602.

Mannino, A., S. R. Signorini, M. G. Novak, J. Wilkin, M. A. M. Friedrichs, and R. G. Najjar (2016). Dissolved organic carbon fluxes in the Middle Atlantic Bight: An integrated approach based on satellite data and ocean model products. *J. Geophys. Res. Biogeosciences* 121:2, 312–336. DOI: `10.1002/2015JG003031`.

Margvelashvili, N., J. Andrewartha, M. Herzfeld, B. J. Robson, and V. E. Brando (2013). Satellite data assimilation and estimation of a 3D coastal sediment transport model using error-subspace emulators. *Environ. Modell. Software* 40: 191–201.

Marinov, I., S. C. Doney, I. D. Lima, K. Lindsay, J. K. Moore, and N. Mahowald (2013). North-South asymmetry in the modeled phytoplankton community response to climate change over the 21st century. *Global Biogeochem. Cycles* 27: 1274–1290. DOI: `10.1002/2013GB004599`.

Maritorena, S., D. A. Siegel, and A. R. Peterson (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Appl. Opt.* 41: 2705–2714.

Marra, J. (2002). Approaches to the measurement of plankton production. In: *Phytoplankton Productivity: Carbon Assimilation in Marine and Freshwater Ecosystems*. Ed. by P. J. Le, B. Williams, D. N. Thomas, and C. S. Reynolds. Blackwell Science Ltd.

Marshall, J. C., A. Adcroft, C. Hill, L. Perelman, and C. Heisey (1997a). A finite volume, incompressible Navier-Stokes model for studies of the oceans on parallel computers. *J. Geophys. Res.* 102: 5753–5766.

Marshall, J. C., C. Hill, L. Perelman, and A. Adcroft (1997b). Hydrostatic, quasi-hydrostatic and non-hydrostatic ocean modelling. *J. Geophys. Res.* 102: 5733–5752.

Martinez, E., D. Antoine, F. D'Ortenzio, and B. Gentili (2009). Climate-driven basin-scale decadal oscillations of oceanic phytoplankton. *Science* 326:5957, 1253–1256. DOI: 10.1126/science.1177012.

Martinez-Vicente, V., G. Dall'Olmo, G. Tarran, E. Boss, and S. Sathyendranath (2013). Optical backscattering is correlated with phytoplankton carbon across the Atlantic Ocean. *Geophys. Res. Lett.* 40: 1154–1158. DOI: 10.1002/grl.50252.

Martinez-Vicente, V. et al. (2017). Intercomparison of ocean color algorithms for picophytoplankton carbon in the ocean. *Front. Mar. Sci.* 4: 378.

Martiny, A., J. A. Vrugt, F. W. Primeau, and M. W. Lomas (2013). Regional variation in the particulate organic carbon to nitrogen ratio in the surface ocean. *Glob. Biogeochem. Cycles* 27:3, 723–731.

Mason, S. J. and D. B. Stephenson (2008). How do we know whether seasonal climate forecasts are any good? In: *Seasonal Climate: Forecasting and Managing Risk.* Springer, 259–289.

Masotti, I., C. Moulin, S. Alvain, L. Bopp, A. Tagliabue, and D. Antoine (2011). Large-scale shifts in phytoplankton groups in the Equatorial Pacific during ENSO cycles. *Biogeosciences* 8:3, 539–550.

Matear, R. J. (1995). Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P. *J. Mar. Res.* 53:4, 571–607. DOI: 10.1357/0022240953213098.

Matear, R. J. and E. Jones (2011). *Marine Biogeochemical Modelling and Data Assimilation.* Ed. by A. Schiller and G. B. Brassington. Operational Oceanography in the 21st Century.

Matsumoto, K. et al. (2004). Evaluation of ocean carbon cycle models with data-based metrics. *Geophys. Res. Lett.* 31: L07303. DOI: 10.1029/2003GL018970.

Mattern, J. P. and C. A. Edwards (2017). Simple parameter estimation for complex models — Testing evolutionary techniques on 3-dimensional biogeochemical ocean models. *J. Mar. Syst.* 165: 139–152. DOI: 10.1016/j.jmarsys.2016.10.012.

Mattern, J. P., C. A. Edwards, and A. M. Moore (2018). Improving variational data assimilation through background and observation error adjustments. *Monthly Weather Review* 146: 485–501.

Mattern, J. P., K. Fennel, and M. Dowd (2012). Estimating time-dependent parameters for a biological ocean model using an emulator approach. *J. Mar. Syst.* 96: 32–47.

Mattern, J. P., H. Song, C. A. Edwards, A. M. Moore, and J. Fiechter (2017). Data assimilation of physical and chlorophyll a observations in the California Current System using two biogeochemical models. *Ocean Modell.* 109: 55–71.

Matthews, M. W. and D. Odermatt (2015). Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. *Remote Sens. Environ.* 156: 374–382.

McDonald, C. P., V. Bennington, N. R. Urban, and G. A. McKinley (2012). 1-D test-bed calibration of a 3-D Lake Superior biogeochemical model. *Ecol. Modell.* 225: 115–126.

McGill, B. J., B. J. Enquist, E. Weiher, and M. Westoby (2006). Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* 21: 178–185.

McKee, D. et al. (2014). Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. *J. Geophys. Res.: Oceans* 119: 9013–9025. DOI: 10.1002/2014JC009909.

McKinley, G. A., A. Ritzer, and N. Lovenduski (2018). Mechanisms of northern North Atlantic biomass variability. *Biogeosciences* 15: 6049–6066.

McKinley, G., M. J. Follows, and J. C. Marshall (2004). Mechanisms of air-sea CO_2 flux variability in the Equatorial Pacific and the North Atlantic. *Glob. Biogeochem. Cycles* 18: DOI: 10.1029/2003GB002179.

Mcowen, C. J., W. W. Cheung, R. R. Rykaczewski, R. A. Watson, and L. J. Wood (2015). Is fisheries production within Large Marine Ecosystems determined by bottom-up or top-down forcing? *Fish and Fisheries* 16:4, 623–632.

Measures, C. I. and S. Vink (1999). Seasonal variations in the distribution of Fe and Al in the surface waters of the Arabian Sea. *Deep-Sea Res. II* 46:8–9, 1597–1622.

Mélin, F. (2016). Impact of inter-mission differences and drifts on chlorophyll-a trend estimates. *Int. J. Remote Sens.* 37: 2061–2079.

