Principal Component Inversion

Dr. A. Neumann, H. Krawczyk
German Aerospace Centre DLR
Remote Sensing Technology Institute
Marine Remote Sensing

- Principal Components - Properties
- The Linear Inversion Algorithm
- Optimisation of the PCI: Segmentation and Semi-logarithmic Approach
- Implementation Scheme
- PCI and Atmospheric Correction
- Examples

Introduction

- The classical way to develop ocean colour algorithms is to correlate experimental data sets, i.e. in-situ analyses of concentrations with remote sensing measurements

  ⇨ empirical algorithms, usually band ratios
  ⇨ work good under case-1 conditions

- this approach fails if we want to develop algorithms for spectral high resolution data (large number of channels) and/or several independently but simultaneously varying geo-physical parameters

  ⇨ multidimensional, multivariate approach is necessary looking for an algorithm sensitive to the shape of the spectrum

andreas.neumann@dlr.de

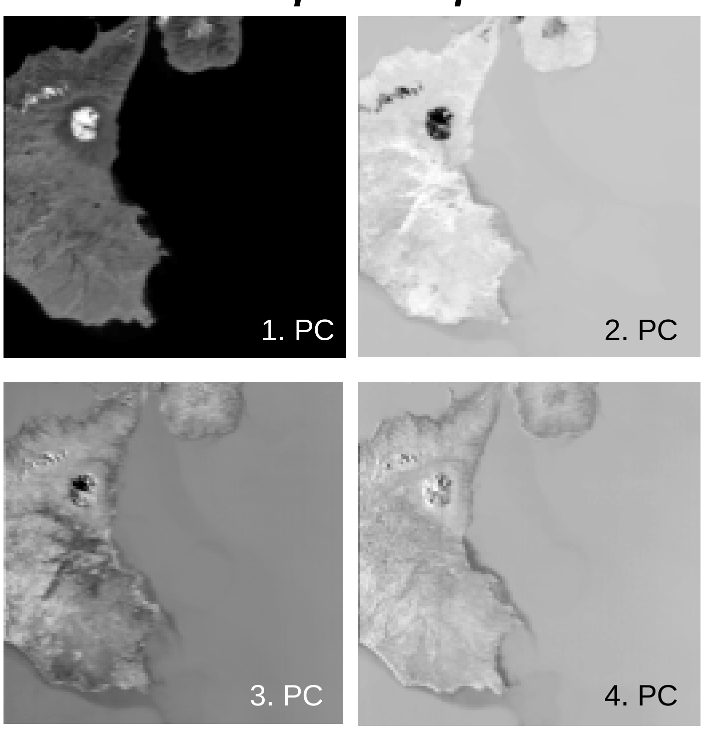
Principal Components

- PCs are an effective tool handle high dimensional and correlated data, to analyse the information content and dimensionality of multivariate data sets
- however, a physical interpretation of PCs is problematic, there is no direct link to (geo-) physical parameters
- some physical understanding can be introduced through the corresponding eigenvectors (brightness, greenness, ...)
- eigenvectors and therewith the PCs depend on the covariance matrix, therefore the eigenvector system will be different for every scene or data set to be analysed

⇨ the idea is to use the PCA as a mathematical tool to develop an algorithm for the physical interpretation of multivariate, high dimensional data which can be optimised for different water types

Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
andreas.neumann@dlr.de

Principal Components - Example (I)

