

**Predicting different pigments from remote sensing reflectances**

A. Bracher et al.

# Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of concentrations of phytoplankton pigments

A. Bracher<sup>1,2</sup>, M. H. Taylor<sup>3</sup>, B. Taylor<sup>1</sup>, T. Dinter<sup>1,2</sup>, R. Röttgers<sup>4</sup>, and F. Steinmetz<sup>5</sup>

<sup>1</sup>Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bussestraße 24, 27570 Bremerhaven, Germany

<sup>2</sup>Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, 28359 Bremen, Germany

<sup>3</sup>ZMT – Leibniz Center for Tropical Marine Ecology, Fahrenheitstraße 6, 28359 Bremen, Germany

<sup>4</sup>HZG – Helmholtz Zentrum Geesthacht, Center of Materials and Coastal Research, Max-Planck-Str., 21502 Geesthacht, Germany

<sup>5</sup>HYGEOS, Euratechnologies, 165 Avenue de Bretagne, 59000 Lille, France

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Received: 30 July 2014 – Accepted: 26 August 2014 – Published: 11 September 2014

Correspondence to: A. Bracher (astrid.bracher@awi.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

**OSD**

11, 2073–2117, 2014

---

**Predicting different pigments from remote sensing reflectances**

A. Bracher et al.

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Abstract

The composition and abundance of algal pigments provide information on characteristics of a phytoplankton community in respect to its photoacclimation, overall biomass, and taxonomic composition. Particularly, these pigments play a major role in photoprotection and in the light-driven part of photosynthesis. Most phytoplankton pigments can be measured by High Performance Liquid Chromatography (HPLC) techniques to filtered water samples. This method, like others when water samples have to be analysed in the laboratory, is time consuming and therefore only a limited number of data points can be obtained. In order to receive information on phytoplankton pigment composition with a higher temporal and spatial resolution, we have developed a method to assess pigment concentrations from continuous optical measurements. The method applies an Empirical Orthogonal Function (EOF) analysis to remote sensing reflectance data derived from ship-based hyper-spectral underwater radiometric and from multispectral satellite data (using the MERIS Polymer product developed by Steinmetz et al., 2011) measured in the Eastern Tropical Atlantic. Subsequently we developed statistically linear models with measured (collocated) pigment concentrations as the response variable and EOF loadings as predictor variables. The model results, show that surface concentrations of a suite of pigments and pigment groups can be well predicted from the ship-based reflectance measurements, even when only a multi-spectral resolution is chosen (i.e. eight bands similar to those used by MERIS). Based on the MERIS reflectance data, concentrations of total and monovinyl chlorophyll *a* and the groups of photoprotective and photosynthetic carotenoids can be predicted with high quality. The fitted statistical model constructed on the satellite reflectance data as input was applied to one month of MERIS Polymer data to predict the concentration of those pigment groups for the whole Eastern Tropical Atlantic area. Bootstrapping explorations of cross-validation error indicate that the method can produce reliable predictions with relatively small data sets (e.g., < 50 collocated values of reflectance and pigment concentration). The method allows for the derivation of time series from continuous reflectance

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



data of various pigment groups at various regions, which can be used to study variability and change of phytoplankton composition and photo-physiology.

## 1 Introduction

Optical measurements taken from various platforms have been successfully used to determine the total chlorophyll *a* (TChl *a*) concentration, e.g. see summary by McClain (2009). Those measurements can be taken continuously and, thereby, estimate TChl *a* concentration at a much higher temporal and spatial resolution than it is possible from chemical measurements in the laboratory, e.g., by High Performance Liquid Chromatography (HPLC) analysis of discrete water samples. Chl *a* is the major pigment in all phytoplankton species and is often used as an indicator of phytoplankton biomass. When pigments are measured by HPLC, TChl *a* is defined as the sum of monovinyl Chl *a* (MVChl *a*), divinyl Chl *a* (DVChl *a*) and chlorophyllide *a* (which is mainly formed as artefact of the former two during the extraction process and therefore included in the calculation). DVChl *a* exists only in the prokaryotic genus *Prochlorococcus*. MVChl *a* is the Chl *a* pigment for all other phytoplankton (other cyanobacteria and eukaryotes). Besides Chl *a* there are many other pigments in phytoplankton that are either involved in light harvesting (such as chlorophyll *b* (Chl *b*), chlorophyll *c* (Chl *c*) and several carotenoids, called photosynthetic carotenoids (PSC)), or protecting Chl *a* and other sensitive pigments from photodamage (photoprotective carotenoids, PPC). Some pigments, e.g., zeaxanthin (Zea) in cyanobacteria, only occur in certain phytoplankton groups and are used as marker pigments to identify them (e.g., via the program CHEMTAX developed by Mackey et al., 1996).

When analysing biogeochemical fluxes in the oceans, however, it is inadequate to consider phytoplankton as a single variable (i.e. TChl *a*) because various groups have different roles in the biogeochemical processes (such as carbon fixation and export, nitrogen fixation, and silicon uptake). Their overall biomass and primary production is not well correlated to their TChl *a* concentration due to variations in pigment concentration

### Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



in response to several factors (e.g., light, temperature, and nutrients). The knowledge of the distribution of different phytoplankton pigments gives insight into phytoplankton composition, overall light absorption, and physiological state.

Several researchers lately have investigated the potential to derive pigments other than TChl *a* from continuous optical data which have the potential to deliver a data set with much better spatial and temporal coverage than obtained with analysing water samples. Chase et al. (2013) decomposed a large global data set of hyperspectral particulate absorption measurements into Gaussian function components and assessed the magnitude of specific Gaussian functions to the absorption by specific pigments or pigment groups. The method provided robust results for obtaining concentrations of TChl *a*, TChl *b* (sum of different types of Chl *b*), TChl *c* (sum of different types of Chl *c*), PSC, PPC and PE (phycoerythrine). Organelli et al. (2013) used a multivariate approach applied to fourth derivative spectra of phytoplankton or particulate absorption ( $a_{ph}$  and  $a_p$ , respectively) data to retrieve TChl *a*, the total concentrations of seven diagnostic pigments and three phytoplankton size classes. However,  $a_p$  and  $a_{ph}$  are inherent optical properties (IOP) which cannot be determined from satellite ocean colour measurements directly where only after successful atmospheric correction water leaving reflectance ( $\rho_w$ ), an apparent optical property (AOP), is derived.  $\rho_w$  is not only related to phytoplankton absorption and therefore the imprints of different types of pigments, which in addition correlate in most parts of the spectrum among each other, are masked not only by scattering and absorption of other water constituents and water itself but also by changing radiance distribution in response to varying environmental conditions, e.g., observation geometry, surface waves, atmospheric conditions, etc..

Pan et al. (2010) developed empirical algorithms based on reflectance ratios to approximate key phytoplankton pigment concentrations. The band ratio algorithms were developed from underwater radiometric measurements collocated to pigment data taken at the US northeast coast and were successful in deriving the concentration of TChl *a*, TChl *b*, TChl *c* and nine different carotenoids. However, such band ratio algorithms require a very large data base (> 400 collocations with satellite data) from a certain

## OSD

11, 2073–2117, 2014

### Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



region to derive robust results. Pan et al. (2013) later described that the algorithm had to be adapted by modifying the pigment specific coefficients based on a regional specific data set.

Craig et al. (2012) developed local models to estimate TChl  $a$  and  $a_{ph}$  at different wavelengths from hyperspectral in situ measurements of remote sensing reflectance ( $R_{RS}(\lambda)$ ) in an optically complex water body. The models were based on empirical orthogonal functions (EOF) analysis of normalized  $R_{RS}(\lambda)$  spectra and subsequently linear fitting of measured TChl  $a$  concentration and  $a_{ph}(\lambda)$ , respectively, as response variables to EOF loadings as predictor variables. Taylor et al. (2013) showed that the method could be used similarly to derive PE concentration from underwater upwelling radiance spectra ( $L_u(\lambda)$ ) which enabled continuous profile predictions of PE concentrations. The present study aims to use the spectral information contained in reflectance data to derive the optical signature of different pigments by applying an automatic and generic technique, and with an additional focus on evaluating performance as a function of sample size. The EOF analysis is applied to  $R_{RS}$  and to  $\rho_{wN}$  (i.e. normalized  $\rho_w$  just above surface) data measured in the field and by satellite sensors, respectively, in the Atlantic Ocean in order to predict the concentrations of several phytoplankton pigments and pigment groups. In addition, the application of our statistical method to study the large scale distribution and photo-physiology of phytoplankton based on various pigments concentrations is investigated.

## 2 Material and methods

Two sets of optical and pigment data from the Atlantic Ocean were used in the analysis. A first model set-up used a data set which included only optical measurements taken in situ (as depth profiles) and collocated surface pigment data collected during the transatlantic RV *Polarstern* cruises ANT XXVI/4, ANTXXV/1 and ANTXXVI74. In the following, we call this data set “field data set”. For a second data set, the “satellite-based data set”, we considered water reflectance measurements from the satellite sensor

### Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



MERIS collocated to pigment data from various researchers in the tropical Atlantic Ocean.

## 2.1 Field data set

Samples for the field data set were collected during three RV *Polarstern* cruises: the expeditions ANTXXIV/4 in April/May 2008 and ANTXXVI/4 in April/May 2010 followed a South-to-North transect through the Atlantic Ocean from Punta Arenas (Chile) to Bremerhaven (Germany); ANTXXV/1 in November 2008 followed a North-to-South transect through the eastern Atlantic Ocean from Bremerhaven to Cape Town (South Africa) (see Fig. 1). Sampling was generally conducted at noon local time and involved CTD casts with water samplers as well as below-water radiance and irradiance and above-water irradiance measurements. Water samples from surface water (< 10 m) for pigment analysis and for PE analysis were filtered on GF/F filters and on 0.4 µm polycarbonate filters, respectively. Filters were immediately shock-frozen in liquid nitrogen and stored at -80 °C until further analysis at the home laboratories at Alfred-Wegener-Institute Helmholtz Centre of Polar and Marine Research (AWI).

### 2.1.1 Pigment data

The composition of pigments which are soluble in organic solvents was analysed by HPLC following the method by Barlow et al. (1997) adjusted to our temperature-controlled instruments as detailed in Taylor et al. (2011). We determined the list of pigments shown in Table 1 of Taylor et al. (2011) and applied the method by Aiken et al. (2009) for quality control of the pigment data. HPLC data for ANTXXV/1 were already published in Taylor et al. (2011) and are available from PANGAEA (doi.pangaea.de/10.1594/PANGAEA.819070). The relative concentration of PE was taken from the data set published for all three cruises in PANGAEA (doi.pangaea.de/10.1594/PANGAEA.819624) and analysed in Taylor et al. (2013). As outlined in Taylor et al. (2013), the PE

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



concentration is expressed as a relative value, while all other pigments concentrations are directly measured values.