Mélin, F. and B. A. Franz (2014). Assessment of satellite ocean colour radiometry and derived geophysical products. In: *Optical Radiometry for Oceans Climate Measurements.* Ed. by G. Zibordi, C. Donlon, and A. Parr. Vol. 47. Experimental Methods in the Physical Sciences. Academic Press. Chap. 6, 609–638.

Menden-Deuer, S. and E. J. Lessard (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnol. Oceanogr.* 45: 569–579.

Menemenlis, D. et al. (2008). ECCO2: High resolution global ocean and sea ice data synthesis. *Mercator Ocean Quarterly Newsletter* 31: 13–21.

Menzel, D. W. and R. F. Vaccaro (1964). The measurement of dissolved organic and particulate carbon in seawater. *Limnol. Oceanogr.* 9: 138–142.

Merino, G. et al. (2012). Can marine fisheries and aquaculture meet fish demand from a growing human population in a changing climate? *Glob. Environ. Chang.* 22:4, 795–806.

Mishra, S. and D. R. Mishra (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ.* 117: 394–406.

Mobley, C. D. (1994). *Light and Water: Radiative Transfer in Natural Waters.* Academic.

— (2011). Fast light calculations for ocean ecosystem and inverse models. *Opt. Express* 19:20, 18927–18944.

Mobley, C. D., L. K. Sundamn, W. P. Bissett, and B. Cahill (2009). Fast and accurate irradiance calculations for ecosystem models. *Biogeosci. Discuss.* 5: 10624–10662.

Mobley, C. D. and L. K. Sundman (2008a). *Hydrolight 5 Technical Documentation.* Sequoia Scientific.

— (2008b). *Hydrolight 5 Users' Guide.* Sequoia Scientific.

Mobley, C. D., J. Werdell, B. Franz, Z. Ahmad, and S. Bailey (2016). *Atmospheric Correction for Satellite Ocean Color Radiometry.* NASA Tech. Memo. 2016-217551. NASA Goddard Space Flight Center, Greenbelt, Maryland.

Mobley, C. D. et al. (1993). Comparison of numerical models for computing underwater light fields. *Appl. Opt.* 32: 7484–7504. DOI: 10.1364/AO.32.007484.

Mobley, C. D. et al. (2005). Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. *Appl. Opt.* 44: 3576–3592.

Mobley, C., F. Chai, P. Xiu, and L. Sundman (2015). Impact of improved light calculations on predicted phytoplankton growth and heating in an idealized upwelling-downwelling channel geometry. *J. Geophys. Res. Oceans* 120: DOI: 10.1002/2014JC010588.

Moffett, J. W., J. Vedamati, T. J. Goepfert, A. Pratihary, M. Gauns, and S. W. A. Naqvi (2015). Biogeochemistry of Iron in the Arabian Sea. *Limnol. Oceanogr.* 60:5, 1671–1688. DOI: 10.1002/lno.10132.

Moline, M. A. and B. B. Prezelin (2000). Optical fractionation of chlorophyll and primary production for coastal waters of the Southern Ocean. *Polar Biol.* 23: 129–136.

Montes-Hugo, M. A., M. Vernet, R. Smith, and K. Carder (2008). Phytoplankton size-structure on the western shelf of the Antarctic Peninsula: A remote-sensing approach. *Int. J. Remote Sens.* 29: 801–829.

Moore, A. M. et al. (2019). Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean. *Front. Mar. Sci.* 6: 90.

Moore, G., J. Aiken, and S. Lavender (1999). The atmospheric correction of water colour and the quantitative retrieval of suspended particulate matter in Case II waters: application to MERIS. *Int. J. Remote Sensing* 20:9, 1713–1733.

Moore, J. K., S. C. Doney, and K. Lindsay (2004). Upper ocean ecosystem dynamics and iron cycling in a global 3D model. *Glob. Biogeochem. Cycles* 18: GB4028. DOI: 10.1029/2004gb002220.

Moore, J. K., K. Lindsay, S. C. Doney, M. C. Long, and K. Misumi (2013). Marine ecosystem dynamics and biogeochemical cycling in the Community Earth System Model [CESM1 (BGC)]: Comparison of the 1990s with the 2090s under the RCP4. 5 and RCP8. 5 scenarios. *J. Climate* 26:23, 9291–9312.

Moore, K. D., K. J. Voss, and H. R. Gordon (2000). Spectral reflectance of whitecaps: Their contribution to water-leaving radiance. *J. Geophys. Res.* 105: 6493–6499.

Moore, T. S., J. W. Campbell, and M. D. Dowell (2009). A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sens. Environ.* 113:11, 2424–2430.

Moore, T. S., J. W. Campbell, and H. Feng (2001). A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Trans. Geosci. Remote Sens.* 39:8, 1764–1776.

Moore, T. S., M. D. Dowell, S. Bradt, and A. R. Verdu (2014). An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Remote Sens. Environ.* 143: 1–15. DOI: 10.1016/j.rse.2013.11.021.

Morel, A. (1980). In-water and remote measurements of ocean color. *Boundary Layer Meteorol.* 18: 177–201.

— (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 waters). *J. Geophys. Res.* 93: 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. *Appl. Opt.* 41:30, 6289–6306.

Morel, A. and B. Gentili (1993). Diffuse reflectance of oceanic waters (2): Bi-directional aspects. *Appl. Opt.* 32: 6864–6879.

Morel, A. and S. Maritorena (2001). Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.* 106: 7163–7180.

Moses, W. J., A. A. Gitelson, S. Berdnikov, V. Saprygin, and V. Povazhnyi (2012). Operational MERIS-based NIR-red algorithms for estimating chlorophyll-a concentrations in coastal waters — The Azov Sea case study. *Remote Sens. Environ.* 121: 118–124.

Mouw, C. B., J. A. Yoder, and S. C. Doney (2012). Impact of phytoplankton community size on a linked global ocean optical and ecosystem model. *J. Mar. Syst.* 89: 61–75. DOI: 10.1016/j.jmarsys.2011.08.002.

Mouw, C. B. et al. (2015). Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* 160: 15–33. DOI: 10.1016/j.rse.2015.02.001.

Mouw, C. B. et al. (2017). A consumer's guide to satellite remote sensing of multiplephytoplankton groups in the global ocean. *Front. Mar. Sci.* 4:41, 497–19. DOI: 10.3389/fmars.2017.00041.