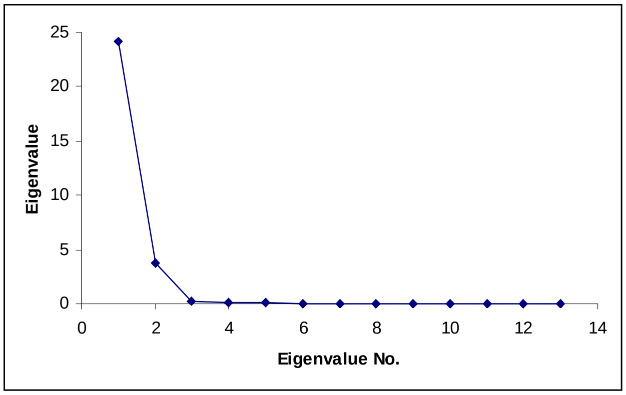


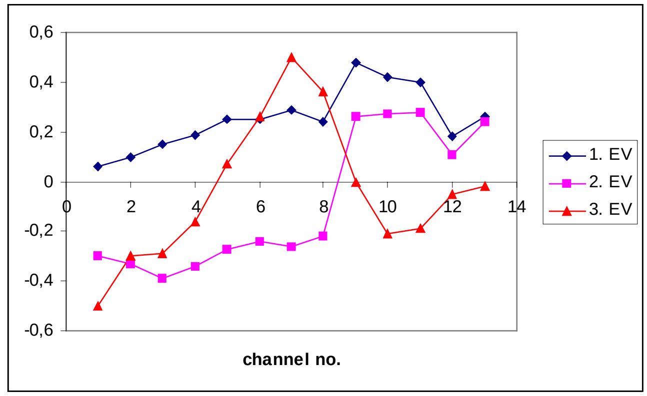
First 4 PCs of MOS-IRS data over Sicilia

andreas.neumann@dlr.de

Principal Components - Example (II)

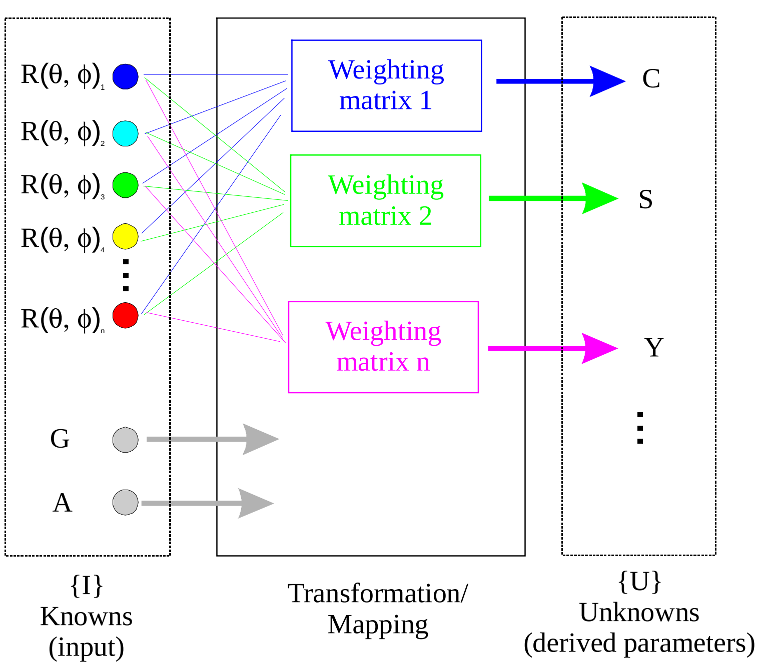
Eigenvalues and Eigenvectors of the
MOS-IRS Sicilia Scene





andreas.neumann@dlr.de

The Remote Sensing Problem



- ! The inverse task is analytically not solvable
- ? Which and how many parameters can be retrieved
- ? Which and how many channels contribute how much to each parameter
- ? How to construct an „optimal“ algorithm

The Linear Inversion Algorithm

- Basic ideas developed and evaluated for MOS-IRS, operational implementation for MERIS
- goal: quantitative retrieval of the main components for case-2 remote sensing using a linear estimator:

$$\hat{p}_i = \sum_j k_{ij} L_j + A_i \qquad (1)$$

- where
  - $\hat{p}_i$ - estimate of the geophysical parameter, e.g. chlorophyll, Gelbstoff and sediments
  - k_{ij} - weightening coefficient in band j for parameter i
  - L_j - measured radiance/reflectance in band j
  - A_i - offset value for parameter i
  - j - spectral band number, from 1 to N