### 2.1.2 Reflectance data field data set

We used  $RRS(\lambda)$  data obtained for all three cruises as AOP input data.  $RRS$  data of AN-TXXV/1 were already published in Taylor et al. (2011) and are available from PANGAEA (doi.pangaea.de/10.1594/PANGAEA.819506). For the other two cruises we applied the same technique and instrumentation as in Taylor et al. (2011) to derive the  $RRS$  spectrum at each station. To test the influence of spectral resolution of AOPs, the hyperspectral field  $RRS(\lambda)$  data were reduced to the multispectral bands of MERIS (412, 443, 490, 510, 560, 620, 665 and 681 nm) by taking the integral over all wavebands within one band ( $\pm 10$  nm around the center wavelength; except for 681 nm  $\pm 7.5$  nm was used).

To allow direct comparisons to MERIS normalized water-leaving reflectance, from the  $RRS(\lambda)$  the water-leaving reflectance,  $\rho_{w\_in\ situ}$ , was calculated as

$$\rho_{w\_insitu}(\lambda) = \pi \cdot RRS(\lambda). \quad (1)$$

$\rho_{w\_insitu}(\lambda)$  was then normalized to the sun and sensor position at nadir ( $\rho_{wN}(\lambda)$ ) according to (Barker et al., 2008) using the solar zenith angle at observation and corrections for the bidirectional structure of the light field (lookup tables for  $f/Q$  and  $R$  factors) as provided in Morel and Gentili (1993, 1996) and Morel et al. (1995). The later were only available for the first seven wavebands but not for 681 nm.

### 2.2 Satellite-based data set

A large data set of phytoplankton pigment data has been compiled (for more details on the data set see Supplement Table S1). The pigment concentrations had been determined from the sea surface ( $< 10$  m) with HPLC by several investigators within the area of 35° N to 10° S and 42° W to 3° E during the MERIS/ENVISAT mission lifetime



(2002–2012). A large part of those data are publicly available from the SEABASS and BODC databases. The other part consists of pigment data from the field data set within this area, including additional data from stations where no radiometric measurements had been taken, and from four other cruises: pigment data from the RV *Maria S. Merian* cruise MSM-18/3 were analysed by AWI as described above in Sect. 2.1.1; those from two RV *Polarstern* cruises (ANTXXIII/1 and ANTXXIV/1) were analysed by HZG following Zapata et al. (2000); data of the *Bonus Good Hope* (BGH) cruise, conducted by the Laboratoire d’Océanographie de Villefranche, (LOV), have been acquired as outlined in Speich et al. (2008) and analysed following the method by Ras et al. (2008).

As AOP input data we used the MERIS Polymer level-2  $\rho_{\text{wN}}(\lambda)$  product. The Polymer algorithm (for details see Steinmetz et al., 2011) provides a powerful atmospheric correction. It is an iterative spectral matching method over the whole available sensor spectrum and uses two decoupled models: first, the water reflectance is modelled using as two parameters the Chl *a* concentration and the particle backscattering coefficient. Second, the reflectance of the atmosphere including aerosols and contamination by sun-glint is simplified by using an analytical expression, that can account for multiple interactions between molecular and aerosol scatterings (and glitter) without reference to a specific aerosol model. Hence, it allows to retrieve large amounts of MERIS observations in sun-glint, thin clouds or heavy aerosol plumes contaminated conditions which could not be treated correctly by standard atmospheric correction schemes extrapolating from the near infra-red. MERIS Polymer products thus improve the spatial coverage by almost a factor of two and have been proven successful for retrieving MERIS Ocean Colour products: Polymer was selected as the MERIS processor for atmospheric correction in the frame of the Ocean Colour Climate Change Initiative after an extensive validation and inter-comparison with other atmospheric correction algorithms in which each algorithm’s uncertainty was assessed (Müller and Krasemann, 2012). However, still uncertainties probably result from the different size of foot-prints (1 km by 1 km) from the satellite data and about 20 cm by 20 cm sampled area for the water sample.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Matchups between pigment data and MERIS Polymer  $\rho_{\text{WN}}(\lambda)$  and TChl  $a$  products were determined according to the MERMAID data base as  $1 \times 1$  (within the MERIS pixel), as  $3 \times 3$  and  $5 \times 5$  pixels around the field observation (see Barker et al., 2008). For the  $3 \times 3$  and  $5 \times 5$  MERIS pixel match-ups the mean  $\rho_{\text{WN}}(\lambda)$  and TChl  $a$  concentrations were calculated. Then the  $1 \times 1$ , mean  $3 \times 3$  and mean  $5 \times 5$  MERIS  $\rho_{\text{WN}}(\lambda)$  matchup data were used for deriving predicted (modelled) pigment concentrations, as outlined in Sect. 2.3. The mean Polymer TChl  $a$  data were validated as outlined in Sect. 2.4.

### 2.3 Statistical methods to retrieve pigment concentrations from reflectance

Different combinations of collocated pigment and AOP data sets were used for the construction of the statistical models. For the field data set, we used to set-ups of RRS( $\lambda$ ) data with collocated surface pigment data: RRS( $\lambda$ ) data in hyperspectral (1 nm resolved, “hyper\_RRS”) and in MERIS band resolution (“band\_RRS”). The satellite-based  $\rho_{\text{WN}}$  data set (“satellite\_ $\rho_{\text{WN}}$ ”) was run for the  $1 \times 1$ , the  $3 \times 3$  and the  $5 \times 5$  collocated data sets containing pigments and MERIS Polymer  $\rho_{\text{WN}}(\lambda)$  data. The following pigments were included in the construction of all different statistical models (further explained below): MVChl  $a$ , DVChl  $a$ , chlorophyllide  $a$ , MChl  $b$ , DVChl  $b$ , Chl  $c_{1/2}$ , (Chl  $c$  type 1 and 2), Chl  $c_3$  (Chl  $c$  type 3), Hex (19-hexanoyl-fucoaxanthin), But (19-butanoyl-fucoaxanthin), Allo (alloxanthin), Diadino (diadinoxanthin), Diato (diatoxanthin), Fuco (fucoxanthin), Lut (lutein), Peri (peridinin), Prasino (prasincoxanthin), Viola (violaxanthin), Zea (zeaxanthin),  $\alpha$ -caroten,  $\beta$ -caroten. In addition, the concentrations of different pigment groups were considered, such as the sum concentration of  $\alpha$ - and  $\beta$ -caroten (Caro), of MVChl  $a$ , DVChl  $a$  and chlorophyllide  $a$  (TChl  $a$ ), of MVChl  $b$  and DVChl  $b$  (TChl  $b$ ), of all pheopigments (TPheo) and according to Hooker et al. (2005) and Roy et al. (2011) the photosynthetically active carotenoids (PSC: Fuco, 19BF, 19HF, Peri), and the photoprotective carotenoids (PPC: Allo, Diadino, Diato, Viola, Lut, Zea, Caro). Also PE, although measured by a different technique, was included since it also had a major impact on the optical  $L_{\text{u}}(\lambda)$  data of the field data set (see Taylor et al., 2013). Chlorophyllide  $a$ , Neo, and PE were only considered in the models using the field data



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



al. (2013), where a covariance matrix of the dataset was subjected to Eigen decomposition with subsequent projection of data onto EOFs to derive PCs. The results of both approaches are similar, except that PCs  $\mathbf{U}$  derived via SVD are unitary and  $\Sigma$  contains standard deviation rather than variance. The SVD method is presented here due to its more straightforward notation; EOFs and PCs are determined in a single step whereas the alternate Eigen decomposition is a three-step calculation.

### 2.3.2 Log transformed general linear model

A general linear model was used to predict log-transformed pigment concentrations of each pigment  $y_p$  based on a subset of PCs,  $\mathbf{U}$ , as covariates. The linear model uses log-transformed pigment concentrations. Since only positive, non-zero values are permissible with this transformation, a small value was added to all concentrations ( $0.00001 \text{ mg m}^{-3}$ ) to allow for the inclusion of samples where pigment concentrations were essentially zero or below the detection limit. A truncated subset of PCs was used as defined by the magnitude of their standard deviation. PCs with standard deviations of  $\leq 0.0001$  times the standard deviation of the first component were omitted. The resulting multiple regression had the form:

$$\log(y_p) = a + b_1 u_1 + b_2 u_2 + \dots + b_n u_n, \quad (3)$$

where  $\log(y_p)$  is the natural log-transformed concentration of pigment  $p$ ,  $e_{1,2,\dots,n} u_{1,2,\dots,n}$  are the leading  $n$  PC scores from  $\mathbf{U}$ ,  $a$  is the intercept, and  $b_{1,2,\dots,n}$  are the regression coefficients. A bidirectional stepwise routine was used to search for smaller multiple regression models based on fewer predictor terms. Best linear models were selected through minimization of the Akaike information criterion (AIC). Once the best linear model was determined, the relative importance of included terms was defined by the change in AIC ( $\Delta\text{AIC}$ ) following each term's removal.

Since the range of concentration varies a lot among the different pigments, we calculated mainly relative error statistics. According to the GlobColour full validation report (ACRI, 2007), the coefficient of determination ( $R^2$ ), the slope ( $S$ ) and intercept ( $I$ ) of the

linear regression based on the log-scaled predicted ( $\log(y_p)$ ) versus the log-scaled observed ( $\log(y_o)$ ) pigment concentration data, and the root mean square error (RMSE), the mean percent difference (MPD), the percent bias (PB) and the median percent difference (MDPD) for the non-log transformed data were determined. The following equations for these statistics have been used:

$$R^2 = 1 - \frac{\sum_{i=1}^N [\log(y_{pi}) - \log(y_{oi})]^2}{\sum_{i=1}^N [\log(y_{oi}) - \log(y_o^-)]^2}, \quad (4)$$

with  $y_o^-$  as the mean value of the observed specific pigment concentration and  $i$  identifying the specific sample pair.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{pi} - y_{oi})^2} \quad (5)$$

$$\text{MPD} = \frac{100}{N} \sum_{i=1}^N |(y_{pi} - y_{oi})| / (y_{oi}) \quad [\%] \quad (6)$$

$$\text{PB} = \frac{100}{N} \sum_{i=1}^N (y_{pi} - y_{oi}) / (y_{oi}) \quad [\%] \quad (7)$$

$$\text{MDPD} = \text{Median of } \left[ |(y_{pi} - y_{oi})| / (y_{oi}) \cdot 100 \right]_{i=1, N} \quad [\%]. \quad (8)$$

### 2.3.3 Model prediction error

In addition to the statistics performed for each pigment linear model (Sect. 2.3.2), we performed a cross validation of the linear model fitting in order to better test the robustness of the models' prediction error. Data was split into two groups: the first part of the

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



data was used for model fitting, while the second part was used for prediction validation. According to Craig et al. (2012), we assessed the number of observations required to achieve adequate predictions by the pigment linear models using the variable jack-knife procedure of Wu (1986). So the data splitting for the cross validation procedure was varied as follows, with  $n$  = total number of samples,  $tp$  = number of training points and  $vp$  = number of points used for validation:

$$tp = n \cdot d, \text{ with } d = 0.1, 0.15, 0.2, \dots, 0.9 \quad (9)$$

$$vp = n(1 - d). \quad (10)$$

Since the number of permutations for data splitting definitely is restricted by computing time, the procedure was run for 500 permutations, similar to what was recommended by Craig et al. (2012). Such a high number of permutations rules out that the model error is assessed based on a spatially or temporally biased data set.