Mouw, C. and J. Yoder (2010). Optical determination of phytoplankton size composition from global SeaWiFS imagery. *J. Geophys. Res.* 115:C12018. DOI: 10.1029/2010JC006337.

Mueller, J. L. (2000). *SeaWiFS algorithm for the diffuse attenuation coefficient, K(490), using water-leaving radiances at 490 and 555 nm.* Sea-WiFS Postlaunch Calibration and Validation Analyses, part 3, pp.24–27. NASA Goddard Space Flight Cent., Greenbelt, MD, USA.

Mueller, J. L., R. R. Bidigare, C. T. W. M. B. J. D. D. Drapeau, D. K. L. V. Heukelem, and J. Perl (2003). *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation.* NASA/TM-2003 Revision 5, Volume V. NASA Goddard Space Flight Space Center.

Nakicenovic, N. and R. Swart, eds. (2000). *Emissions Scenarios: IPCC Special Report.* Cambridge University Press.

NASA (2014). *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Ocean Color Data.* Tech. rep. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, NASA OB.DAAC. DOI: 10.5067/ORBVIEW-2/SEAWIFS_OC.2014.0.

National Academies of Sciences, E. and Medicine (2018). *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space.* Washington, DC: The National Academies Press. ISBN: 978-0-309-46757-5. DOI: 10.17226/24938.

Natvik, L. J. and G. Evensen (2003). Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments. *J. Mar. Syst.* 40: 127–153.

Nelson, N. B., D. A. Siegel, C. A. Carlson, and C. M. Swan (2010). Tracing global biogeochemical cycles and meridional overturning circulation using chromophoric dissolved organic matter. *Geophys. Res. Lett.* 37: L03610. DOI: 10.1029/2009GL042325.

Nelson, N. and D. Siegel (2013). The global distribution and dynamics of chromophoric dissolved organic matter. *Mar. Sci.* 5: 447–476. DOI: 10.1146/annurev-marine-120710-100751.

Nerger, L. and W. Gregg (2007). Assimilation of SeaWiFS data into a global ocean biogeochemical model using a local SEIK filter. *J. Mar. Syst.* 68: 237–254.

Oke, P. R. and P. Sakov (2008). Representation error of oceanic observations for data assimilation. *J. Atmos. Oceanic Technol.* 25:6, 1004–1017.

Oke, P. R. and A. Schiller (2007). Impact of Argo, SST, and altimeter data on an eddy resolving ocean reanalysis. *Geophys. Res. Lett.* 34:19, L19601.

Oke, P. R. et al. (2015). Assessing the impact of observations on ocean forecasts and reanalyses: Part 2, Regional applications. *J. Oper. Oceanogr.* 8: s63–s79. DOI: 10.1080/1755876X.2015.1022080.

Ono, T., T. Saino, N. Kurita, and K. Sasaki (2004). Basin-scale extrapolation of shipboard pCO_2 data by using satellite SST and Chl a. *Int. J. Remote Sens.* 25:19, 3803–3815.

O'Reilly, J. E. et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.* 103: 24937–24953.

O'Reilly, J. E. et al. (2000). *SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3.* Ed. by S. B. Hooker and E. R. Firestone. Vol. NASA Tech. Memo. 2000-206892.

Orr, J. C. et al. (2017). Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geosci. Model Dev.* 10: 2169–2199.

Oschlies, A. and M. Schartau (2005). Basin-scale performance of a locally optimized marine ecosystem model. *J. Mar. Res.* 63:2, 335–358.

OSPAR (2013). *Common Procedure for the Identification of the Eutrophication Status of the OSPAR Maritime Area.* Tech. rep. OSPAR Commission.

PACE (2012). *Pre-Aerosol, Clouds, and ocean Ecosystem (PACE) Mission Science Definition Team Report.* Tech. rep. PACE Mission Science Definition Team.

Pahlevan, N., Z. Lee, J. Wei, C. B. Schaaf, J. R. Schott, and A. Berk (2014). On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic remote sensing. *Remote Sens. Environ.* 154:C, 272–284. DOI: 10.1016/j.rse.2014.08.001.

Pahlow, M. and A. Oschlies (2009). Chain model of phytoplankton P, N and light colimitation. *Mar. Ecol. Prog. Ser.* 376: 69–83.

Palacz, A. P., M. A. St. John, R. J. W. Brewin, T. Hirata, and W. W. Gregg (2013). Distribution of phytoplankton functional types in high-nitrate low-chlorophyll waters in a new diagnostic ecological indicator model. *Biogeosciences* 10: 7553–7574. DOI: 10.5194/bg-10-7553-2013.

Palmer, J. R. and I. J. Totterdell (2001). Production and export in a global ocean ecosystem model. *Deep Sea Res. I* 48: 1169–1198.

Palmer, S. J., P. Hunter, T. Lankester, S. Hubbard, E. Spyrakos, and A. Tyler (2015). Validation of Envisat MERIS algorithm for chlorophyll retrieval in a large, turbid and optically-complex shallow lake. *Remote Sens. Environ.* 157:C, 158–169.

Park, J. Y. et al. (2018). Modeling global ocean biogeochemistry with physical data assimilation: A pragmatic solution to the equatorial instability. *J. Adv. Model. Earth Syst.* 10:3, 891–906.

Park, J.-Y., C. A. Stock, J. Dunne, X. Yang, and A. Rosati (2019). Seasonal to multiannual marine ecosystem prediction with a global Earth system model. *Science* 365:6450, 284–288. DOI: 10.1126/science.aav6634.

Parslow, J., N. Cressie, E. P. Campbell, E. Jones, and L. Murray (2013). Bayesian learning and predictability in a stochastic nonlinear dynamical model. *Ecol. Appl.* 23:4, 679–698.

Pastor, M. V., J. B. Palter, J. L. Pelegri, and J. Dunne (2013). Physical drivers of interannual chlorophyll variability in the eastern subtropical North Atlantic. *J. Geophys. Res. Oceans* 118:8, 3871–3886.

Patara, L., M. Visbeck, S. Masina, G. Krahmann, and M. Vichi (2011). Marine biogeochemical responses to the North Atlantic Oscillation in a coupled climate model. *J. Geophys. Res. Oceans* 116: C07023. DOI: 10.1029/2010JC006785.