Step 1 - Forward Modelling

- The principal Component Inversion algorithm was developed using simulated data sets containing radiance vectors and the corresponding geo-physical parameters:

$$\{p_i\} \xmapsto[\text{forward model}]{} \{L_j\},\ j \text{ from 1 to N}$$

- the p_i are Chlorophyll concentration C, Gelbstoff attenuation a_y(440nm), sediment scattering b_s(550nm) and aerosol-optical thickness τ_A (750), varying independently and simultaneously
- bio-optical model after Morel/Prieur/Sathyendranath $R_V = 0.33\frac{b_B}{a}$
  in the beginning simplest atmospheric modelling after Gordon and Sturm using Angstroem-law to describe the Aerosol

Important remark: *the used forward model is not a principle issue of the algorithm, any other forward model can be used!*

Step 2 - Principal Component Analysis (PCA)

- The forward model generates a large set (10^4 .. 10^6) of vectors

  $\{C, a_y, b_s, \tau_A, L_1 .. L_N\}$

- variability ranges and, possibly correlation between single parameters, are chosen corresponding to the area/season of interest, specific optical properties of water constituents can be accounted for by the IOPs in the bio-optical model

- ***Note:*** *depending of the goal other combinations of parameters are possible, atmosphere may be absent when developing algorithms for BOA-measurements or atmospherically corrected data, also reflectance values may be used*

- a PCA as described earlier is applied to the radiance values yielding $\{PC_k\}$ and the corresponding eigenvalues λ_k, k=1..N

IOCCG International Ocean Colour Coordinating Group
Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
andreas.neumann@dlr.de

Step 3 - Instrinsic Dimensionality

- The intrinsic dimensionality of the simulated radiance data set now can be determined as:

  $D = \max(k)$ with $\sqrt{\lambda_k} >> 1, D < N.$

- thus, two groups of PCs are separated
  - ones representing useful measurement information (k ☼ D) and
  - ones containing non-interpretable variations due to noise, quantisation error etc.

- For the following analyses only the first group of PCs is used. This enhances the stability of the inversion, suppresses noise and reduces the data set to the usable information in the mathematical sense

Step 4 - Reverse Correlation (I)

- The PCs represent the *identical information as the radiance data*, but in a mathematically defined, orthogonal coordinate system, except the small noise-like portion of information that is suppressed by reducing the number of used components to the "significant" ones
- therefore it must be possible to retrieve the geo-physical parameters from the PCs as well, as long as it is possible at all:

$$\hat{p}_i \sim \sum_{m=1}^{D} C_{im} PC_m \qquad (2)$$

where C_{im} is the correlation coefficient between the ith parameter and the mth principal component

- this is already close to an inverse relationship of equ. (1) !

Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
andreas.neumann@dlr.de

Step 4 - Reverse Correlation (II)

- Composing now a data set containing the PCs and the corresponding physical values we can use a regression to determine the coefficients C_{im}

$$\frac{\hat{p}_i - \bar{p}_i}{\sigma_i} = \sum_m \frac{C_{im} PC_m}{\sqrt{\lambda_m}} \quad \text{for} \rightarrow \text{MSE min} \qquad (3)$$

where	$\hat{p}_i$	- estimate of the parameter
	$\bar{p}_i$	- mean value of the parameter
	σ_i	- variance of the parameter
	C_{im}	- correlation coefficient
	MSE	- mean square error

- so we can estimate the parameters from principle components - but this is not what we really want ...