Each cross validation procedure was as follows:

1. For 500 permutations, do steps 2–9.
2. Randomly select  $n \cdot d$  of collocated samples to include in training sets  $\mathbf{X}^{\text{train}}$  and  $\mathbf{Y}^{\text{train}}$  for spectra and pigment data, respectively. Remaining  $n(1 - d)$  of samples are allocated to the validation sets  $\mathbf{X}^{\text{valid}}$  and  $\mathbf{Y}^{\text{valid}}$ .
3. Standardize  $\mathbf{X}^{\text{train}}$  and perform EOF following Eq. (2) to obtain  $\mathbf{U}^{\text{train}}$ ,  $\boldsymbol{\Sigma}^{\text{train}}$ , and  $\mathbf{V}^{\text{train}}$ .
4. For each pigment concentration  $y_p^{\text{valid}}$  of  $\mathbf{Y}^{\text{valid}}$  do steps 5–9.
5. Fit linear model to log-transformed pigment concentrations using selected  $\mathbf{U}^{\text{train}}$  as in Eq. (3).

$$\log(y_p^{\text{train}}) = a + b_1 u_1^{\text{train}} + b_2 u_2^{\text{train}} + \dots + b_n u_n^{\text{train}} \quad (11)$$

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



6. Perform bidirectional stepwise search for smaller linear model.

7. Standardize validation set and project  $\mathbf{X}^{\text{valid}}$  onto the EOFs  $\mathbf{V}^{\text{train}}$  and the inverse of singular values  $\mathbf{\Sigma}^{\text{train}^{-1}}$  to derive their PCs  $\mathbf{U}^{\text{valid}}$ :

$$\mathbf{U}^{\text{valid}} = \mathbf{X}^{\text{valid}} \cdot \mathbf{V}^{\text{train}} \cdot \mathbf{\Sigma}^{\text{train}^{-1}}$$

$$\log(y_p^{\text{valid}}) = a + b_1 u_1^{\text{valid}} + b_2 u_2^{\text{valid}} + \dots + b_n u_n^{\text{valid}} \quad (12)$$

8. Record pairs of observed and predicted validation pigment concentrations  $y_o$  and  $y_p^{\text{valid}}$  in a new object for all permutations for later calculation of prediction error.

For each permutation the  $R^2$  based on the log-scaled predicted ( $\log(y_p^{\text{valid}})$ ) versus the log-scaled measured ( $\log(y_p)$ ) were derived and finally over all permutations the mean value,  $R^2_{\text{cv}}$ , was calculated. Prediction error was described in terms of absolute squared difference and relative difference,  $(y_p^{\text{valid}} - y_o)^2$  and  $(y_p^{\text{valid}} - y_o)/y_o$ , respectively. Mean and median relative difference (MPDcv and MDPDcv, respectively) and the root mean square absolute difference (RMSEcv) over all permutation were determined, as follows:

$$\text{MPDcv} = \frac{1}{N} \sum_{i=1}^N [(y_p^{\text{valid}} - y_o)/y_o] \quad (13)$$

$$\text{RMSEcv} = \sqrt{\frac{1}{N} \sum_{i=1}^N [y_p^{\text{valid}} - y_o]^2} \quad (14)$$

$$\text{MDPDcv} = \text{median value of } \left( \frac{y_{i,p}^{\text{pred}} - y_{i,p}^{\text{valid}}}{y_{i,p}^{\text{valid}}} \cdot 100 \right), \quad i = 1, N[\%]. \quad (15)$$

### 2.3.4 Pigment concentration predictions with MERIS reflectance data

In order to predict pigment concentration from MERIS  $\rho_{wN}(\lambda)$  data where we did not have corresponding pigment measurements, we projected standardized MERIS  $\rho_{wN}(\lambda)$  data onto the EOF loading ( $\mathbf{V}$ ) to derive their principal components ( $\mathbf{U}$ ), which were subsequently used for the prediction with the fitted linear model (as in Sect. 2.3.3, step7, Eq. 12).

### 2.4 Validation of MERIS Polymer products with in-situ data

MERIS  $1 \times 1$  pixel matchup data and mean values of  $3 \times 3$  and  $5 \times 5$  pixel matchups were validated with the in situ TChl *a* data of the satellite-based data set. The  $R^2$ , PB, MPD and the RMSE between the two collocated data sets were calculated, as outlined in the GlobColour full validation report (ACRI, 2007) and used to determine pigment prediction full fit statistics (see Sect. 2.3.2).

## 3 Results and discussion

### 3.1 Characteristics of input data sets

Figure 1 presents the distribution of collocated pigment and reflectance measurements for both data sets which were used as input for the EOF analysis. The field data set with 53 collocations has been obtained in two seasons, spring and fall, in 2008 and 2010, while the satellite-based data set consisted of 155, 150 and 135 collocated samples from 2002 to 2012 for the  $5 \times 5$ ,  $3 \times 3$  and  $1 \times 1$  pixel collocation, respectively, covering all months except January, March and December. Fig. 2 shows the original and standardized spectra of the field and satellite-based data sets. Considering the conversion between  $RRS(\lambda)$  to  $\rho_{wN}(\lambda)$  data by a factor of  $\pi$ , magnitude and shape of the original and standardized spectra are similar for the band resolved data sets, except that the standardized satellite- $\rho_{wN}$  data set contains only one spectrum with maximum

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





reflectance in the green, at 560 nm, while the standardized field data set contains four spectra with maxima at 510 nm.

The composition and range of pigments is as well similar for both data sets (as detailed in Supplement Table S2). However, for all pigments (except for Fuco for which it is equal, and for Zea for which it is vice versa) the collocations to the field data set contain higher maxima and minima than the collocations to the satellite-based data set. The higher concentration of total pigments in the field data set may explain the small differences in the shape of the reflectance spectra of the two (field versus satellite-based) data sets. However, DVChl *b*, MVChl *b*, TChl *b*, Allo, Diato, Lut, Neo, Peri, Viola, Pras, Chlorophyllide *a* and TPheo had values of 0 mg m<sup>-3</sup> in more than 20 % of all stations in both data sets. Also Chl *c*<sub>3</sub> had a concentration of 0 mg m<sup>-3</sup> in one sample collocated to the field and in over 30 % of samples collocated to the satellite-based data set. Several pigments had occasionally (< 10 %) concentrations of 0 mg m<sup>-3</sup> in samples collocated to the satellite-based data set: Caro, Chl *c*<sub>1/2</sub>, 19BF, 19HF, Zea, DVChl *a*, Diadino and Fuco, the later three also for the field data sets. All other pigments not listed here reached in all samples concentrations higher than 0 mg m<sup>-3</sup>.

### 3.2 EOF analysis – shape of modes and relevance for predictions

The decomposition of the standardized spectra by EOF analysis returned nine significant modes (EOF-1 to EOF-9) for the hyper\_RRS and seven significant modes for the band\_RRS and satellite\_ρ<sub>WN</sub> data sets (the first four modes are presented in Fig. 3) given our inclusion criterion based on the explained standard deviation relative to EOF-1 (see Sect. 2.3.2). For all data sets, the first three included modes explain over 99.8 % of the variance for all three data sets and EOF-1 explains already between 94.5 and 96 % of the variance (Table 1).

The shapes of the first three EOF modes are very similar among all three reflectance data sets. They are nearly identical for the band\_RRS and the satellite\_ρ<sub>WN</sub> data sets, but show smoother shapes and peaks for hyper\_RRS for the first two modes. Still one has to bear in mind that although the RAMSES measurements deliver 1 nm resolved

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



data, the real spectral resolution of the sensors is 3.3 nm. the hyper\_RRS data. Be-  
cause of the limited number of wavelengths for the two multispectral data sets, starting  
from EOF-3 their peaks are clearly shifted (peak at 412 and 443 nm for EOF-3 and  
EOF-4, respectively) as compared to hyper\_RRS (peak at 360 and 410 nm for EOF-3  
and EOF-4, respectively) where the narrow spectral resolution allows for more pre-  
cision in identifying spectral regions of higher variance. For EOF-4, the satellite\_ρ<sub>WN</sub>  
mode is much flatter beyond 500 nm and shows no trough between 600 and 650 nm as  
opposed to the EOF-4 for the two other data sets. No much similarity is seen among  
the EOF-5 of the different spectra data sets, while for EOF-6 the two field data sets  
are similar in the overall shape and peak positions which are in contrast shifted to-  
wards longer wavelengths for the satellite data set. EOF-7 and EOF-8 show very simi-  
lar shapes for hyper\_RRS and deviate from EOF-7 of the band data sets, while EOF-9  
from hyper\_RRS looks much more like the later ones.

The EOF analyses deliver modes of oscillation which can be interpreted as imprints  
of changes in the optical properties of water constituents in the water column: Com-  
pared to the shape of spectra obtained in case-2 waters by Lubac and Loisel (2007)  
and Craig et al. (2012) only our reflectance spectra taken in high TChl *a* waters with  
mineral fraction (identified as cluster V for the ANTXXV/1 data in Taylor et al., 2011)  
correspond to part of the spectra presented in those studies (e.g. class 5 in Lubac and  
Loisel 2007). While all our other spectra (typical case-1-water) are not contained in the  
data set of those studies. This explains the minor differences in the shape and loading  
of EOFs between their and our data set. In the following we focus the discussion on  
our hyper\_RRS data set results since also the Craig et al. (2012) study was based on  
hyper-spectral RRS data.

Our first three EOF modes more or less correspond to the ones derived for the hy-  
perspectral case-2 reflectance data set of Craig et al. (2012). As pointed in their study,  
EOF-1 is likely the signature of bulk oscillations in phytoplankton biomass concentra-  
tion. However, our EOF-1 already explains much more of the variance as compared to  
Craig et al. (2012) where it only accounted for 72.4 % and showed much more structure

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



and a weaker exponential decrease from 400 to 550 nm. This indicates that in our open ocean data set, the change in total attenuation is the main difference among the different sampled stations, which is mainly reflecting the attenuation as affected by the total pigment concentration. Our data set was largely composed of samples from waters with lower TChl *a* concentration, ranging from 0.005 to 3.553 mg m<sup>-3</sup>, while in the study of Craig et al. (2012) it ranged from 0.584 to 18.020 mg m<sup>-3</sup>. EOF-2 superficially resembles the oscillation in the amplitude of RRS which also is affected by overall changes in the total absorption over broad band structures. It is strongly decreasing from 350 to 510 nm and again increasing above 570 nm, which is connected to total pigment and water absorption, respectively. There is a peak around 683 nm which can be linked to MVChl *a* and DVChl *a* fluorescence. While this peak is present in EOF-1 and EOF-2 in the Craig et al. (2012) data set, it is only apparent in EOF-2 of our data set probably because of the lower TChl *a* concentrations.