Pelegri, J. L., A. Marrero-Diaz, and A. W. Ratsimandresy (2006). Nutrient irrigation of the North Atlantic. *Prog. Oceanogr.* 70:2–4, 366–406.

Platt, T. (1986). Primary production of the ocean water column as a function of surface light intensity: algorithms for remote sensing. *Deep Sea Res. A* 33: 149–163.

Platt, T., C. Fuentes-Yaco, and K. T. Frank (2003). Marine ecology: spring algal bloom and larval fish survival. *Nature* 423:6938, 398.

Polimene, L. et al. (2018). Modelling marine DOC degradation time scales. *Natl. Sci. Rev.* 5:4, 468–474.

Polovina, J. J., J. P. Dunne, P. A. Woodworth, and E. A. Howell (2011). Projected expansion of the subtropical biome and contraction of the temperate and equatorial upwelling biomes in the North Pacific under global warming. *ICES J. Mar. Sci.* 68:6, 986–995.

Polovina, J. J., E. A. Howell, and M. Abecassis (2008). Ocean's least productive waters are expanding. *Geophys. Res. Lett.* 35:3, L03618. DOI: 10.1029/2007GL031745|.

Powell, T. M., C. V. Lewis, E. N. Curchitser, D. B. Haidvogel, A. J. Hermann, and E. L. Dobbins (2006). Results from a three dimensional, nested biological physical model of the California Current System and comparisons with statistics from satellite imagery. *J. Geophys. Res. Oceans* 111:C7.

Pradhan, H. K., C. Völker, S. N. Losa, A. Bracher, and L. Nerger (2019). Assimilation of global total chlorophyll OC-CCI data and its impact on individual phytoplankton fields. *J. Geophys. Res.: Oceans* 124:1, 470–490.

Quay, P., J. Stutsman, and T. Steinhoff (2012). Primary production and carbon export rates across the subpolar N. Atlantic Ocean basin based on triple oxygen isotope and dissolved O_2 and Ar gas measurements. *Global Biogeochem. Cycles* 26: DOI: 10.1029/2010GB004003.

Radenac, M. H. et al. (2001). Modeled and observed impacts of the 1997–1998 El Niño on nitrate and new production in the equatorial Pacific. *J. Geophys. Res. Oceans* 106:C11, 26879–26898.

Raicich, F. and A. Rampazzo (2003). Observing system simulation experiments for the assessment of temperature sampling strategies in the Mediterranean Sea. In: *Annales Geophysicae, Copernicus GmbH*.

Raitsos, D., S. Lavender, C. Maravelias, J. Haralabous, A. Richardson, and P. Reid (2008). Identifying four phytoplankton functional types from space: An ecological approach. *Limnol. Oceangr.* 53: 605–613.

Randall, D. A. et al. (2007). Climate models and their evaluation. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Ed. by S. [Solomon et al. Cambridge University Press, 589–662.

Redfern, J. V. et al. (2013). Assessing the risk of ships striking large whales in marine spatial planning. *Conserv. Biol.* 27:2, 292–302.

Redfield, A. C. (1934). On the proportions of organic derivations in sea water and their relation to the composition of plankton. In: *James Johnstone Memorial Volume.* Ed. by R. Daniel. University Press of Liverpool, 176–192.

Redfield, A. C., B. H. Ketchum, and F. A. Richards (1963). The Influence of Organisms on the Composition of the Sea Water. In: *The Sea, Vol. 2.* Ed. by M. Hill. Interscience Publishers, 26–77.

Regaudie-de-Gioux, A., S. Lasternas, S. Agustí, and C. M. Duarte (2014). Comparing marine primary production estimates through different methods and development of conversion equations. *Front. Mar. Sci.* 1: 19. DOI: 10.3389/fmars.2014.00019.

Resplandy, L., J. Vialard, M. Levy, O. Aumont, and Y. Dandonneau (2009). Seasonal and intraseasonal biogeochemical variability in the thermocline ridge of the southern tropical Indian Ocean. *J. Geophys. Res. Oceans* 114: C07024. DOI: 10.1029/2008JC005246.

Riley, G. A. (1946). Factors controlling phytoplankton populations on Georges Bank. *J. Mar. Res.* 6: 5473.

Rodgers, K., J. Lin, and T. Frolicher (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences* 12:11, 3301.

Roesler, C. S. and M. J. Perry (1995). In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance. *J. Geophys. Res.* 100: 13279–13294.

Roesler, C. et al. (2017). Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors. *Limnol. Oceanogr.: Methods* 15: 572–585.

Rose, K. A., D. Justic, K. Fennel, and R. D. Hetland (2017). Numerical modeling of hypoxia and its effects: Synthesis and going forward. In: *Modeling Coastal Hypoxia.* Springer, 401–421.

Rose, K. A., B. M. Roth, and E. P. Smith (2009). Skill assessment of spatial maps for oceanographic modeling. *J. Mar. Syst.* 76:1–2, 34–48.

Rose, K. A. et al. (2010). End-To-End Models for the Analysis of Marine Ecosystems: Challenges, Issues, and Next Steps. *Mar. Coast. Fish.* 2:1, 115–130. DOI: 10.1577/C09-059.1.

Rousseaux, C. S. and W. W. Gregg (2012). Climate variability and phytoplankton composition in the Pacific Ocean. *J. Geophys. Res. Oceans* 117: C10006. DOI: 10.1029/2012JC008083.

— (2014). Interannual variation in phytoplanktonprimaryproduction at a global scale. *Remote Sens.* 6: 1–19.

— (2015). Recent decadal trends in global phytoplankton composition. *Global Biogeochem. Cy.* 29: 1674–1688. DOI: 10.1002/2015GB005139.

— (2017). Forecasting ocean chlorophyll in the Equatorial Pacific. *Front. Mar. Sci.* 4: 236.

Roy, S., S. Sathyendranath, H. Bouman, and T. Platt (2013). The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data. *Remote Sens. Environ.* 139: 185–197. DOI: 10.1016/j.rse.2013.08.004.

Roy, S., S. Sathyendranath, and T. Platt (2011). Retrieval of phytoplankton size from bio-optical measurements: theory and applications. *J. R. Soc. Interface* 8:58, 650–660. DOI: 10.1098/rsif.2010.0503.

— (2017). Size-partitioned phytoplankton carbon and carbon-to-chlorophyll ratio from ocean colour by an absorption-based bio-optical algorithm. *Remote Sens. Environ.* 194: 177–189. DOI: 10.1016/j.rse.2017.02.015.