Step 5 - Reverse Transformation to Radiances

- Since the PCs are computed from the radiance vectors one can substitute them by the inverse transformation and get the regression formula in terms of radiances:

$$\frac{\hat{p}_i - \overline{p}_i}{\sigma_i} = C_{i1} \sum_{j=1}^{D} \frac{U_{1j}(L_j - \overline{L}_l)}{\Delta L_j \sqrt{\lambda_j}} + C_{i2} \sum_{j=1}^{D} \frac{U_{2j}(L_j - \overline{L}_j)}{\Delta L_j \sqrt{\lambda_j}} + \ldots \quad (4)$$

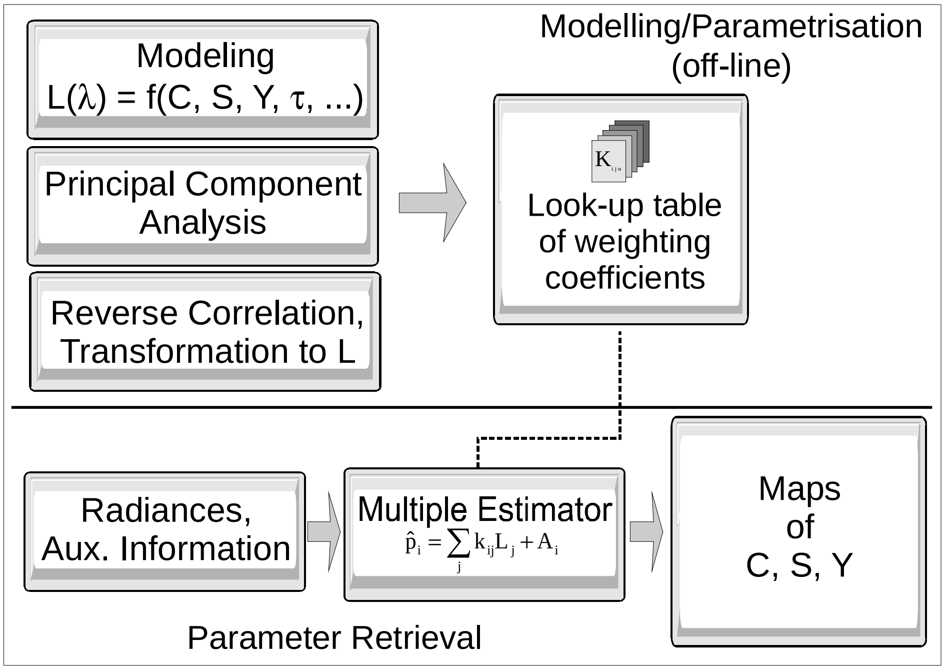
Step 6 - Determination of Coefficients

- from (4) we can compute the desired coefficients k_{ij} and A_i for the estimator in equ. (1) by regression
- the computed coefficient set is stored in a look-up-table (LUT) and represents the inversion algorithm for the situation corresponding to the modelling: geometry, IOPs, atmosphere, variability ranges and correlation properties etc.
- by repeating the analysis for different simulations, the algorithm can be adapted to geometry, specific regions, seasons, species compositions etc.

PCI Basics - Summary

- PCI is a fast and easy to implement algorithm, all time-consuming computations can be performed off-line
- PCs as an orthogonal representation of the data allows the multivariate inversion, what would be hard to do with radiances due to high spectral correlation (ill-posed problem of inverting Cov{$L_j/\Delta L_j$})
- the approach allows to account for measurement accuracy
- PCI is adaptable to wide variety of specifics and parameters and allows a detailed analysis of influencing factors

Implementation



IOCCG International Ocean Colour Coordinating Group
Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
andreas.neumann@dlr.de

PCI - Practical Considerations

- PCI is an optimal strategy for *linear relationships* - but the relation between radiances and geo-physical parameters is *non-linear*, especially for wide ranges of variability
- PCI technique gives an optimal estimate in the sense of MSE, i.e. works most optimal for the mean value of the parameters, at the borders the retrieval errors will be larger

- therefore additional optimization was developed:
  - ⇨ dividing the complete variability range into subranges for which separate coefficient tables are computed
  - ⇨ realize a semi-logarithmic expression of the parameters

Segmentation Procedure

- Investigations on simulated data showed, that for case-2 waters the sediment scattering is the dominating factor in BOA radiance data
- therefore a hierarchical segmentation is applied:
  - the entire range of b_s is divided in 5 subranges
  - within each of these subranges C an a_y vary over their total range respectively
  - further optimisation can be achieved by additional segmentation of C
- during retrieval the results for all subranges and all pixels are computed. Valid in a subrange image are only pixel values within the defined subrange, the others are masked out.
- Superimposing the resulting subimages gives the final parameter map