EOF-3 of our data set as compared to the one of Craig et al. (2012) shows a much steeper decrease with wavelength in the blue spectral range. These changes may reflect concomitant changes of absorption by chlorophyll and non-algal particles which are expected to be co-varying and of much lower concentration in our case-1-waters, as the scattering by other particles than phytoplankton was much higher in the case-2-water of Craig et al. (2012) leading to a less steep slope of this EOF mode. EOF-4 appears different in the relation of the three peaks. Similar as for EOF-2 and EOF-3, these differences are caused by the different composition and overall loading of water constituents of our and their sampled stations. All higher EOF were not presented in Craig et al. (2012) because they were not used to predict TChl *a* from RRS data, as it was the case for our TChl *a* (and MVChl *a*) linear model predictions (Sect. 3.3.3). EOF modes higher than four were probably reflecting imprints of specific pigment groups or pigments, as indicated by the results of the  $\Delta$ AIC values and further pointed out in the end of the next section (Sect. 3.3.3).

### 3.3 Pigment prediction by linear models

#### 3.3.1 Field data set linear models

All pigments which were apparent in all samples of the field data set were well predicted by linear models based on hyperspectral (hyper\_RRS) or the reduced eight band (band\_RRS) resolution. Correlations between predicted and observed pigment concentrations were based on a significance level of  $p < 0.0001$  and cross validation statistics reached reasonable quality with  $R^2_{cv} \geq 0.5$ , MDPD<sub>cv</sub>  $\leq 45\%$  and MPD<sub>cv</sub>  $\leq 60\%$  (Table 2a, upper part). For some pigments (TChl *a*, MVChl *a*, Hex, PSC) EOFs based on 380 to 700 nm produced much better linear model results using hyper\_RRS data than based on 350 to 700 nm. Plots of observed versus predicted values for the full data set of pigments TChl *a*, PSC, PPC, Hex and Zea are shown in Fig. 4. Lower quality for one statistical parameter for both linear models was reached for Zea ( $R^2_{cv}$  0.31 and 0.27), But (MPD<sub>cv</sub> 81 and 95 %) and for two parameters for PE (MDPD<sub>cv</sub> 65 and 67 %, MPD<sub>cv</sub> 139 and 156 %).

For all other pigments, predictions were of low quality (results not shown), demonstrating that the linear model approach does not produce robust predictions as soon as a pigment is not present (i.e.,  $0 \text{ mg m}^{-3}$ ) in every sample. The replacement of concentration of  $0 \text{ mg m}^{-3}$  with  $0.00001 \text{ mg m}^{-3}$  for specific pigments did not enable robust linear model construction and produced large errors, especially for the cross validation statistical parameters. We re-ran the predictions for specific pigments where only a few samples ( $< 10\%$ ) had concentrations of  $0 \text{ mg m}^{-3}$ , as it was the case for DVChl *a*, Fuco, Diadino and Chl *c*<sub>3</sub> (see Supplement Table S2). In those specific linear model runs we only included as input data the data points where the specific pigment concentrations were  $> 0 \text{ mg m}^{-3}$ . The resulting predictions (Table 2a, lower part, and for DVChl *a* see full-fit results in Fig. 4d) from using the adjusted input data for those pigments show robust and significant cross validation results within the same quality range as for the pigments which were detected in all data. No robust predictions were obtained for all other pigments which reached in less than 80 % of all samples concentrations  $> 0 \text{ mg m}^{-3}$ ,

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



even when in the specific linear model runs we only included as input data the data points with specific pigment concentrations  $> 0 \text{ mg m}^{-3}$  (results not shown).

Cross validation results of well predicted pigments (Table 2a) show that, especially regarding the  $R^2_{cv}$  and  $RMSE_{cv}$  values, the hyper\_RRS based linear models perform either slightly better (PPC, PSC, Chl  $c_{1/2}$ , But, Chl  $c_3$ ) or much better (TChl  $a$ , MVChl  $a$ , Hex, Zea, Caro, PE, DVChl  $a$ , Diadino, Fuco) than reducing the data set to eight wavelengths (band\_RRS data set). Especially,  $RMSE_{cv}$  is much lower for pigment predictions where  $RMSE_{cv}$  reaches high values ( $> 0.10 \text{ mg m}^{-3}$ ), i.e., for TChl  $a$ , Fuco, MVChl  $a$ , PE, Diadino, and Hex. Considering the  $MDPD_{cv}$  and  $MPD_{cv}$  values it is less clear for MVChl  $a$ , Chl  $c_{1/2}$ , TChl  $a$  and PSC predictions. For the later the multispectral resolution seems to be sufficient to obtain similar robust linear models. TChl  $a$  (MVChl  $a$  in line with that) and PSC are dominating the overall phytoplankton pigment composition and absorption. TChl  $a$  has been shown to be well retrieved by band-ratio algorithms for the main phytoplankton biomass indicator (e.g. see Brewin et al., 2014). For pigments which are very similar in their spectral range, such as But, Hex and Fuco, the hyperspectral resolution of the linear models provides much more robust pigment predictions (Table 2a). The hyper\_RRS linear models also produced better predictions for DVChl  $a$ , Zea, Diadino and PPC where the specific linear models require more than the first seven EOF modes (see Sect. 3.3.3). These are not available using the multispectral resolution of RRS data.

### 3.3.2 Satellite-based data set linear models

Results for the models predicting pigment concentration from the satellite-based data set were very similar when using  $1 \times 1$ ,  $3 \times 3$  or  $5 \times 5$  collocated MERIS  $\rho_{WN}$  data. Deviations were within 1 to 3% for all statistical parameters.  $R^2_{cv}$  values were best in all cases for well predicted pigment concentrations in the  $1 \times 1$  collocations, while  $MPD_{cv}$  was best in the  $3 \times 3$  collocations. For simplicity, in the following we are presenting and discussing the results of the  $1 \times 1$  collocated reflectance data only.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



In line with field data linear model results, pigment groups and pigments which reached in every sample concentrations  $> 0 \text{ mg m}^{-3}$  (MVChl *a*, TChl *a*, PSC and PPC; the full-fit linear model results are shown in Fig. 5a–c) were well predicted with similar cross validation statistics values using the satellite  $\rho_{wN}$  data set (Table 2b, upper part). Also good predictions for some pigments (DVChl *a*, Zea, Diadino, Hex, But, Fuco and Chl  $c_{1/2}$ ) could be obtained by re-running the linear model analysis by excluding collocations with respective pigment concentrations of  $0 \text{ mg m}^{-3}$  (Table 2b, lower part). For DVChl *a*, Hex and Zea exemplarily results of the full-fit linear model are shown in Fig. 5d–f, respectively. Though, some of these pigments show only medium quality for one cross validation statistical parameter (lower  $R^2_{cv}$  for DVChl *a* and Zea, higher MPD<sub>cv</sub> for Fuco, Chl  $c_{1/2}$  and Diadino). Similar to the field data linear models, no robust predictions were obtained for all other pigments which reached in less than 80 % of all samples concentrations  $> 0 \text{ mg m}^{-3}$ , even when only data points with specific pigment concentrations  $> 0 \text{ mg m}^{-3}$  were included (results not shown).

### 3.3.3 EOF modes relevant for pigment predictions

Results of the  $\Delta AIC$  showing the significance of each EOF mode for the main pigment prediction linear models are presented in Table 3. For the hyper\_RRS data set, the prediction linear models used EOF-2 and EOF-3 for all pigments. EOF-2 was the most relevant in the respective models for all pigment prediction, except for Zea and DVChl *a* where EOF-3 was the most important and closely followed by several other EOF modes. For all other well-predicted pigments, EOF-3 followed EOF-2 by importance, except for Chl  $c_3$  (EOF-4) and for PE (EOF-1). Besides PE, EOF-1 only was used (with medium importance) for But, DVChl *a* and Zea linear models.

As discussed Sect. 3.2, EOF-2 is reflecting the optical imprint of all phytoplankton pigments. The high  $\Delta AIC$  value of EOF-2 for most pigments' linear models is probably caused by that the concentration of these specific pigments and most phytoplankton groups increase when TChl *a* increases. In contrast to that, cyanobacteria

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



and especially its subgroup *Prochlorococcus*, containing the marker pigments Zea and DVChl *a*, respectively, are the most abundant phytoplankton under low TChl *a* concentrations. This has been manifested in the abundance-based algorithms to retrieve picoplankton from TChl *a* data (Uitz et al., 2006; Hirata et al., 2011). This may explain why predictions of those marker pigments by our linear models show lower  $\Delta$ AIC for EOF-2 and require several different EOF modes in their linear models.

For DVChl *a* all nine EOF-9 and for Zea EOF-1 to EOF-4 and EOF-5 to EOF-8 were incorporated in their respective linear model. As in Craig et al. (2012), EOF-2 to EOF-4 were relevant for our hyper\_RRS based TChl *a* and MVChl *a* predictions. EOF models developed by Taylor et al. (2013) to predict PE concentrations based on *Lu* data required the first four EOF modes, while our PE prediction based on RRS data required the first three EOFs only. For all other pigments, also the higher EOFs were necessary for robust predictions.

Similarly to the hyper\_RRS linear models, the two multispectral linear models also showed EOF-2 to be the most important predictor for specific pigment models, except for DVChl *a* (both models) and Zea (only band\_RRS). However, compared to the hyper\_RRS linear models, much more EOF modes from the multispectral data were needed for all specific pigment models.

### 3.4 Number of data points to construct robust models

Our presented linear models to predict specific pigment or pigment group concentration are calibrated for an ocean colour data set of a specific region with coincident pigment measurements. Results of the variable jack-knife procedure indicate that the selection of minimal training points to set up a robust linear model for predictions varies among pigments and pigment groups and also among all three statistical parameters: the ratio of  $R^2_{cv}$  to  $R^2$  ( $R^2_{cv}/R^2$ ), the ratio of MPD<sub>cv</sub> to MPD ( $MPD_{cv}/MPD$ ) and RMSE<sub>cv</sub>, shown exemplarily for predicting TChl *a*, PSC, PPC and PE in Fig. 6. While  $R^2/R^2_{cv}$  (Fig. 6a, d) already drops below 0.8 and then decreases exponentially for PPC with each step diminishing the number of training points below 50 for all linear models, for

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



all other pigment predictions this is the case when it drops below 30 data points and even 15 data point for the hyper-RR<sub>S</sub> PE linear model. The slope of increase in RMSE<sub>cv</sub> (Fig. 6c, f) varies among pigments and linear models. It is especially high for band-RR<sub>S</sub> TChl *a* and satellite-ρ<sub>wN</sub> PSC predictions (< 45 and < 70 training points, respectively). For the other predictions, RMSE<sub>cv</sub> indicate that more than 40 and 50 training points are required for the field and satellite-based data linear models, respectively. MPD<sub>cv</sub>/MPD below 1.4 indicating robust fits considering this criterion only are obtained for all pigments above 40 training points for the satellite-ρ<sub>wN</sub> (Fig. 6e) and above 30 for the hyper-RR<sub>S</sub> data sets (Fig. 6b). Generally, we observe that the band-RR<sub>S</sub> as compared to the hyper-RR<sub>S</sub> linear model results deteriorate faster with a decreasing number of samples used for training, especially for TChl *a* and PE. For the set-up of linear models at least 45 to 50 training data points are required, while for some pigments (e.g. TChl *a*) using the hyper-RR<sub>S</sub> data as input only 25 training data points are necessary. Based on the results we are confident that for both, the field and the satellite-based, data sets our number of data points used for linear model construction and cross validation, chosen for results presented in Sect. 3.3, had been adequate for robust predictions. The number of collocated PE samples seems to have been too small, especially for the multispectral resolution, for predicting robust PE concentrations.