Ryan, J. P., I. Ueki, Y. Chao, H. C. Zhang, P. S. Polito, and F. P. Chavez (2006). Western Pacific modulation of large phytoplankton blooms in the central and eastern equatorial Pacific. *J. Geophys. Res.* 111:G2. DOI: 10.1029/2005JG000084|.

Ryther, J. and C. Yentsch (1957). The estimation of phytoplankton production in the ocean from chlorophyll and light data. *Limnol. Oceanogr.* 281–286. DOI: 10.1002/lno.1957.2.3.0281.

Ryu, J. H., H. J. Han, S. Cho, Y. J. Park, and Y. H. Ahn (2012). Overview of geostationary ocean color imager (GOCI) and GOCI data processing system (GDPS). *Ocean Sci. J* 47: 223–233. DOI: 10.1007/s12601-012-0024-4.

Saba, V. S. et al. (2010). Challenges of modeling depth integrated marine primary productivity over multiple decades: A case study at BATS and HOT. *Global Biogeochem. Cycles* 24:3. DOI: 10.1029/2009GB003655.

Saba, V. S. et al. (2011). An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe. *Biogeosciences* 8: 489–503.

Sadeghi, A. et al. (2012). Improvements to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data. *Ocean Sci.* 8: 1055–1070.

Sakov, P. (2018). *EnKF-C user guide. arXiv:1410.1233.*

Salisbury, J. et al. (2017). Coastal observations from a new vantage point. *Eos*, 1–7.

Santoleri, R., V. Banzon, S. Marullo, E. Napolitano, F. D'Ortenzio, and R. Evans (2003). Year-to-year variability of the phytoplankton bloom in the southern Adriatic Sea (1998–2000): Sea-viewing Wide Field-of-view Sensor observations and modeling study. *J. Geophys. Res. Oceans* 108:C9. DOI: 10.1029/2002JC001636.

Sarma, V. V. S. S. (2003). Monthly variability in surface pCO$_2$ and net air-sea CO$_2$ flux in the Arabian Sea. *J. Geophys. Res. Oceans* 108:C8.

Sathyendranath, S., R. J. W. Brewin, T. Jackson, F. Melin, and T. Platt (2017). Ocean-colour products for climate-change studies: what are their ideal characteristics? *Remote Sens. Environ.* 203: 125–138.

Sathyendranath, S., F. E. Hoge, T. Platt, and R. N. Swift (1994). Detection of phytoplankton pigments from ocean color: Improved algorithms. *Appl. Opt.* 33: 1081–1089. DOI: 10.1029/2009GB003655.

Sathyendranath, S. and T. Platt (1988). The spectral irradiance field at the surface and in the interior of the ocean: a model for applications in oceanography and remote sensing. *J. Geophys. Res.: Oceans* 93:C8, 9270–9280.

— (2007). Spectral effects in bio-optical control on the ocean system. *Oceanologia* 49: 5–39.

Sathyendranath, S., T. Platt, C. M. Caverhill, R. E. Warnock, and M. R. Lewis (1989). Remote sensing of oceanic primary production: computations using a spectral model. *Deep Sea Res. A.* 36: 431–453.

Sathyendranath, S., L. Watts, E. Devred, T. Platt, C. Caverhill, and H. Maass (2004). Discrimination of diatoms from other phytoplankton using ocean colour data. *Mari. Ecol. Prog. Ser.* 272: 59–68.

Sathyendranath, S. et al. (2009). Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. *Mar. Ecol.Prog. Ser.* 383: 73–84. DOI: 10.3354/meps07998.

Sathyendranath, S. et al. (1991). Estimation of new production in the ocean by compound remote sensing. *Nature* 353: 129–133. DOI: 10.1038/353129a0.

Saux Picart, S., M. Butenschon, and J. Shutler (2012). Wavelet-based spatial comparison technique for analysing and evaluating two-dimensional geophysical model fields. *Geosci. Model Dev.* 5:1, 223–230.

Schiller, H. and R. Doerffer (1999). Neural network for emulation of an inverse model - operational derivation of Case II water properties from MERIS data. *Int. J. Remote Sens.* 20: 1735–1746.

Schlitzer, R. (2004). Export production in the equatorial and North Pacific derived from dissolved oxygen, nutrient and carbon data. *J. Oceanogr.* 60:1, 53–62.

Schneider, B. et al. (2008). Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models. *Biogeosciences* 5:2, 597–614.

Schofield, O. et al. (2004). Watercolors in the coastal zone: what can we see? *Oceanography* 17: 24–31.

Scott, J. D., M. A. Alexander, D. R. Murray, D. Swales, and J. Eischeid (2016). The climate change web portal: A system to access and display climate and Earth system model output from the CMIP5 archive. *Bull. Am. Meteorol. Soc.* 97:4, 523–530.

Séférian, R. et al. (2014). Multiyear predictability of tropical marine productivity. *Proc. Natl. Acad. Sci.* 111:32, 11646–11651.

She, J. et al. (2016). Developing European operational oceanography for Blue Growth, climate change adaptation and mitigation, and ecosystem-based management. *Ocean Sci.* 12: 953–976. DOI: 10.5194/os-12-953-2016.

Sherman, K. and L. Alexander (1986). *Variability and management of large marine ecosystems*. Westview Press, Inc., Boulder, CO.

Shimoda, Y. and G. B. Arhonditsis (2016). Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledgeel. *Ecol. Modell.* 320: 29–43.

Shulman, I. et al. (2013). Impact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions. *J. Geophys. Res.* 118: 2215–2230.

Shutler, J. D. et al. (2016). Fluxengine: a flexible processing system for calculating atmosphere-ocean carbon dioxide gas fluxes and climatologies. *J. Atmos. Oceanic Technol.* 33:4, 741–756.

Shutler, J. et al. (2011). Evaluating the ability of a hydrodynamic ecosystem model to capture inter-and intra-annual spatial characteristics of chlorophyll-a in the north east Atlantic. *J. Mar. Syst.* 88:2, 169–182.

Sieburth, J., V. Smetacek, and J. Lenz (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol. Oceanogr.* 23: 1256–1263.

Siegel, D. A., S. Maritorena, N. B. Nelson, D. A. Hansell, and M. Lorenzi-Kayser (2002). Global distribution and dynamics of colored dissolved and detrital organic materials. *J. Geophys. Res. Oceans* 107:3228. DOI: 10.1029/2001JC000965.

Siegel, D. A., M. Wang, S. Maritorena, and W. Robinson (2000). Atmospheric correction of satellite ocean colour imagery: The black pixel assumption. *Appl. Opt.* 39: 3582–3591.

Siegel, D., S. Maritorena, N. Nelson, M. Behrenfeld, and C. McClain (2005). Coloured dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere. *Geophys. Res. Lett.* 32: L20605. DOI: 10.1029/2005GL024310.

Siegel, D. et al. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sens. Environ.* 135: 77–91. ISSN: 0034-4257. DOI: 10.1016/j.rse.2013.03.025.

Silsbe, G. M., M. J. Behrenfeld, K. H. Halsey, A. J. Milligan, and T. K. Westberry (2016). The CAFE model: A net production model for global ocean phytoplankton. *Global Biogeochem. Cycles* 30:12, 1756–1777. DOI: 10.1002/2016GB005521.

Simon, E. and L. Bertino (2009). Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Sci.* 5:4, 495–510.

Skákala, J. et al. (2018). The assimilation of phytoplankton functional types for operational forecasting in the northwest European Shelf. *J. Geophys. Res. Oceans* 123: 5230–5247.

Smith, R. and K. Baker (1978). Optical classification of natural waters. *Limnol. Oceanogr.* 23:2, 260–267.

Smith, S. L. (2001). Understanding the Arabian Sea: Reflections on the 1994–1996 Arabian Sea expedition. *Deep-Sea Res. II* 48:6–7, 1385–1402.

Smyth, T. J., G. F. Moore, T. Hirata, and J. Aiken (2006). Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment. *Appl. Opt.* 45: 8116–8131.

Sokolov, A. P. et al. (2009). Probabilistic forecast for 21st century climate based on uncertainties in emissions (without policy) and climate parameters. *J. Climate* 22: 5175–5204. DOI: 10.1175/2009JCLI2863.1.

Song, H., C. A. Edwards, A. M. Moore, and J. Fiechter (2016a). Data assimilation in a coupled physical–biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 1 — Model formulation and biological data assimilation twin experiments. *Ocean Modell.* 106: 131–145. DOI: 10.1016/j.ocemod.2016.04.001.

— (2016b). Data assimilation in a coupled physical-biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 2 — Joint physical and biological data assimilation twin experiments. *Ocean Modelling.* 106: 146–158. DOI: 10.1016/j.ocemod.2016.09.003.

— (2016c). Data assimilation in a coupled physical–biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 3 — Assimilation in a realistic context using satellite and *in situ* observations. *Ocean Modell.* 106: 159–172. DOI: 10.1016/j.ocemod.2016.06.005.

Stanski, H. R., W. R. Burrows, and L. J. Wilson (1989). Survey of common verification methods in meteorology. *WMO World Weather Watch Tech. Rep.*

Stedmon, C. A. and S. Markager (2003). Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing. *Estuar. Coast. Shelf Sci.* 57: 973–979.

Steemann Nielsen, E. (1952). The use of radio-active carbon (C^{14}) for measuring organic production in the sea. *ICES J. Mar. Sci.* 18:2, 117–140. DOI: 10.1093/icesjms/18.2.117.

Steinmetz, F., P. Deschamps, and D. Ramon (2011). Atmospheric correction in presence of sun glint: Application to MERIS. *Opt. Express* 19: 571–587.

Stephens, M. P., G. Samuels, D. B. Olson, R. A. Fine, and T. Takahashi (1995). Sea-air flux of CO_2 in the North Pacific using shipboard and satellite data. *J. Geophys. Res. Oceans* 100:C7, 13571–13583.

Stock, C. (2019). Comparing apples to oranges: perspective on satellite-based primary production estimates from a global biogeochemical model. In: *The Sea.* Ed. by P. Glibert and D. M. Jr. John Wiley and Sons.

Stock, C. A. et al. (2015). Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Prog. Oceanogr.* 137: 219–236.

Stock, C. A. et al. (2017). Reconciling fisheries catch and ocean productivity. *Proc. Natl. Acad. Sci. U.S.A.* 114:8, E1441–E1449.

Stock, C., J. Dunne, and J. John (2014a). Drivers of trophic amplification of ocean productivity trends in a changing climate. *Biogeosciences* 11:24, 7125–7135.

— (2014b). Global-scale carbon and energy flows through the marine planktonic food web: An analysis with a coupled physical-biological model. *Prog. Oceanogr.* 120: 11128.

Stock, C. A. et al. (2011). On the use of IPCC-class models to assess the impact of climate on Living Marine Resources. *Prog. Oceanogr.* 88:1-4, 1-27. DOI: 10.1016/j.pocean.2010.09.001.

Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda (1998). Global seasonal rainfall forecasts using a coupled ocean atmosphere model. *Nature* 392: 370-373. DOI: 10.1038/32861.

Stow, C. A. et al. (2009). Skill assessment for coupled biological/physical models of marine systems. *J. Mar. Syst.* 76:1-2, 4-15.

Stramski, D. (1999). Estimation of particulate organic carbon in the ocean from satellite remote sensing. *Science* 285: 239-242.

Stramski, D., E. Boss, D. Bogucki, and K. J. Voss (2004). The role of seawater constituents in light backscattering in the ocean. *Prog. Oceanogr.* 61: 27-56.

Stramski, D., A. Bricaud, and A. Morel (2001). Modeling the inherent optical properties of the ocean based on the detailed composition ofthe planktonic community. *Appl. Opt.* 40: 2929-2945.

Stramski, D., R. A. Reynolds, M. Babin, S. Kaczmarek, M. R. Lewis, and R. Rottgers (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences* 5:1, 171-201. DOI: 10.5194/bg-5-171-2008.

Stramski, D., R. A. Reynolds, M. Kahru, and B. G. Mitchell (1999). Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing. *Science* 285:5425, 239-242. DOI: 10.1126/science.285.5425.239.

Stumpf, R. P. et al. (2016). Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae* 54: 160-173.

Stumpf, R. and M. Tomlinson (2005). Use of remote sensing in monitoring and forecasting of harmful algal blooms. *Proc. SPIE Int. Soc. Opt. Eng.* 5885: DOI: 10.1117/12.614376.

Subramaniam, A., C. Brown, R. Hodd, E. Carpenter, and D. Capone (2002). *Trichedesmium* blooms in SeaWiFS imagery. *Deep-Sea Res. II* 49: 107-121.

Suggett, D. J., C. M. Moore, A. E. Hickman, and R. J. Geider (2009). Interpretation of fast repetition rate (FRR) fluorescence: signatures of phytoplankton community structure versus physiological state. *Mar. Ecol. Prog. Ser.* 376: 1-19.

Sverdrup, H. U. (1953). On conditions for the vernal blooming of phytoplankton. *ICES J. Mar. Sci.* 18:3, 287-295.

Szeto, M., P. J. Werdell, T. S. Moore, and J. W. Campbell (2011). Are the world's oceans optically different? *J. Geophys. Res.* 116: C00H04. DOI: 10.1029/2011JC007230.

Taboada, F. G., A. D. Barton, C. A. Stock, J. Dunne, and J. G. John (2019). Seasonal to interannual predictability of oceanic net primary production inferred from satellite observations. *Prog. Oceanogr.* 170: 28-39.

Tagliabue, A. et al. (2016). How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Glob. Biogeochem. Cycles* 30:2, 149-174.

Tagliabue, A. et al. (2018). The role of external inputs and internal cycling in shaping the global ocean cobalt distribution: insights from the first cobalt biogeochemical model. *Global Biogeochem. Cycles* 32:4, 594-616. DOI: 10.1002/2017GB005830.

Tarran, G. A., J. L. Heywood, and M. V. Zubkov (2006). Latitudinal changes in the standing stocks of nano- and picoeukaryotic phytoplankton in the Atlantic Ocean. *Deep-Sea Res. II* 53: 1516-1529.

Taucher, J. and A. Oschlies (2011). Can we predict the direction of marine primary production change under global warming? *Geophys. Res. Lett.* 38: LO2603. DOI: 10.1029/2010GL045934.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmosphere* 106:D7, 7183-7192.

Teruzzi, A., S. Dobricic, C. Solidoro, and G. Cossarini (2014). A 3-D variational assimilation scheme in coupled transport-biogeochemical models: forecast of Mediterranean biogeochemical properties. *J. Geophys. Res. Oceans* 119: 200-217.

Teruzzi, A., G. Bolzon, S. Salon, P. Lazzari, C. Solidoro, and G. Cossarini (2018). Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea. *Ocean Modell.* 132: 46-60.

Thuillier, G. et al. (2003). The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS and EURECA missions. *Sol. Phys.* 214: 1-22.

Tomlinson, M. C., T. T. Wynne, and R. P. Stumpf (2009). An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, *Karenia brevis*. *Remote Sens. Environ.* 113: 598–609. DOI: 10.1016/j.rse.2008.11.003.

Tommasi, D., C. A. Stock, M. A. Alexander, X. Yang, A. Rosati, and G. A. Vecchi (2017a). Multi-annual climate predictions for fisheries: an assessment of skill of sea surface temperature forecasts for large marine ecosystems. *Front. Mar. Sci.* 4: 201.

Tommasi, D. et al. (2017b). Managing living marine resources in a dynamic environment: the role of seasonal to decadal climate forecasts. *Prog. Oceanogr.* 152: 15–49.

Trees, C. C., M. C. Kennicutt, and J. M. Brooks (1985). Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments. *Mar. Chem.* 17: 1–12.

Trenberth, K. E. and J. G. Olson (1988). An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bull. Am. Meteorol. Soc.* 69: 1047–1057.

Turi, G. et al. (2018). Response of O_2 and pH to ENSO in the California Current System in a high-resolution global climate model. *Ocean Sci.* 14: 69–86.

Twardowski, M. S., E. Boss, J. M. Sullivan, and P. L. Donaghay (2004). Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Mar. Chem.* 89: 69–88.

Tzortziou, M., J. Herman, A. Cede, C. Loughner, N. Abuhassan, and S. Naik (2013). Spatial and temporal variability of ozone and nitrogen dioxide over a major urban estuarine ecosystem. *J. Atmos. Chem.* DOI: 10.1007/s10874-013-9255-8.

Uitz, J., H. Claustre, A. Morel, and S. Hooker (2006). Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *J. Geophys. Res.* 111: C08005. DOI: 10.1029/2005JC003207.

Vaillancourt, R. D., C. W. Brown, R. L. Guillard, and W. M. Balch (2004). Light backscattering properties of marine phytoplankton: relationships to cell size, chemical composition and taxonomy. *J. Plankton Res.* 26: 191–212.

Vallina, S. M., B. A. Ward, S. Dutkiewicz, and M. J. Follows (2014). Maximal foraging with active prey-switching: a new "kill the winner" functional response and its effect on global species richness and biogeography. *Prog. Oceanogr.* 120: 93–109.

Van Oostende, N. et al. (2018). Simulating the ocean's chlorophyll dynamic range from coastal upwelling to oligotrophy. *Prog. Oceanogr.* 232–247. DOI: 10.1016/j.pocean.2018.10.009.

Vanhellemont, Q. and K. Ruddick (2018). Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. *Remote Sens. Environ.* 216: 586–597. DOI: 10.1016/j.rse.2018.07.015.

Vantrepotte, V. and F. Mélin (2011). Inter-annual variations in the SeaWiFS global chlorophyll a concentration (1997–2007). *Deep-Sea Res. I* 58:4, 429–441.

Vantrepotte, V. et al. (2015). CDOM-DOC relationship in contrasted coastal waters: implication for DOC retrieval from ocean color remote sensing observation. *Opt. Express* 23:1, 33. DOI: 10.1364/OE.23.000033.

Vichi, M., N. Pinardi, and S. Masina (2007). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *J. Mar. Syst.* 64:1, 89–109.

Vogt, M. et al. (2013). The distribution, dominance patterns and ecological niches of plankton functional groups in Dynamic Green Ocean Models and satellite estimates. *Biogeosci. Discuss.* 10: 17193–17247. DOI: 10.5194/bgd-10-17193-2013.

Von Storch, H. and F. W. Zwiers (2002). *Statistical Analysis in Climate Research*. Citeseer.

Voss, K. J., A. Morel, and D. Antoine (2007). Detailed validation of the bidirectional effect in various Case 1 waters for application to ocean color imagery. *Biogeosciences* 4: 781–789. DOI: 10.5194/bg-4-781-2007.

Wakelin, S. L., J. T. Holt, J. C. Blackford, J. I. Allen, M. Butenschon, and Y. Artioli (2012). Modeling the carbon fluxes of the northwest European continental shelf: Validation and budgets. *J. Geophys. Res. Oceans* 117:C5.

Wang, M. (2005). A refinement for the Rayleigh radiance computation with variation of the atmospheric pressure. *Int. J. Remote Sens.* 24: 5651–5663.

Wang, M. and S. W. Bailey (2001). Correction of sun glint contamination on the SeaWiFS ocean and atmospheric products. *Appl. Opt.* 40: 4790–4798.

Wang, M. et al. (2013). Ocean colour products for the Korean Geostationary Ocean Color Imager (GOCI). *Opt. Express* 21: 3835–3849.

Ward, B. A., S. Dutkiewicz, O. Jahn, and M. J. Follows (2012). A size structured food-web model for the global ocean. *Limnol. Oceanogr.* 57: 1877–1891.

Ward, B. A. and M. J. Follows (2016). Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc. Natl. Acad. Sci. U S A.* 113:11, 2958–63. DOI: 10.1073/pnas.1517118113.

Ward, B., M. Friedrichs, T. Anderson, and A. .Oschlies (2010). Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *J. Mar. Syst.* 81: 34–43.

Watson, J. R., C. A.Stock, and J. L.Sarmiento (2015). Exploring the role of movement in determining the global distribution of marine biomass using a coupled hydrodynamic-size-based ecosystem model. *Prog. Oceanogr.* 138: 521–532.

Weatherhead, E. C. et al. (1998). Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.* 103:D14, 17149–17161. DOI: 10.1029/98JD00995.

Weitz, J. W. et al. (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J.* 9:6, 1352.

Welschmeyer, N. (1994). Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnol. Oceanogr.* 39: 1985–1992.

Werdell, J. et al. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Appl. Optics* 52: 2019–2037. DOI: 10.1364/ao.52.002019.

Werdell, P. J., L. I. W. McKinna, E. Boss, S. G. Ackleson, S. E. Craig, and W. W. Gregg (2018). An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. *Prog. Oceanogr.* 160: 186–212. DOI: 10.1016/j.pocean.2018.01.001.

Westberry, T., M. J. Behrenfeld, D. A. Siegel, and E. Boss (2008). Carbon based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochem. Cycles* 22: GB2024. DOI: 10.1029/2007GB003078.

Wiggert, J. D. and R. G. Murtugudde (2007). The sensitivity of the Southwest Monsoon phytoplankton bloom to variations in aeolian iron deposition over the Arabian Sea. *J. Geophys. Res.* 112: DOI: 10.1029/2006JC003514.

Wiggert, J. D., R. G. Murtugudde, and J. R. Christian (2006). Annual ecosystem variability in the tropical Indian Ocean: Results of a coupled bio-physical ocean general circulation model. *Deep-Sea Res. II* 53:5–7, 644–676.

Wiggert, J. D., J. Vialard, and M. Behrenfeld (2009). Basinwide modification of dynamical and biogeochemical processes by the positive phase of the Indian Ocean Dipole during the SeaWiFS era. In: *Indian Ocean Biogeochemical Processes and Ecological Variability*. Ed. by J. D. Wiggert, R. R. Hood, S. A. Naqvi, K. H. Brink, and S. L. Smith. Vol. 185. Geophysical Monograph Series. Washington, D. C.: American Geophysical Union. DOI: 10.1029/2008GM000776.

Wild-Allen, K., M. Herzfeld, P. A. Thompson, U. Rosebrock, J. Parslow, and J. K. Volkman (2010). Applied coastal biogeochemical modelling to quantify the environmental impact of fish farm nutrients and inform managers. *J. Mar. Sys.* 81: 134–147.

Wilks, D. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

Williams, P. J. L. and D. A. Purdie (1991). *In vitro* and *in situ* derived rates of gross primary production, net community production and respiration of oxygen in the oligotrophic subtropical gyre of the North Pacific ocean. *Deep Sea Res. I* 38: 891–190.

Wolanin, A., M. A. Soppa, and A. Bracher (2016). Investigation of spectral band requirements for improving retrievals of phytoplankton functional types. *Remote Sens.* 8: 871. DOI: 10.3390/rs8100871.

Xiao, Y. and M. A. M. Friedrichs (2014a). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *J. Geophys. Res. Oceans* 119: 2691–2712. DOI: 10.1002/2013JC009433.

— (2014b). Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models: effects of increasing the complexity of the planktonic food web. *Biogeosciences* 11:11, 3015–3030.

Xing, X., A. Morel, H. Claustre, F. D'Ortenzio, and A. Poteau (2012). Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Coloured dissolved organic matter absorption retrieval. *J. Geophys. Res.: Oceans* 117: C04022. DOI: 10.1029/2011JC007632.

Xiu, P. and F. Chai (2014). Connections between physical, optical and biogeochemical processes in the Pacific Ocean. *Prog. Oceanogr.* 122: 30–53.

Yool, A., E. E. Popova, and T. R. Anderson (2013). MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies. *Geosci. Model Dev.* 6: 1767–1811. DOI: 10.5194/gmd-6-1767-2013.

Zakem, E. et al. (2018). Ecological control of nitrite in the upper ocean. *Nat. Commun.* 9: 1206. DOI: 10.1038/s41467-018-03553-w.

Zhao, J. et al. (2013). Assessment of satellite-derived diffuse attenuation coefficients and euphotic depths in south Florida coastal waters, *Remote Sens. Environ.* 131: 38–50.

Zheng, G. and P. M. DiGiacomo (2017). Uncertainties and applications of satellite-derived coastal water quality products. *Prog. Oceanogr.* 159: 45–72. DOI: 10.1016/j.pocean.2017.08.007.

Zielinski, O., O. Llinás, A. Oschlies, and R. Reuter (2002). Underwater light field and its effect on a one-dimensional ecosystem model at station ESTOC, north of the Canary Islands. *Deep Sea Res. II* 49:17, 3529–3542.

Zwolinski, J. P., R. L. Emmett, and D. A. Demer (2011). Predicting habitat to optimize sampling of Pacific sardine (*Sardinops sagax*). *ICES J. Mar. Sci.* 68:5, 867–879.