IOCCG International Ocean Colour Coordinating Group
Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
DLR
andreas.neumann@dlr.de

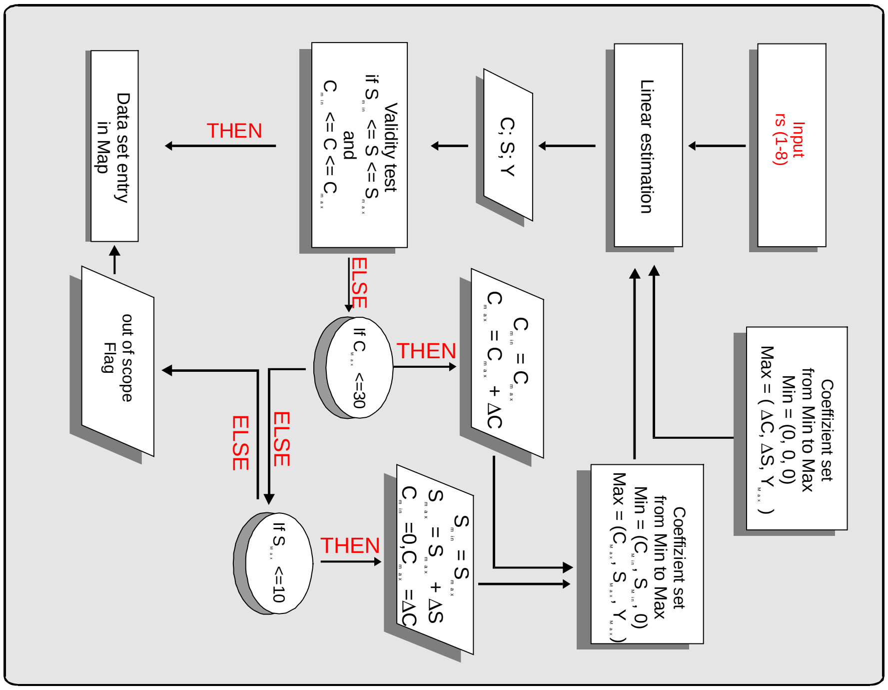
Semi-logarithmic Parameters

- Because the original relationship between parameters and radiances is non-linear, it was found by numerical tests that the algorithm performance increases if the PCI is applied to auxiliary parameters defined by

$$q_i = p_i + a \ln(p_i) \quad \text{with } a = 0.1$$

- this accounts for non-linearity for small parameter values and increases the retrieval accuracy significantly

Implementation Scheme



IOCCG International Ocean Colour Coordinating Group
Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
DLR
andreas.neumann@dlr.de

PCI and Atmospheric Correction

- the PCI algorithm was developed especially for case-2 waters (i.e. turbid, high scattering, sometimes shallow)
- under these conditions the usual atmospheric correction schemes fail because the “black water” condition is invalid
- one possible solution is to treat water and atmosphere in an integrated inversion procedure, accounting for these effects

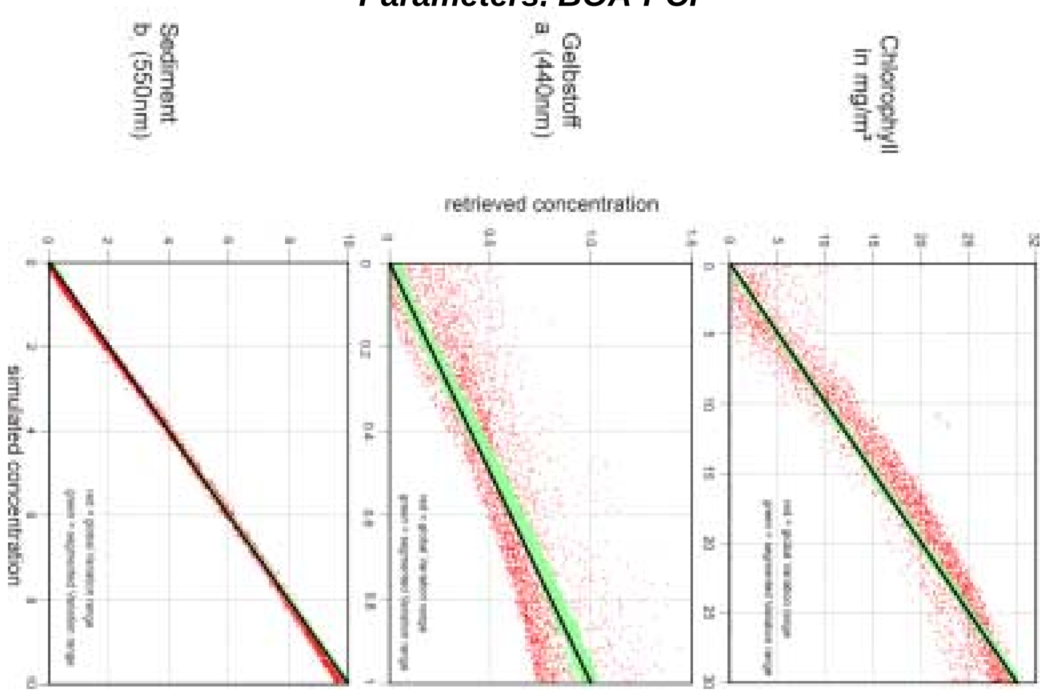
- for PCI we therefore simulated TOA radiances, including the atmosphere (aerosol) as an additional parameter in the forward model (cp. Step 2)

⇨ ***the result is a retrieval algorithm for water constituents, that is directly applied to TOA radiances and does not need an extra atmospheric correction, the inversion “automatically” accounts for the atmospheric influence***

Summary

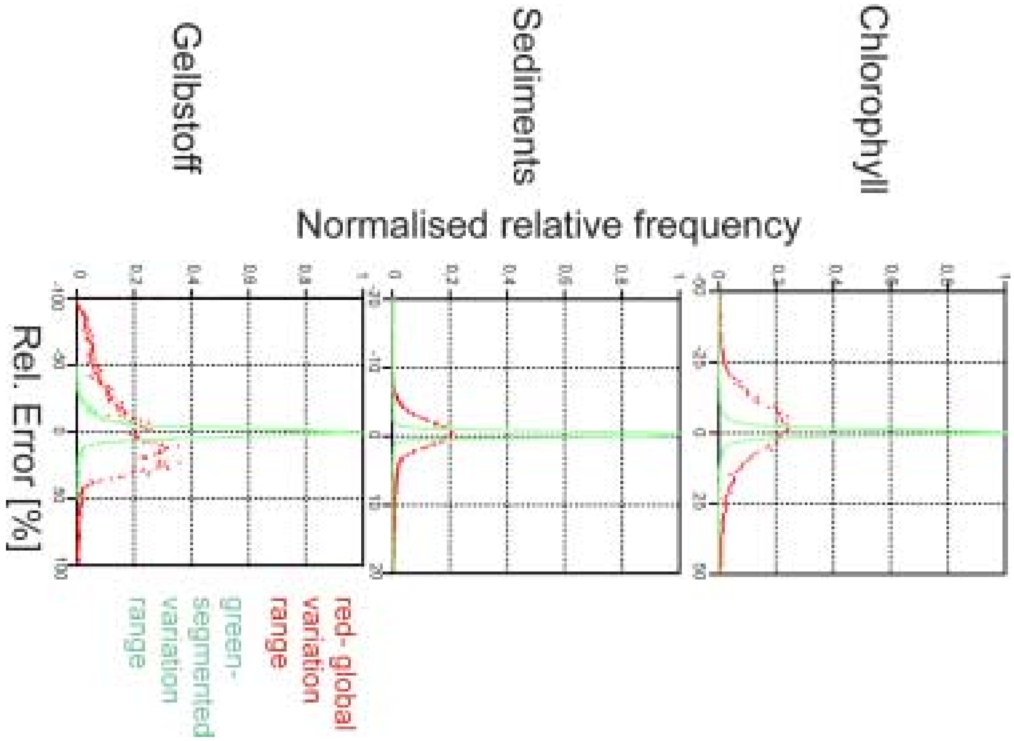
- With PCI inversion we have a very stable and fast algorithm for quantitative retrieval of water constituents from spectral high resolution remote sensing data
- it can easily be adapted to different sensors, water types or geo-physical situations
- cp. to NN the “training” is much simpler and faster
- the inversion is very stable with respect to noise
- TOA PCI provides a fast retrieval procedure ***without atmospheric correction*** for both case-1 and case-2 waters. The accuracy for the in-situ data available for MOS is ~30%
- the optimised (i.e. segmented semi-log) BOA PCI promises extreme good accuracy, but still needs to be evaluated for experimental data

Examples - Scatterplot of Simulated and Retrieved Parameters. BOA-PCI



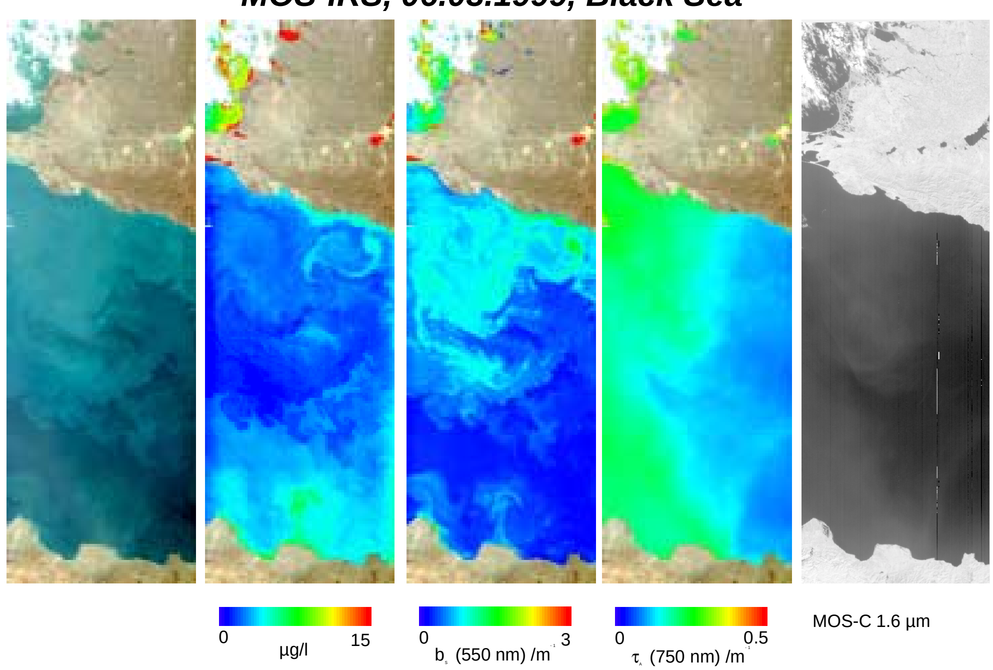
andreas.neumann@dlr.de

Examples - Relative Retrieval Error Histograms, BOA-PCI



IOCCG International Ocean Colour Coordinating Group
Training Course on Remote Sensing of Ocean Colour
Ahmedabad, India, February 2001
Principal Component Inversion (PCI)
andreas.neumann@dlr.de

Examples - TOA-retrieved Parameters
MOS-IRS, 06.03.1999, Black Sea



Colour Composite | Chlorophyll | Sediments | aer.opt. Thickness

andreas.neumann@dlr.de

Examples - TOA-retrieved Parameters
MOS-IRS, 06.03.1999, German Bight