### 3.5 Comparison to other approaches deriving pigment concentration

Our hyper-RR<sub>S</sub> TChl *a* linear model results ( $R^2 = 0.82$ , RMSE = 0.30,  $R^2_{cv} = 0.77$ , RMSE<sub>cv</sub> = 0.41; Fig. 4 and Table 2) are comparable to results by Craig et al. (2012;  $R^2 = 0.84$ , RMSE = 0.30,  $R^2_{cv} = 0.76$ , RMSE<sub>cv</sub> = 0.21). Even though Craig et al. (2012) used measurements only from one location and sampled about weekly throughout one year, while our field data set was from a much larger region (covering 95° in latitudes and 85° in longitude) and sampled at two seasons in 2008 and 2010 only. In their study the same linear model set-up was used with collocated in situ reflectance and TChl *a* data sampled at Compass Buoy Station in the Bedford Basin near Halifax as input



data. The number of collocations used for training to obtain robust results for TChl *a* predictions was also similar for both studies, with more than 25 recommended for our hyper-RR<sub>S</sub> linear model and more than 15 for the Craig et al. (2012) linear model.

Chase et al. (2013) used Gaussian functions to derive different chlorophylls, PSC and PPC concentrations from a large global data set of hyperspectral particulate absorption measurements. Their validation results showed MDPD values between predicted and observed concentrations of 30–36, 40–53, 49 and 51 % for TChl *a*, Chl *c*, PSC and PPC, respectively. Our three linear models show similar (TChl *a* 27–32 %) or even much better MDPD<sub>cv</sub> values (Chl *c*<sub>1/2</sub>: 33–41 %, PSC: 32–42 %, PPC: 25–27 %) which indicates that our method produces robust results, also considering the fact that we use a more indirect measure of pigments, an AOP (reflectance), as opposed to the IOPs used in their study.

Pan et al. (2010) developed pigment specific band-ratio algorithms with collocated in situ RRS( $\lambda$ ) and pigment measurements from the United States northeast coast. Those algorithms are based on deriving pigment specific coefficients for third order polynomial functions using the band ratio of either 490 to 550 nm or 490 to 670 nm (for SeaWiFS; for MODIS changed accordingly to MODIS bands 488 and 547 nm). Validation of results with collocated satellite (SeaWiFS and MODIS) reflectance data and pigment concentrations showed very good quality (MPD, RMSE and  $R^2$  range from 36 to 48 %, 0.23 to 0.29 and 0.65 to 0.90, respectively, for SeaWiFS, with similar results for MODIS) for several pigments, among them TChl *a*, TChl *c*, Caro, Fuco, Diadino and Zea. This method was modified to the Northern South China Sea accordingly using globally derived relationships and locally identified links between pigment concentration and sea surface temperature (Pan et al., 2013). They obtained similar validation results as in Pan et al. (2011). Compared to our linear model results the quality to predict pigment concentration is of similar quality: while our results for MDPD<sub>cv</sub> and  $R^2$ <sub>cv</sub> are slightly worse (42 to 50 % and 0.61 to 0.80, respectively), our results for RMSE<sub>cv</sub> (0.06 to 0.18 mg m<sup>-3</sup>, except for TChl *a* 0.41 mg m<sup>-3</sup>) are much better.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



PE is not too well predicted by both our linear models based on the field data set. Still hyper- $RR_{5}$  linear model cross validation measures are much better than the PE band- $RR_{5}$  linear model. In Taylor et al. (2013), PE concentrations were predicted from the same underwater light measurements but using  $L_u$  instead of RRS data. Their pigment data included besides surface samples also samples taken from deeper depths. No cross validation was performed within their study. Our results for  $R^2_{cv}$  (0.69) are even better than their results for using the data from all three cruises for predictions ( $R^2$  of 0.58). The data set of Taylor et al. (2013) was nearly three times larger than our field data set and a log-link Generalized Linear Model (GLM) was used instead of a log-transformed linear model. For the later we tested both settings for our pigment linear models. Cross validation results were similar for PE using the log-link GLM instead of log-transformed linear model, but worse with GLM for the other pigment predictions.

As for TChl  $a$  predictions from the satellite- $\rho_{wN}$  linear model, validation results of the MERIS Polymer TChl  $a$  product considering collocations with in situ TChl  $a$  from the satellite-based data set showed marginal differences for the  $1 \times 1$ ,  $3 \times 3$  or  $5 \times 5$  MERIS pixel collocations (Table 4, upper panel). The TChl  $a$  Polymer product obtained 3% higher MPD, similar  $R^2$  and PB, and an about one third lower RMSE values than the TChl  $a$  linear model predictions. In the global validation by Brewin et al. (2014) among various TChl  $a$  satellite products the OC4V6 (Ocean-Chlorophyll-4 algorithm version 6, O'Reilly et al., 2000) was selected as the best TChl  $a$  algorithm. This algorithm is used to produce the MERIS Polymer TChl  $a$  from atmospheric corrected MERIS Polymer data. Global validation by Brewin et al. (2014), with 1039 collocations and retrieving TChl  $a$  directly from in situ  $\rho_{wN}(\lambda)$  data, showed for OC4V6  $R^2$  of 0.87 and RMSE of 0.29. We conclude that both MERIS Polymer TChl  $a$  products, level-2 and linear model, although they are retrieved from satellite data and not in situ  $\rho_{wN}$  data, show high quality within the eastern Atlantic Ocean.

All together the comparison to other methods retrieving pigment concentrations from reflectance data shows that our method based on a linear model using EOFs from reflectance data gives robust results for pigment groups and pigments which are always

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



present in the region investigated. The advantage of our linear models, either set-up with reflectance data measured directly in the ocean water or obtained from a satellite ocean colour sensor, is that we can obtain robust results for other pigment groups and some specific pigments as well. For the Eastern Tropical Atlantic Ocean data set, these additional pigments (other than TChl *a*) include PPC, PSC, DVChl *a* and MChl *a*. To some extent we can claim that even more pigments can be predicted when the linear model runs are adjusted to a data set which only incorporates samples from a region where the specific pigment is measured in every sample. Generally, we can also see from the field data linear models, that using a coherent in situ data set where all pigments have been measured by the same method and instrumentation will provide a wider range of pigment predictions because also the pigment data, used for linear model fitting and validation, have a more homogeneous error. An advantage of our linear method to pigment specific band algorithms is that we require a much smaller data set for establishing the prediction (about 50 as opposed to several hundreds) of collocated pigment and reflectance data.

### 3.6 Application of linear model to study large scale pigment distributions

For demonstrating the application of our linear model, we used the satellite  $\rho_{wN}$  specific pigment's full-fit models for TChl *a*, MVChl *a*, PSC and PPC and run these specific models using November 2008 MERIS Polymer  $\rho_{wN}$  level-2 data to retrieve those pigments for an example time period on a larger spatial scale. By subtracting the MVChl *a* value from TChl *a* we also derived concentrations of DVChl *a*. Figure 7 shows the monthly averages for those various pigment groups and pigments. Also the MERIS Polymer TChl *a* concentration for the same time and region is shown.

The distribution of TChl *a* from the EOF model prediction or from the Polymer algorithm are very similar ranging from 0.00003 to 7.52 mg TChl *a* m<sup>-3</sup>. For this particular month, the total biomass of phytoplankton shows a strong phytoplankton bloom (> 2 mg m<sup>-3</sup>) at the Mauritanian Upwelling spread in two parts, 19–24° N and 14–7° N, and high values (> 0.5 mg m<sup>-3</sup>) at all coastal areas of the African continent. Enhanced

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



TChl *a* concentrations  $> 0.3 \text{ mg m}^{-3}$  are also spreading into the open ocean especially at 5–20° N and 30–40° W, along the 0° latitude across from Africa to South America, and South of this at 3–10° S from 3° E to about 25° W. MVChl *a* follows more or less the TChl *a* distribution, however only at the northern bloom it reaches the magnitude indicated by the TChl *a* values. The deviation between TChl *a* and MVChl *a* is obvious in the distribution of DVChl *a* which indicates that at the northern part of the Mauritanian Upwelling bloom *Prochlorococcus* (the only phytoplankton genus which contains DVChl *a*) seems to have contributed to this bloom by only a very minor fraction (few percent), while elsewhere it presents a substantial background of about 30 % of all phytoplankton.

Our predicted PPC concentrations show values in the same range as TChl *a* at the oligotrophic areas and about 50 % in the enhanced TChl *a* areas and the southern part of the bloom. As for DVChl *a*, in the northern part of the bloom PPC concentrations are significantly lower and only contribute to less than 10 % to the total pigment concentrations. PSC concentration in the oligotrophic and enhanced TChl *a* areas are much lower than PPC or even DVChl *a* concentrations, but reflect more or less on the large scale the TChl *a* distribution. Within the northern part of the Mauritanian upwelling PSC concentrations reach even values as high as for TChl *a*, while concentrations at the bloom further south contribute only to less than 10 % of the total pigment concentrations. In Taylor et al. (2011) the analysis of pigment and additional microscopic data clearly showed very high concentrations of Fuco, a main pigment of PSC, and a high dominance of diatoms within water samples at the Northern bloom collected at the same time period.

From our results, we can conclude, that the Northern phytoplankton bloom at the Mauritanian upwelling seems to have been freshly growing with very high photosynthetic activity while for most of the other areas a lot of the energy build up via photosynthesis was used for photo-protection. We have no information on photodegradation since no significant prediction linear model could be developed for pheopigments. These pigments had only been identified in less than 60 % of all samples collocated to

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the field and satellite-based data sets and the results show that this pigment group was not well predicted by the linear model. Based on the biogeography of Longhurst (2006), the oligotrophic areas on our maps fall into the North Atlantic Subtropical Gyre Province East at  $> 25^\circ$  N (the border between the two is the subtropical convergence) and the North Atlantic Tropical Gyre Province at  $25^\circ$  N to about  $12^\circ$  N. At the eastern corner towards the coast of these provinces, the Canary Coastal Province (CNRV), concentrations of all predicted pigments and pigment groups may have been increased due to eddy-driven processes increasing the supply of nutrients. In Taylor et al. (2011), the two blooms analysed by field samples at CNRV have been identified to cluster differently due to their pigment composition. The northern “fresh” bloom with low photo-inhibition, high dominance of phytoplankton, and strong photosynthetic efficiency was related to a major upwelling focusing in the area south of Cape Blanc (Western Sahara) off the coast of Mauritania. DVChl *a* was absent in this bloom which is in line with our results obtained from the linear model. The Southern part of the CNRV bloom was placed within the African dust veil where mineral rich dust fertilizes the ocean. Clearly in the Northern bloom stations the spectral shape and absolute values of particulate non-phytoplankton absorption spectra, presented in Taylor et al. (2011) which were coinciding with the pigment data used in our study, indicated that mineral particles absorption was very high.

## 4 Conclusions

We present robust predictions of concentrations of various pigments and pigment groups from linear models based on fitting empirical orthogonal function on a set of reflectance data to collocated pigment concentrations. Spectral shapes of the reflectance spectra from the Eastern Atlantic and of their derived EOF modes reflect typical case-1 water characteristics. In our study, it was shown that not only hyper-spectral under-water radiometric measurements enable reliable predictions of concentration of nine different pigments/pigment groups (TChl *a*, PPC, PSC, MVChl *a*, Chl *c*<sub>1/2</sub>, But, Hex,

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Zea, Caro, PE), but also multi-spectral reflectance data from field or satellite (MERIS Polymer) data which are collocated to pigment data can be used to establish predictive linear models based on EOFs. A limitation for all predictions is that only pigments can be predicted which have been identified in every collocated sample and that adding a small value ( $0.0001 \text{ mg m}^{-3}$ ) was not an appropriate solution to this problem.

The method proves for the first time to be applicable for predicting concentrations of not only TChl *a* and PE, but also of other pigments and pigment groups with weaker peculiar imprints on the underwater light field. Statistical resampling used for cross-validation indicates that predictions were robust ( $R^2_{cv} \geq 0.5$ , MDPD<sub>cv</sub>  $\leq 44\%$  and MPD<sub>cv</sub>  $\leq 60\%$ ) for all pigments (except for PE, Zea and 19BF, which deviated for one of these measures) and pigment groups. Hyperspectral linear models proved to be already stable with less collocated samples for most pigment or pigment groups used for training ( $n > 30$  to 40) than linear models based on multispectral reflectance data ( $n > 50$ ). The linear models using MERIS Polymer reflectance data as input were applied to one month of satellite data to predict the concentrations of TChl *a*, PSC, PPC, MVChl *a* and DVChl *a* for the whole Eastern Tropical Atlantic. For the first time a consistent picture of several phytoplankton pigments indicating group specific behaviour and photo-physiology on larger spatial scale for this area were shown.

Our presented linear models are generic and can be applied to even a small, consistent collocated reflectance and pigment data set to enable various specific pigment predictions from continuous optical measurements. The optical data can be obtained from radiometric measurements based on various platforms (buoys, gliders, floats or satellite). On a global scale, TChl *a*, PSC and PPC are persistently predicted accurately, while other pigments may be better predicted on smaller spatial scales. Highly temporally resolved time series data, which depending on the platform even may be of good spatial coverage, can be used to study variability and change of overall phytoplankton and photo-physiological response to environmental variables. While we established the linear models for prediction of various pigments in typical case-1 waters, the method should be tested in the future for its applicability in case-2 waters as well.



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Barker, K., Mazeran, C., Lerebourg, C., Bouvet, M., Antoine, D., Ondrusek, M., Zibordi, G. and Lavender, S.: MERMAID: MERis MAtchup *In-situ* Database. Proceedings of the 2nd (A)ATSR and MERIS Workshop, Frascati, Italy, September 2008.

Barlow, R. G., Cummings, D. G. and Gibb, S. W.: Improved resolution of mono- and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC, *Marine Ecol. Prog. Ser.*, 161, 303–307, 1997.

Brewin, R. J. W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B. A., Grant, M. G., Groom, S. B., Horseman, A., Hu, C., Krasemann, H., Lee, Z., Maritorena, S., Mélin, F., Peters, M., Platt, T., Regner, P., Smyth, T., Steinmetz, F., Swinton, J., Werdell, J., and White III, G. N. The Ocean Colour Climate Change Initiative: A round-robin comparison on in-water bio-optical algorithms, *Remote Sens. Environ.*, in press, 2014.

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Dall'Olmo, G., and Westberry, T. K.: Decomposition of in situ particulate absorption spectra, *Methods Oceanogr.*, 7, 110–124, 2013.

Craig, S. E., Jones, C. T., Li, W. K. W., Lazin, G., Horne, E., Caverhill, C., and Cullen, J. J.: Deriving optical metrics of ecological variability from measurements of coastal ocean colour, *Remote Sens. Environ.*, 119, 72–83, 2012.

Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., and Yamanaka, Y.: Synoptic relationships between surface Chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types, *Biogeosciences*, 8, 311–327, doi:10.5194/bg-8-311-2011, 2011.

Hooker, S. B., Van Heukelem, L., Thomas, C. S., Claustre, H., Ras, J., and Barlow, R.: The second SeaWiFS HPLC analysis round robin experiment (SeaHARRE-2), NASA TM/2005-212785, NASA Goddard Space Flight Center, Greenbelt, Maryland, 2005.

Longhurst, A.: *Ecological Geography of the Sea*, 2nd Edn., Elsevier Academic press, USA, 2006.

Lubac, B. and Loisel, H.: Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea, *Remote Sens. Environ.*, 110, 45–48, 2007.

Mackey, M. D., Mackey, D. J., Higgins, H. W., and Wright, S. W.: CHEMTAX – a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton pigments, *Marine Ecol. Progr. Ser.*, 144, 265–283, 1996.



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- McClain, C. R.: A Decade of Satellite Ocean Color Observations, *Ann. Rev. Marine Sci.*, 1, 19–42, doi:10.1146/annurev.marine.010908.163650, 2009.
- Morel, A. and Gentili, B.: Diffuse reflectance of oceanic waters. 2. Bidirectional aspects, *Appl. Opt.*, 32, 6864–6872, 1993.
- 5 Morel, A. and Gentili, B.: Diffuse reflectance of oceanic waters. 3. Implication of bidirectionality for the remote-sensing problem, *Appl. Opt.*, 35, 4850–4862, 1996.
- Morel, A., Voss, K. J., and Gentili, B.: Bidirectional reflectance of oceanic waters: A comparison of modeled and measured upward radiance fields, *J. Geophys. Res.*, 100, 13143–13150, 1995.
- 10 Müller, D. and Krasemann, H.: Product validation and algorithm selection report, part 1 – atmospheric correction. Tech. Rep. AO-1/6207/09/I-LG D2.5, European Space Agency, ESRIN, 2012.
- Müller, D., Krasemann, H., Brewin, R. J. W., Brockmann, C., Deschamps, P.-Y., Doerffer, R., Fomferra, N., Franz, B. A., Grant, M. G., Groom, S. B., Mélin, F., Platt, T., Regner, P., Sathyendranath, S., Steinmetz, F., and Swinton J.: A Methodology for assessing Atmospheric Correction Processors based on in situ measurements, *Remote Sens. Environ.*, in press, 2014.
- 15 O'Reilly, J. E., Maritorea, S., Siegel, D., and O'Brien, M. C.: Ocean color chlorophyll a algorithms for SeaWiFs, OC2, and OC4: Technical report, in: *SeaWiFS postlaunch calibration and validation analyses, Part 3.*, edited by: Toole, D., Mitchell, B. G., Kahru, M., Chavez, F. P., Strutton, P., Cota, G., Hooker, S. B., McClain, C. R., Carder, K. L., Muller-Karger, F., Harding, L., Magnuson, A., Phinney, D., Moore, G. F., Aiken, J., Arrigo, K. R., Letelier, R., Culver, M., Hooker, S. B., and Firestone, E. R., NASA, Goddard Space Flight Center, Greenbelt, Maryland. *SeaWiFS Postlaunch Technical Report Series*, 11, 9–23, 2000.
- 20 Organelli, E., Bricaud, A., Antoinem D., and Uitz, J.: Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site), *Appl. Opt.*, 52, 2257–2273, 2013.
- Pan, X., Mannino, A., Russ, M. E., Hooker, S. B., and Harding, L. W.: Remote sensing of phytoplankton pigment distribution in the United States northeast coast, *Remote Sens. Environ.*, 114, 2403–2416, 2010.
- 30 Ras, J., Claustre, H., and Uitz, J.: Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: comparison between in situ and predicted data, *Biogeosciences*, 5, 353–369, doi:10.5194/bg-5-353-2008, 2008.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.R-project.org/> (last access: 19 December 2013), 2013.

Roy, S., Llewellyn, C. A., and Egeland, E. S.: Phytoplankton pigments: characterization, chemotaxonomy and applications in oceanography, Cambridge University Press, 2011.

Speich, S., Boyé, M., and Dehairs, F.: IPY BONUS-GoodHope cruise Report. May 2008, IPEV-INSU-LEFE-IFREMER-ANR-IRD-IUEM-UBO, 2008.

Steinmetz, F., Deschamps, P. Y., and Ramon, D.: Atmospheric correction in presence of sun glint: application to MERIS, *Opt. Express*, 19, 9783–9800, 2011.

Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., Piera, J., and Bracher, A.: Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance, *Biogeosciences*, 8, 3609–3629, doi:10.5194/bg-8-3609-2011, 2011.

Taylor, B. B., Taylor, M., Dinter, T., and Bracher, A.: Estimation of relative phycoerythrin concentrations from hyperspectral underwater radiance measurements – a statistical approach, *J. Geophys. Res.-Ocean.*, 118, 2948–2960, doi:10.1002/jgrc.20201, 2013.

Uitz, J., Claustre, H., Morel, A., and Hooker, S.: Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll, *J. Geophys. Res.*, 111, C08005, doi:10.1029/2005JC003207, 2006.

Wu, C. F.: Jackknife, bootstrap and other resampling methods in regression analysis, *Ann. Statistic.*, 14, 1261–1295, 1986.

Zapata, M., Rodriguez, F., and Garido, J. L.: Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases, *Marine Ecol. Progr. Ser.*, 195, 29–45, 2000.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

**Table 1.** Percent of total variance explained (expl. variation) by the significant EOFs derived from field RRS spectra in hyper- (hyper\_RRS) and multi- (band\_RRS) spectral resolution and from satellite\_ $\rho_{\text{WN}}$  (from MERIS Polymer) using the  $1 \times 1$  pixel collocation criterion.

% expl. variation	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7	EOF-8	EOF-9
hyper_RRS	95.0	4.1	0.7	0.1	<0.1	<0.1	<0.1	<0.1	<0.1
band_RRS	94.5	5.0	0.4	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
satellite_ $\rho_{\text{WN}}$	95.9	3.8	0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


**Table 2.** Statistics of linear models using EOF modes based on (a) field RRS data in hyper (hyper; normally 350–700 nm – if \* then 380–700 nm) and multi (band)-spectral resolution and (b) the satellite  $\rho_{wN}$  (from MERIS Polymer) using the  $1 \times 1$  pixel collocation criterion data set. Cross validation results are presented with 500 permutations for data splitting into 80 % of the data used for training and 20 % for validation. Only well retrieved pigment prediction results, with correlations being highly significant at  $p < 0.0001$ , are given. Abbreviations of pigments are explained in Sect. 2.3.1. Pigments listed in the upper part of each table show high quality results using the entire data set. In the lower part of each table (listed under “> 0 mg m<sup>-3</sup>”) models are based only on the data set of collocated RRS samples where the respective pigment reached concentrations above 0 mg m<sup>-3</sup>. Bold: here band-model performs better than hyper-model. Italic signifies only medium quality as specified in the text.

(a)									
band/hyper	N	MDPDcv		MPDcv		RMSEcv		$R^2_{cv}$	
		band	hyper	band	hyper	band	hyper	band	hyper
TChl <i>a</i> *	53	<b>28</b>	32	<b>42</b>	43	<i>0.68</i>	0.41	0.72	0.77
PSC*	53	<b>32</b>	34	<b>51</b>	53	0.33	0.32	0.75	0.75
PPC	53	27	27	48	45	0.21	0.18	0.52	0.52
MVChl <i>a</i> *	53	<b>31</b>	34	45	44	0.62	0.35	0.77	0.79
Chl <i>c</i> <sub>1/2</sub> *	53	<b>34</b>	36	<b>48</b>	49	0.08	0.08	0.79	0.80
But	53	44	43	<i>95</i>	<i>81</i>	0.05	0.04	0.50	0.52
Hex*	53	37	36	57	53	0.19	0.13	0.64	0.70
Zea	53	31	28	45	42	0.05	0.05	<i>0.28</i>	<i>0.31</i>
Caroten	53	34	34	55	50	0.17	0.06	0.55	0.61
PE*	53	67	65	<i>156</i>	<i>139</i>	<i>0.23</i>	0.14	0.65	0.69
> 0 mg m <sup>-3</sup> :									
DVChl <i>a</i>	49	26	24	45	39	0.06	0.06	0.59	0.65
Chl <i>c</i> <sub>3</sub> *	52	38	37	61	58	0.08	0.07	0.71	0.72
Fuco*	52	42	39	<i>87</i>	<i>73</i>	<i>0.61</i>	0.18	0.65	0.65
Diadino	52	33	33	59	53	<i>0.25</i>	0.08	0.57	0.65
(b) satellite $\rho_{wN}$									
	N	MDPDcv	MPDcv	RMSEcv		$R^2_{cv}$			
TChl <i>a</i>	135	31	47	0.39	0.67				
MVChl <i>a</i>	135	39	55	0.35	0.71				
PSC	135	42	59	0.25	0.73				
PPC	135	25	36	0.09	0.54				
> 0 mg m <sup>-3</sup> :									
DVChl <i>a</i>	120	41	60	0.03	<i>0.25</i>				
Chl <i>c</i> <sub>1/2</sub>	127	41	89	0.06	0.64				
X19_But	126	39	55	0.03	0.74				
X19_Hex	128	40	55	0.07	0.71				
Fuco	128	40	65	0.43	0.71				
Zea	134	27	45	0.05	<i>0.42</i>				
Diad	129	44	94	0.13	0.56				

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 3.**  $\Delta$ AIC for the robust pigment predictions of the pigment groups TChl *a*, PSC, PPC and the pigments MVChl *a*, Zea, and DVChl *a* by the EOF models based on field RRS in (a) hyper- (hyper\_RRS) and (b) multi- (band\_RRS) spectral resolution and c) the satellite- $\rho_{wN}$  (from MERIS Polymer) using the  $1 \times 1$  pixel collocation criterion. The pigments listed under “no  $0 \text{ mg m}^{-3}$ ” were predicted using a reduced data set where the respective pigment reached concentrations above  $0 \text{ mg m}^{-3}$ . Bold highlights the EOF mode with the highest  $\Delta$ AIC.

(a) hyper_RRS	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7	EOF-8	EOF-9
TChl <i>a</i>		<b>80</b>	16	16					
PSC		<b>78</b>	9	8	4				
PPC		<b>36</b>	34		1		2	1	
MVChl <i>a</i>		<b>88</b>	16	11					
Zea	5	6	<b>21</b>		10	7	1	5	
no $0 \text{ mg m}^{-3}$ : DVChl <i>a</i>	7	10	21	<b>21</b>	1	20	9	4	7
(b) band_RRS	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7		
TChl <i>a</i>		<b>86</b>	10	15	3	1			
PSC		<b>80</b>	5	5	6				
PPC		<b>42</b>	21	9		1			
MVChl <i>a</i>		<b>90</b>	10	9	1				
Zea	2	5	6	<b>10</b>	4	7	0.3		
no $0 \text{ mg m}^{-3}$ : DVChl <i>a</i>	12	17		20	<b>23</b>				
(c) satellite- $\rho_{wN}$	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7		
TChl <i>a</i>	6	<b>144</b>			0.3	0.4			
PSC	6	<b>158</b>				14			
PPC	4	<b>47</b>	1	1			6		
MVChl <i>a</i>	4	<b>148</b>				4			
no $0 \text{ mg m}^{-3}$ : Zea	13	<b>31</b>	3	1		14	9		
DVChl <i>a</i>	12		<b>24</b>	<b>33</b>	4				

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

**Table 4.** TChl *a* validation statistics for MERIS Polymer TChl *a* (left panel) and TChl *a* obtained from full fit linear model on satellite\_ρ<sub>wN</sub> (from MERIS Polymer data; EOF full-fit model; right panel) with different collocation criteria (either 1 × 1 or the mean of 3 × 3 or 5 × 5 pixel values) for the MERIS Polymer data compared to the in situ (from HPLC) value. Bold highlights that TChl *a* validation results for MERIS Polymer are better than EOF model predictions.

MERIS collocation	<i>N</i>	Polymer				EOF full-fit model			
		<i>R</i> <sup>2</sup>	RMSE	PB	MPD	<i>R</i> <sup>2</sup>	RMSE	PB	MPD
1 × 1 pixel	135	0.74	<b>0.21</b>	<b>6</b>	39	0.73	0.32	10	37
3 × 3 pixel	150	0.75	<b>0.20</b>	10	40	0.74	0.33	10	36
5 × 5 pixel	155	0.74	<b>0.20</b>	11	40	0.74	0.32	10	36

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



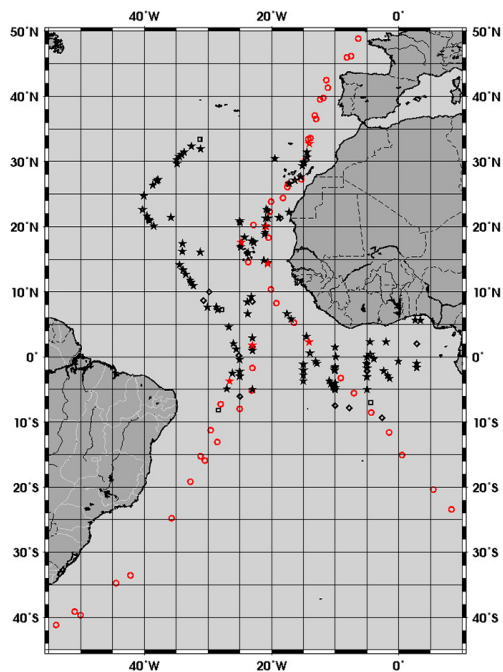
Back

Close

Full Screen / Esc

Printer-friendly Version

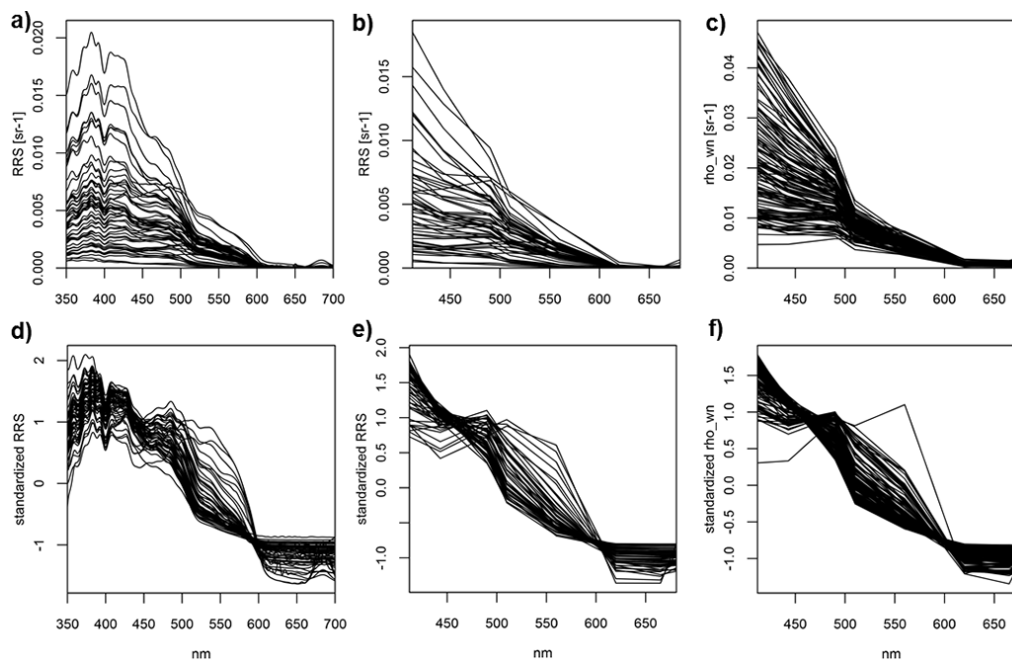
Interactive Discussion



**Figure 1.** Position of pigment samples used in this study. Red: field data set, black: samples which are only collocated to satellite-based but not to field reflectance data, circles: samples which are only collocated to field but not to satellite-based reflectance data. Stars, diamonds and squares show collocations to MERIS Polymer data based on the  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$  pixel criteria, respectively.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.



**Figure 2.** (a)–(c) Original (in  $\text{sr}^{-1}$ ) and (d)–(f) standardized (subtracted mean) reflectance spectra of hyper\_RRS (a) and (d), band\_RRS (b) and (e) and satellite\_ρ<sub>wN</sub> (from MERIS Polymer) data within the  $1 \times 1$  pixel collocation box (c) and (f).

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)




## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

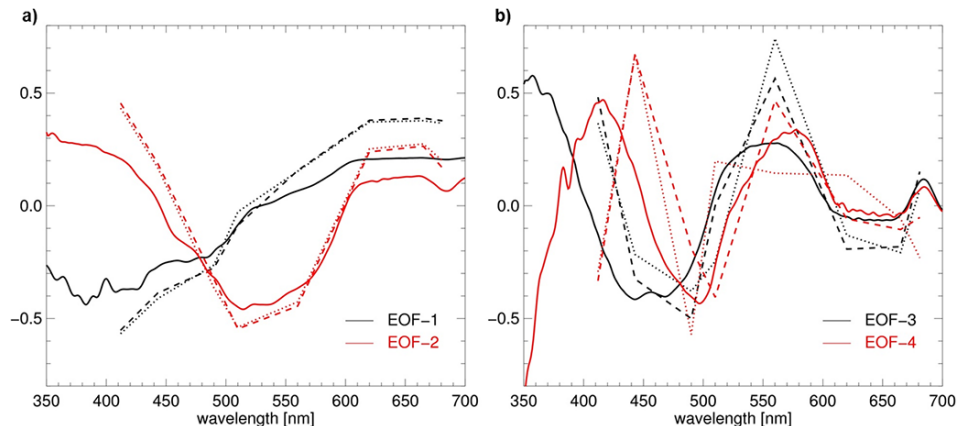
Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Figure 3.** First four EOF modes (EOF-1 to EOF-4) derived from field RRS data set in hyper-spectral resolution (hyper\_RRS, solid lines) and in multispectral band resolution (band\_RRS, dashed lines) and from using satellite  $\rho_{wN}$  (from MERIS Polymer) data within the  $1 \times 1$  pixel collocation box.

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



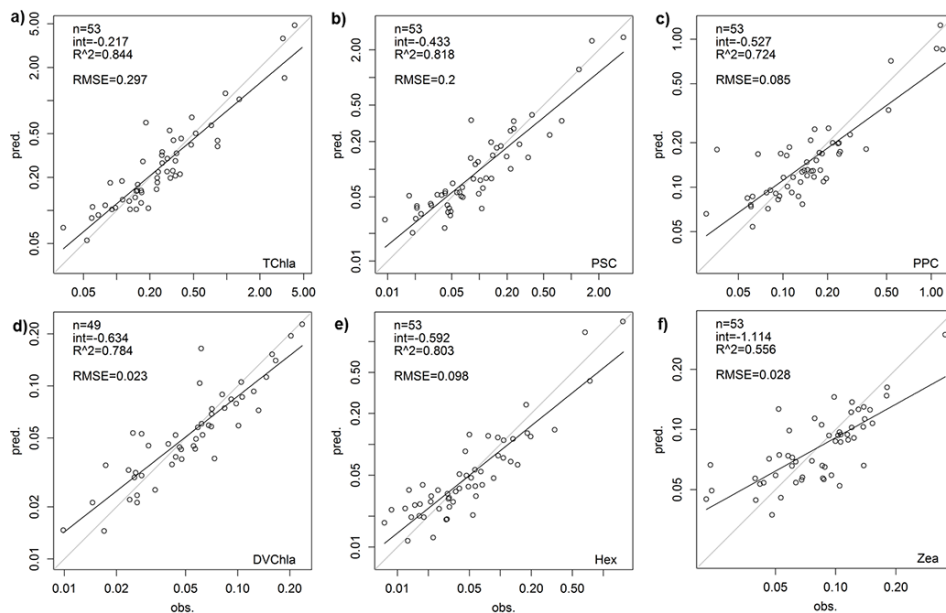
Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Figure 4.** Examples of pigment group's **(a)** TChl *a*, **(b)** PSC, **(c)** PPC) and specific pigment's **(d)** DVChl *a*, **(e)** Hex, **(f)** Zea) concentrations measured by HPLC (obs.) versus predicted (pred.) from a pigment or pigment group specific linear model based on EOF modes using field RRS data in hyper-spectral resolution (hyper\_RRS). For DVChl *a* the model data set was reduced by excluding collocated samples where DVChl *a* had concentrations of  $0 \text{ mg m}^{-3}$ .

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



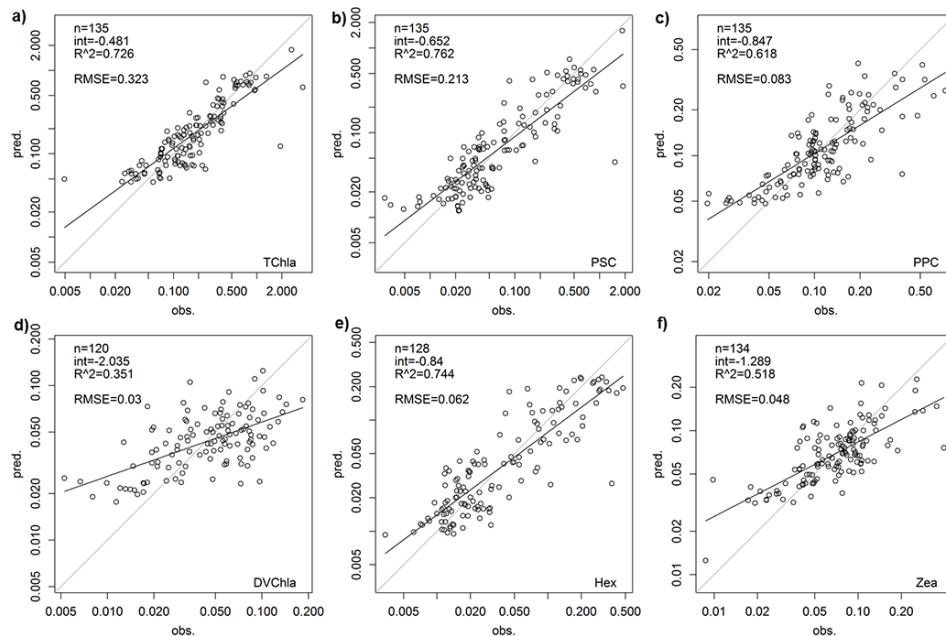
Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Figure 5.** Examples of pigment group's **(a)** TChl *a*, **(b)** PSC, **(c)** PPC) and specific pigment's **(d)** DVChl *a*, **(e)** Hex, **(f)** Zea) concentrations measured by HPLC (obs.) versus predicted (pred.) from a pigment or pigment group specific linear model based on EOF modes using satellite  $\rho_{wN}$  (from MERIS Polymer) data within the  $1 \times 1$  pixel collocation box. For DVChl *a*, Hex and Zea the model data set was reduced by excluding collocated samples where DVChl *a* had concentrations of  $0 \text{ mg m}^{-3}$ .

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



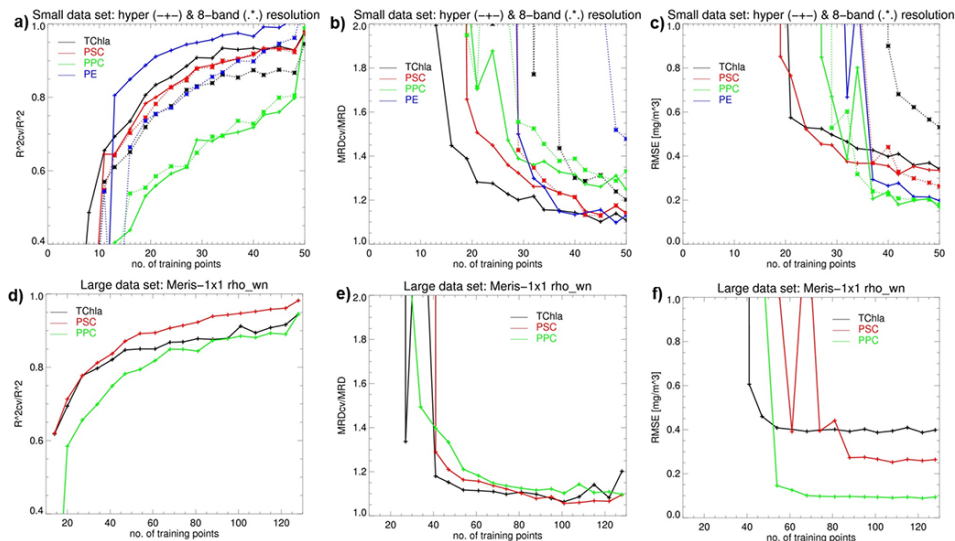
Back

Close

Full Screen / Esc

Printer-friendly Version

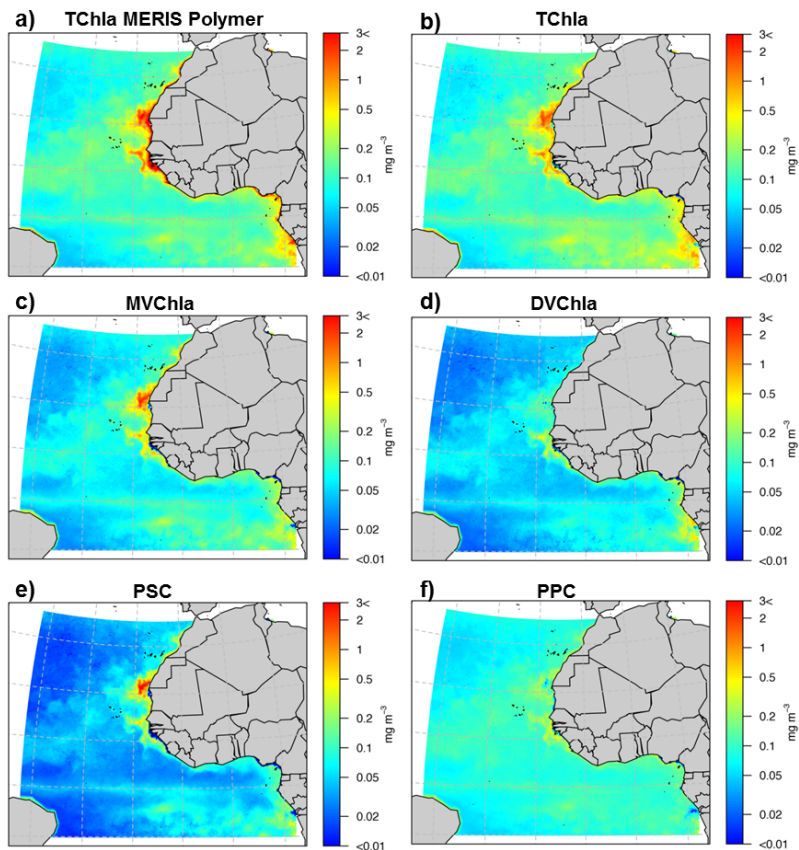
Interactive Discussion



**Figure 6.** The ratio of  $R^2_{cv}$  to  $R^2$  (**a**) and (**d**), of MRD<sub>cv</sub> to MRD (MRD<sub>cv</sub> / MRD, (**b**) and (**e**) and RMSE<sub>cv</sub> (**c**) and (**f**) as a function of number of training points (tp) for the linear models. Shown are results for specific models for TChl *a*, PSC, PPC and PE using reflectance data (**a**) to (**c**): from the field in hyperspectral (hyper\_RRS, solid lines) and multispectral band (band\_RRS, dotted lines) resolution and (**d**) to (**f**): from satellite MERIS Polymer within the  $1 \times 1$  pixel collocation box (satellite\_ $\rho_{wN}$ ). The number of total samples points were  $n = 53$  for hyper\_RRS and band\_RRS, and  $n = 135$  for satellite\_ $\rho_{wN}$ . Cross validation were done based on 500 permutations using tp for training and as number of validation points (vp):  $vp = n - tp$ .

## Predicting different pigments from remote sensing reflectances

A. Bracher et al.



**Figure 7.** Monthly mean concentrations (gridded in  $0.25^\circ$  lat/lon boxes) for Nov 2008 of (a) TChl *a* of the MERIS Polymer algorithm (TChl *a* MERIS Polymer) and predicted (b) TChl *a*, (c) MVChl *a*, (d) DVChl *a*, (e) PSC, and (f) PPC by the LM based on the full fit of satellite  $\rho_{+wN}$  data within the  $1 \times 1$  pixel collocation box and the EOFs of this month's MERIS Polymer  $\rho_{+wN}$  data.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion