Global Retrieval Algorithms for Phytoplankton Functional Types (PFTs): toward the Applications to OLCI and GlobColour Merged Products

Hongyan Xi[1], Svetlana N. Losa[1], Antoine Mangin[3], Mariana A. Soppa[1], Philippe Garnesson[3], Julien Demaria[3], Odile Hembise Fanton d'Andon[3], Astrid Bracher[1,2]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[3]ACRI-ST, 06904 Sophia Antipolis Cedex, France

Abstract: This study focuses on PFT retrieval algorithms that are then applied to Sentinel-3A (S3) OLCI data and merged ocean colour (OC) products from CMEMS GlobColour archive. The main retrieved PFTs include diatoms, haptophytes, and prokaryotic phytoplankton (cyanobacteria). Previously investigated retrieval methods, empirical orthogonal functions (EOF) for pigment concentrations estimation (Bracher et al. 2015) and generalized OC inversion model for inherent optical properties (GIOP) (Werdell et al. 2013, 2014) for PFT discrimination, are tested and adapted potentially with full use of our current available *in situ* measurements from various campaigns worldwide, in which we have a number of collocated remote sensing reflectance spectra (R_{rs}) and PFT data based on HPLC pigments in addition to other bio-optical measurements. Algorithms are tested and compared by both taking hyperspectral and multispectral *in situ* R_{rs} spectra as input data, and the multispectral based approach is applied later on to the above-mentioned satellite products. Performances of both EOF- and GIOP-based approaches are assessed statistically and cross-validated, with results showing that both could well predict chlorophyll-a concentrations for diatoms and haptophytes but less accurate for prokaryotes. In a next step these algorithms are adapted to satellite OC data collocated to an even much larger in-situ PFT database derived from HPLC phytoplankton pigments. This is to eventually develop the global satellite PFT products for long-term observation, updated timely with more available OLCI data in the future, and intercompared to the results with other existing PFT products (e.g. PhytoDOAS, OC-PFT, SynSenPFT, PHYSAT).

1. Background

Over the past decades, satellite ocean colour remote sensing has been widely utilized as a method for estimating chlorophyll a concentration, which is often used as an indicator of phytoplankton biomass. Beyond that, extracting information on phytoplankton community structure (e.g., phytoplankton functional types (PFTs), size classes (PSCs) and composition) has become a research topic of priority, as they are playing an important role in understanding the marine food web and provide aids in the modelling associated with climate change impacts on biogeochemical and ecological cycling of oceans (IOCCG, 2014; Bracher et al., 2017; Correa-Ramirez et al., 2018). Satellite OC data enables observation of PFTs over large areas or even at global scale. With the successful launch of Sentinel 3A in 2016 and 3B in 2018, both OLCI sensors are providing high quality OC data which allow us to continue developing methods and their application to satellite data for identifying and estimating PFTs with respect to global observation. There is a clear need to implement a sound PFT retrieval algorithm to the OLCI data, as well as to previous and current OC products, such as GlobColour merged OC data.

In the present study, we investigate two algorithms for their capability in PFT retrievals, namely the adapted generalized IOP (AGIOP) and the EOF-based algorithm, with the use of extensive *in situ* measurements, matchups between *in situ* and satellite data, and satellite OC products. The retrieved PFTs (diatoms, haptophytes, prokaryotic phytoplankton (cyanobacteria)) based on *in situ* data sets are compared with the results of diagnostic pigment analysis (DPA) from HPLC measurements. The two algorithms are also preliminarily applied to the GlobColour merged OC products and OLCI data.

2. Data and Method

2.1. *In situ* datasets

We have collected our current available *in situ* measurements from various campaigns covering different regions worldwide, including the Atlantic (ANT24-1 (7/2007), ANT24-4 (4-5/2008), ANT25-1 (11/2008), ANT26-4 ((4-5/2010) and a very recent one PS113 (5-6/2018); ~77 profiles), North Atlantic MSM9-3 (summer 2008), Arctic (PS76 (summer 2010), ARK26-3 (8-9/2011), PS93.2, PS99, and PS107 (summers 2015-2017), ~ 81 profiles, Pacific regions (SO218 South China/Zulu Sea (11/2011), SO243 at East tropical Pacific (10/2015); 12 profiles), the Southern Ocean (ARK28-3 (1-3/2012); 17 profiles), and the North Sea (Heincke462 (4-5 2016), ~19 profiles), in which we have a number of remote sensing reflectance spectra (R_{rs}) and PFT data based on HPLC pigments in addition to other bio-optical measurements (mainly the absorption spectra by the total particles, non-algal particles and phytoplankton). From this collection we have 212 R_{rs} profiles with 208 HPLC pigment and absorption measurements collocated data sets.

A large database of the quality controlled near surface (first 12 m) HPLC phytoplankton pigments built for the SynSenPFT Project (SynSenPFT, 2017) is also used with the collocated R_{rs} spectra from GlobColour merged OC products for the EOF model training. The HPLC pigment database includes 15,176 sets of phytoplankton pigments spanning 15 years from 1988 to 2012 covering the global ocean, collected from SEABASS, MAREDAT, LTER, BATS, AESOP-CSIRO, LOV and also from own data at PANGAEA (see Table 1 in Losa et al., 2017).

2.2. Satellite data and matchups

The R_{rs} spectra at multispectral bands collocated with the HPLC pigments from the large database mentioned in Section 2.1 were extracted from the merged ocean color products (including SeaWiFS, MODIS, MERIS, VIIRS) from 1997 to 2012, with the help of ACRI-ST. This data set with collocated HPLC pigment data is used for the EOF model training for global PFT retrievals. The extracted R_{rs} matchups include single pixel, 3x3, and 5x5 pixels with the mean and the standard deviation for each matchup for the 3x3 and 5x5 cases. However, not always the same wavebands for R_{rs} data are available due to that different sensors with different spectral coverage have measured at different period (in addition also the exclusion of data with bad quality). Table 1 lists the numbers of matchups with different band combinations (from 5 bands to 12 bands) for R_{rs} matchups with single pixel and 3x3 (5x5) pixels, respectively.

Satellite products were also obtained from the GlobColour data archive (http://www.globcolour.info/). For the first application of AGIOP and EOF-based algorithms to OLCI and GlobColour merged products, scenes of S3A OLCI L3 8-day averaged global R_{rs} product in 2017, and monthly merged R_{rs} products in 2011 were acquired with resolutions at 25 km and 100 km, respectively.

2.3. Theoretical basis of retrieval algorithms for PFTs

2.3.1. Adapted generalized ocean colour inversion model (AGIOP)

GIOP was initially developed under the framework of relating R_{rs} to the inherent optical properties (IOPs). GIOP integrates all the existing semi-analytical algorithms (SAA) and allows for the construction of different SAAs at runtime by selecting different parameterizations. Figure 2 gives the flowchart of the GIOP and the detailed description can be found in Werdell et al. (2013). Further investigations by Werdell et al. (2014) and Wolanin et al. (2016) on the capability of GIOP in discriminating phytoplankton groups implied that it is potentially possible to quantify the biomass of phytoplankton types by characterizing the absorption by phytoplankton with consideration of different phytoplankton populations. It is assumed that the phytoplankton absorption component is a linear sum of subcomponents with unique spectral dependencies. Therefore, in the adapted GIOP (AGIOP) we decomposed the phytoplankton absorption into three PFTs, that are diatoms, haptophytes, and cyanobacteria. Specific absorption of the three PFTs used in the AGIOP were obtained from natural water samples where one of the three PFTs was dominating. Using R_{rs} at different wavebands and the spectral shapes of the inherent optical properties as eigenvectors as input, eigenvalues for absorption (which are the chlorophyll concentrations of PFTs for phytoplankton absorption and absorption of coloured detrital matter at 440nm) and particle backscattering at 440 ($b_{bp}(440)$) can be derived via linear or nonlinear least squares inversions. Note that different inversion methods may lead to differences in the retrievals.

2.3.2. Empirical Orthogonal Functions (EOF) based algorithm

Empirical Orthogonal Function (EOF) analysis which is based on principal component analysis, has been previously used for assessing variance of structures in spectral R_{rs} or water leaving radiance data (e.g., Taylor et al. 2013, Craig et al. 2012, Lubac and Loisel 2007). The spectral data are subject to EOF analysis, in order to reduce the high dimensionality of the data and derive the dominant signals ("modes") that best describe variance within the data set (Taylor et al. 2013). In addition to dimension reduction of spectral data, the use of EOF modes in statistical model building also avoids problems associated with multicollinearity amongst the original predictor variables. In a recent study by Bracher et al. (2015) the reflectance data were used to derive the optical signature of different pigments by an automatic and generic technique. The EOF analysis is applied to R_{rs} data obtained in-situ and by satellite sensors in the Atlantic Ocean. The dominant EOF loadings were subsequently assessed as predictors in a multiple linear regression for the concentration of phytoplankton pigments. A permuted cross-validation routine was then performed to evaluate the prediction error of each model, to estimate the critical sample sizes necessary for reliable prediction.

Given that EOFs derived from both hyperspectral underwater radiometric measurements and multispectral reflectance data from field or satellite (MERIS Polymer) enable reliable predictions of the concentration of nine different pigments/pigment groups (Bracher et al. 2015), it is worthwhile to investigate the capability of EOF analysis on reflectance data in predicting the concentrations of PFTs. A recent study by Correa-ramirez et al. (2018) showed that the EOF method could improve the remote sensing identification of PFTs (only in their dominance but not quantitatively), which suggests the possibility of the EOF in retrieving the PFT Chl-a concentrations. Following the steps used in model building and predictions in Bracher et al (2015), we applied the EOF analysis to collocated R_{rs} data from GlobColour merged products with *in situ* HPLC (globally) from 1997–2012. With the use of the collocated *in situ* pigment data we established the prediction functions between the EOF modes and the HPLC based PFTs, which will be further implemented to the satellite images (merged OC products and OLCI) to enable the global retrieval of PFTs.

3. Results

3.1. AGIOP with three PFTs included

The *in situ* data sets of R_{rs} was used as input for the AGIOP with three PFTs included (as described in section 2.3.1). Based on the improved diagnostic pigment analysis (DPA) (Soppa et al., 2014, 2017; Losa et al., 2017), the *in situ* HPLC pigments were used to estimate the chlorophyll concentrations of the three PFTs, which were taken as observations and compared with the retrievals from the AGIOP to validate the performance of the algorithm.

Both nonlinear least square (NLSQ) inversion and non-negative linear matrix inversion (NonnegLMI) were used in the AGIOP, in order to have a comparison between the two inversion methods. Besides, the following aspects were also considered:

1. Hyperspectral R_{rs} data set with different band ranges (400 – 700 nm, 450 – 650 nm, 400 – 650 nm, and 400 – 600 nm) as input
2. Excluding or including data points in optically complex case 2 waters (points of case 1 and case 2 waters were separated using the criteria proposed by Lee & Hu (2006))
3. *In situ* R_{rs} spectra with reduced spectral resolution (specially at OLCI bands according to the OLCI spectral response functions (SRF))

Tests on band ranges showed that there were only slight differences when using R_{rs} data set at different band ranges. However, the range 400 – 600 nm outperformed the others. Excluding the points in case 2 waters improved even more the statistical results comparing the AGIOP retrieved the PFTs and the PFTs from the in-situ data set (results not shown). The retrievals from AGIOP using multispectral R_{rs} with OLCI bands were not deteriorated significantly and were still comparable to the results from hyperspectral R_{rs}. As an example, Figure 1 shows the comparison between the AGIOP retrieved concentrations of total Chl-a and the three PFTs and the HPLC pigment based PFTs. Statistical results for some important cases were listed in Table 2. Overall, the AGIOP performs the best for the total Chl-a, fairly for diatoms and haptophytes, but the worst for cyanobacteria due to the relatively low concentrations. All the retrieved concentrations were generally overestimated compared to the HPLC based PFTs.

3.2. EOF-based Algorithm

According to the band combinations of the R_{rs} from the merged OC products listed in Table 1, we chose the ones marked as blue (8 bands, 9 bands and 11 bands) with corresponding *in situ* pigments as input for EOF model building. The total chlorophyll concentration and the three PFTs were predicted based on the leading EOF modes and linear relationship established by the multilinear regression. R_{rs} at single pixel, 3x3 pixels, and averaged R_{rs} at 3x3 pixels were all taken as input for comparison between the results from different band numbers, pixels and data points. Statistics showed that EOF predicted PFTs displayed slight differences though there were various settings of the input R_{rs} data sets regarding the band number and matchup pixels. Among all the settings, R_{rs} at 9 bands with 1x1 pixel performed the best, slightly better than the R_{rs} at the same 9 bands with 3x3 pixels (all pixels taken into account). Figure 2 shows the regression between the predicted and observed PFTs for the mentioned two R_{rs} data sets. With respect to the

PFTs retrievals, diatoms and haptophytes showed comparable and relatively good predictions but prokaryotes (cyanobacteria) were of the lowest quality. This is mainly due to the low concentrations and the narrow range of the variation compared to other PFTs, leading poor accuracy in the prediction model.

3.3. First PFT results of AGIOP and EOF algorithms applied to OLCI and merged OC data

3.3.1 AGIOP applied to OLCI products

We chose 8-day averaged global R_{rs} product in January, April, June and October 2017 as the first examples to test the GIOP model for PFT retrievals. For the first test and run only the ones with 25 km (8 hours runtime on PC) and 100 km (30 minutes on PC) were used. Only case-1 waters were chosen based on the criteria by Lee and Hu (2006), and all others (case 2 waters, area with unavailable R_{rs}, etc.) were flagged. The NLSQ and Nonneg-LMI inversion methods in the AGIOP model with 3 PFTs included were both tested. Disappointingly, the NLSQ showed totally unreasonable distribution of the 3 PFTs, with diatom extremely high, negative retrievals for haptophytes, and overestimated cyanobacteria. The main reason causing this might be the inappropriate use of nonlinear minimization method in which proper constraints and bounds could not be applied for the retrieved parameters. Nevertheless, the Nonneg-LMI inversion worked fairly well even though a large amount of the pixels were failed to be retrieved for diatom and haptophytes (Figure 3), as most of the retrievals for the two PFTs reached the Nonneg-LMI lower bound '0'. From the distribution it was found that retrieval failures often occurred in oligotrophic regions where the overall phytoplankton biomass was very low, indicating that the algorithm is not efficient in working with small values, and corresponding adjustment and more tests in minimization methods in the AGIOP are still required.

In conclusion, while the whole procedure of AGIOP applied to OLCI data is working further investigations on the model settings need to be carried out. The factors influencing the most are: types and numbers of the PFTs involved in the model, selection of the PFT absorption spectra and their corresponding weightings (e.g., introducing wavelength dependent uncertainties in R_{rs} and a_{ph_PFT}). All these factors limit the AGIOP model which was developed in the first place on exploitation of the *in situ* measurements.

3.3.2. EOF based algorithm applied to GlobColour merged OC products

As Sentinel 3A and 3B were recently launched and matchups between OLCI R_{rs} and *in situ* measurements have not been fully analysed or collected yet, only the matchups between the satellite R_{rs} from merged sensors and *in situ* HPLC pigment data from 1997 to 2012 were available for the EOF training, which therefore only allows the application of the trained EOF modes to the merged OC products in the period of 1997 – 2012. Given that the EOF modes trained by the R_{rs} at 9 bands works the best for PFT prediction compared to other band combinations (band information in Table 1), we applied the EOF modes based on the matchups with 9 R_{rs} bands to the merged OC R_{rs} product at the same bands in the year of 2011. Figure 4 shows the global retrievals of the 3 PFTs in different seasons. Both diatom and haptophytes displayed similar distribution patterns with the results from AGIOP but with much more valid retrievals leading to a better coverage, despite that the magnitudes of the EOF retrieved concentrations were generally lower than that of the AGIOP. Cyanobacteria retrieved by the two algorithms showed apposed patterns in most of the oceans except for the waters above 40°N in some seasons. Regarding cyanobacteria, it is already known from section 3.2 the EOF training using satellite matchups, that the prediction for cyanobacteria was the worst (low R^2) among all the PFTs and predicted concentrations were flattened showing overestimation in smaller values and underestimation in higher values (Figure 2). So far, an accurate cyanobacteria or its marker pigment retrieval has always been a challenge (e.g., Bracher et al., 2015; Losa et al., 2017). Validation of the PFT product are yet to be carried out and the algorithm itself needs to be further improved by considering more mathematical inputs in the EOF transformation such as using nonlinear multi-regression for PFT prediction or scaling the data sets to improve the retrievals for smaller values.

4. Conclusions and outlook

Two algorithms, AGIOP and EOF based PFT retrieval algorithm, were both investigated and tested with different settings and scenarios, and the retrievals were compared with the PFTs derived using the improved DPA algorithm. Both algorithms can well retrieve diatoms and haptophytes but perform less good for cyanobacteria mainly due to their general low concentration resulting weak signal in the reflectance spectra.

Nevertheless, regarding the algorithm development, the following aspects are still worth to be further investigated for AGIOP:

- Specific absorption spectra of PFTs used in the model, parametrizations accounting for the package effect, and spectral regions of no correlation of the absorption spectra by different PFTs.
- Uncertainties introduced into the R_{rs} data set with consideration of specific bands, information obtained from measurements and knowledge of PFT contained in each band.
- As GIOP assumes the absorption by phytoplankton is a linear sum of different phytoplankton components, it is crucial to determine the number of PFTs that are to be included.
- Other minimization methods.

With respect to the EOF based PFT algorithm, it performed better than the AGIOP for both, the *in situ* datasets and the merged OC products. However, the EOF modes applied to the satellite OC images for the global retrieval showed that concentration of cyanobacteria was still overestimated especially for oligotrophic waters. Due to the current lack of a large enough number of matchups between the OLCI R_{rs} and *in situ* pigments, the EOF training for OLCI data so far cannot be carried out. With our recent PS113 trans-Atlantic cruise (see Bracher et al., this conference) and previous PS cruises in Fram Strait since 2016 (see Liu et al., this conference), it is possible to extract matchups of OLCI for the EOF application to S3 OLCI data. Overall, the following aspects are worth to be considered or tested:

- Collect and use matchups between the OLCI R_{rs} and *in situ* pigment data to establish the EOF modes and prediction functions for OLCI bands, and apply the EOF modes to S3A/B OLCI products, and compare the results with the results from AGIOP–OLCI retrieved PFTs.
- Separate two cyanobacteria groups: EOF to DVChla (directly retrieved Prochlorococcus) and PE-data (marker for phycoerythrine containing cyanobacteria, Taylor et al. 2013).
- Try EOF models with different principle components included, this might reduce the flattening of the regression between the predicted and observed data.

Acknowledgements

This work is supported by a collaborative project between ACRI-ST and AWI, OLCI-PFT. We express our gratitude to Jeremy Werdell for the original GIOP algorithm, to Marc Taylor for the EOF-based algorithm. We thank Sonja Wiegmann for her dedication in measuring and processing the absorption spectra and HPLC data for most of our past cruises, and all the previous and current PhytoOptics team members who participated in the past cruises for data collection and analysis. We are thankful to all the scientists and crew involved in the global HPLC data collection and analyses for providing their pigment data. We also thank the SeaWiFS, MODIS, MERIS, VIIRS, and OLCI data, and specially the GlobColour team for providing the OLCI and merged OC L3 products.

References

Bracher, A., Bouman, H., Brewin, R. J., Bricaud, A., Brotas, V., Ciotti, A. M., et al. (2017). Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. Front. Mar. Sci. 4:55. doi: 10.3389/fmars.2017.0005

Bracher, A., Taylor, B.B., Taylor, M., Dinter, T., Röttgers, R., Steinmetz, F. (2015). Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of concentrations of phytoplankton pigments. *Ocean Science* 11: 139-158.

Bracher, A., Xi, H., Wiegmann, S., Dinter, T. (2018). Phytoplankton diversity and its optical properties along different Longhurst provinces in the Atlantic (not Indian) Ocean. Ocean Optics Conference OOXXIV, Dubrovnic, Croatia, 7-12 October, 2018.

Correa-Ramirez, M., Morales, C.E., Letelier, R., Anabalon, V., and Hormazabal, S. (2018). Improving the remote sensing of phytoplankton functional types (PFT) using empirical orthogonal functions: a case study in a coastal upwelling region. *Remote Sensing*, 10, 498.

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.

Franz, B.A. & Werdell, P.J. (2010). A Generalized Framework for Modeling of Inherent Optical Properties in Remote Sensing Applications, Proc. Ocean Optics 2010, Anchorage, Alaska, USA, 27 September – 1 October 2010.

IOCCG (2014). "Phytoplankton functional types from space," in Reports of the International Ocean-Color Coordinating Group, No. 15, ed S. Sathyendranath (Dartmouth: IOCCG).

Lee, Z., & Hu, C. (2006). Global distribution of Case-1 waters: An analysis from SeaWiFS measurements. Remote Sensing of Environment, 101, 270–276. doi:10.1016/j.rse.2005.11.008

Liu, Y., Boss, E., Chase, A., Pan, Y., Xi, H., Roettgers, R., Bracher, A. (2018). Retrieval of phytoplankton pigments and functional types from underway spectrophotometry in the Fram Strait. Ocean Optics Conference OOXXIV, Dubrovnic, Croatia, 7-12 October, 2018.

Losa S. N., Soppa M. A., Dinter T., Wolanin A., Brewin R. J. W., Bricaud A., Oelker J., Peeken I., Gentili B., Rozanov. V. V., Bracher A. (2017). Synergistic exploitation of hyper- and multispectral precursor Sentinel measurements to determine Phytoplankton Functional Types at best spatial and temporal resolution (SynSenPFT). Front. Mar. Sci. 4: 203; doi: 10.3389/fmars.2017.00203

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

SynSenPFT, https://www.awi.de/en/science/climate-sciences/physical-oceanography/main-research-focus/ocean-optics/projects/synsenpft.html, accessed in 2018.

Soppa et al. 2017 PANGAEA data set (see reference in Losa et al, 2017).

Soppa M. A., Hirata T., Brenner S., Dinter T., Peeken I., Wiegmann S., & Bracher A. (2014). Global retiieval of diatoms abundance based on phytoplankton pigments and satellite. Remote Sensing 6 (10), 10089-10106.

Taylor, B. B., Taylor, M., Dinter, T., and Bracher, A. (2013). 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.

Werdell, P. J., Franz, B. A., Bailey, S. W., Feldman, G. C., Boss, E., Brando, V. E., Dowell, M., Hirata, T., Lavender, S. J., Lee, Z. P., Loisel, H., Maritorena, S., Melin, F., Moore, T. S., Smyth, T. J., Antoine, D., Devred, E., d'Andon, O. H. F., and Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties, Appl. Opt., 52, 2019–2037.

Werdell, P. J., Roesler, C. S., and Goes, J. I. (2014). Discrimination of phytoplankton functional groups using an ocean reflectance inversion model. *Appl. Opt.* 53, 4833–4849. doi: 10.1364/AO.53.004833

Wolanin A., Soppa M. A., Bracher A., (2016) Investigation of spectral band requirements for improving retrievals of phytoplankton functional types. *Remote Sensing* 8: 871; doi:10.3390/rs8100871

Table 1. Numbers of available matchups between *in situ* HPLC pigment and GlobColour merged R_{rs} (1x1 pixel, in total 1,607 spectra; 3x3 pixels, in total 2,063 spectra; 5x5 pixels, in total 2360 spectra) with different band combinations from the merged ocean colour products. Blue highlights the matchups used in the EOF based algorithm (SeaW = SeaWiFS, MO = MODIS, ME = MERIS, V = VIIRS).

Sensors	**Number of Matchups**		**Available Wavebands (nm)**												**Number of bands**
	1x1	3x3 5x5	**412**	**443**	**490**	**510**	**531**	**547**	**551**	**555**	**560**	**620**	**670**	**678**	
SeaW	1223	609	X	X	X	X				X			X		**6**
SeaW+MO+ME	**408**	**266**	X	X	X	X	X	X		X			X	X	**9**
SeaW+ME	**502**	**129**	X	X	X	X				X	X	X	X		**8**
SeaW+MO+ME	**212**	**64**	X	X	X	X	X	X		X	X	X	X	X	**11**
MO+ME+V	**3**	**2**	X	X	X	X	X	X	X	X	X	X	X	X	**12**
SeaW+MO+ME	**766**	**516**	X	X	X		X	X		X			X	X	**8**
MO+V	**25**	**27**	X	X	X		X	X	X	X			X	X	**9**
SeaW	**1596**	**880**	X	X	X					X			X		**5**

Table 2: Statistical results of comparison between three PFTs included AGIOP retrieved PFTs (with different methods, band ranges and water types considered) and HPLC-based PFTs. Blue highlights the best statistical result. R = Spearmann correlation coefficient, ME = Mean Relative Error, RMSE = Root-Mean-Squar Error.

Minimization Method	Band range	First guess	Water type	Retrieved parameter	R (Spearman)	R^2	ME	RMSE
NLSQ	400-600	OC4 Chl-a	All	Total Chl-a	0.71	0.26	55.2	1.31
				Diatoms	0.43	0.08	349.7	2.12
				Haptophytes	0.56	0.46	82.5	0.34
				Cyanobacteria	0.42	0.08	272.5	0.27
NLSQ	400-600	OC4 Chl-a	Case 1	Total Chl-a	0.78	0.48	51.1	0.97
				Diatoms	0.60	0.29	211.0	1.23
				Haptophytes	0.69	0.46	59.1	0.36
				Cyanobacteria	0.42	0.17	209.6	0.21
NLSQ	OLCI 10 bands	OC4 Chl-a	Case 1	Total Chl-a	0.80	0.50	77.1	1.15
				Diatoms	0.47	0.43	510.1	1.25
				Haptophytes	0.62	0.42	54.5	0.45
				Cyanobacteria	-0.08	0.00	615.6	1.03
NonnegLMI	OLCI 10 bands	OC4 Chl-a	Case 1	Total Chl-a	0.76	0.45	37.1	0.77
				Diatoms	0.82	0.57	127	0.38
				Haptophytes	0.78	0.50	235	0.43
				Cyanobacteria	0.12	0.00	248	1.34

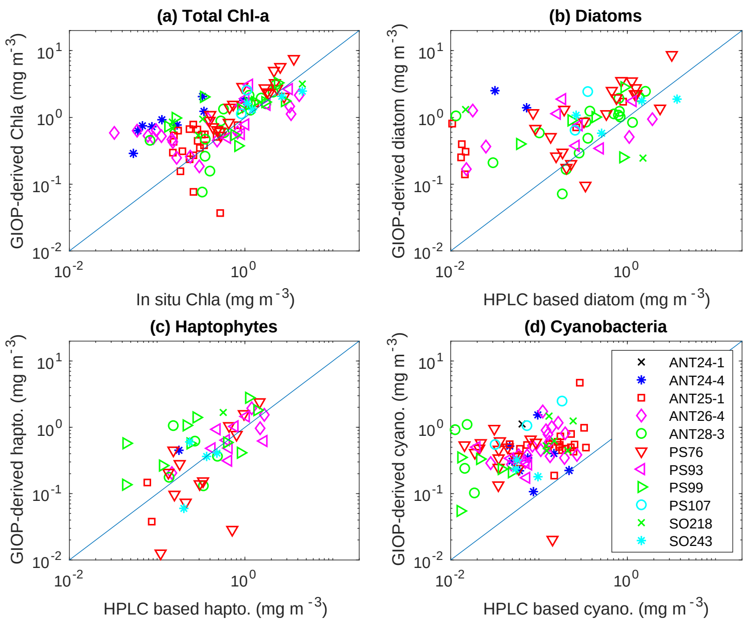

Figure 1: Scatterplots of the AGIOP retrieved (a) total Chl-a, (b) diatoms, (c) Haptophytes, and (d) Cyanobacteria versus HPLC-based Chl-a and PFTs. Here the GIOP was conducted with NLSQ, R_{rs} ranging in 400 – 600 nm and only points in case 1 waters were included (statistics marked blue in Table 2).

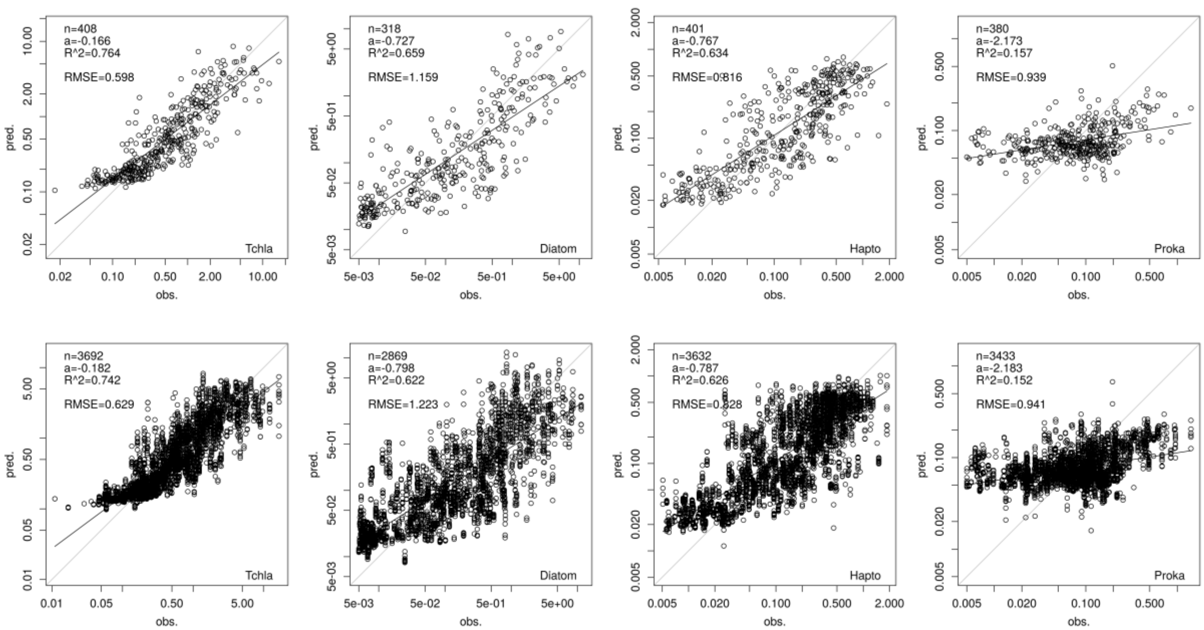

Figure 2: Regressions between observed (obs.) PFTs (based on DPA) and predicted concentrations (pred.) based on EOF modes derived from R_{rs} at 9 wavebands from GlobColour merged products. Top panel: R_{rs} 1x1 pixel at 9 bands, bottom panel: R_{rs} 3x3 pixels at 9 bands.

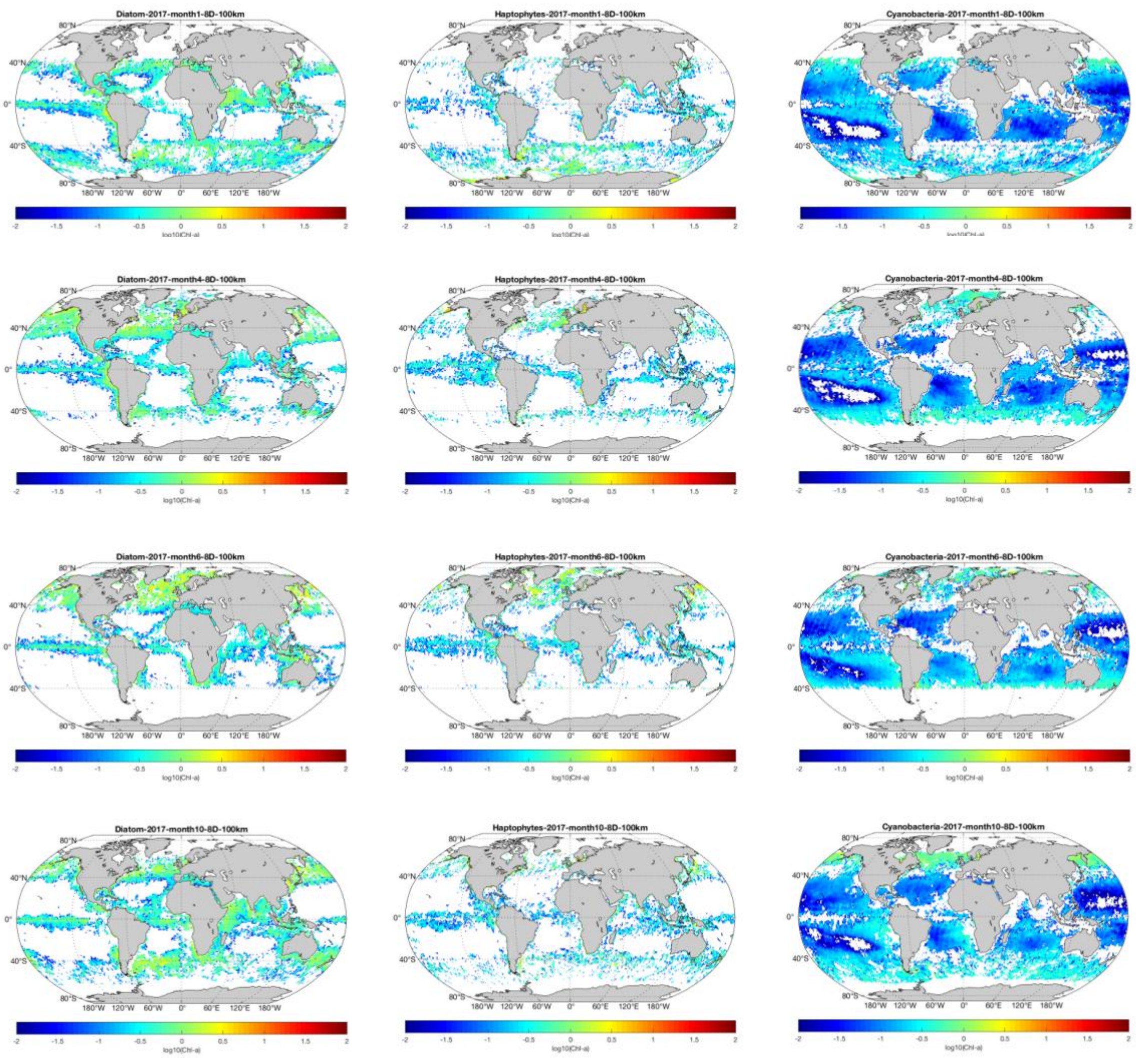

Figure 3: Global PFTs retrieved by AGIOP applied to monthly OLCI R_{rs} data (GlobColour level 3 product) in January, April, June and October 2017. Left panel: diatom, middle panel: hyptophytes, right panel: cyanobacteria.

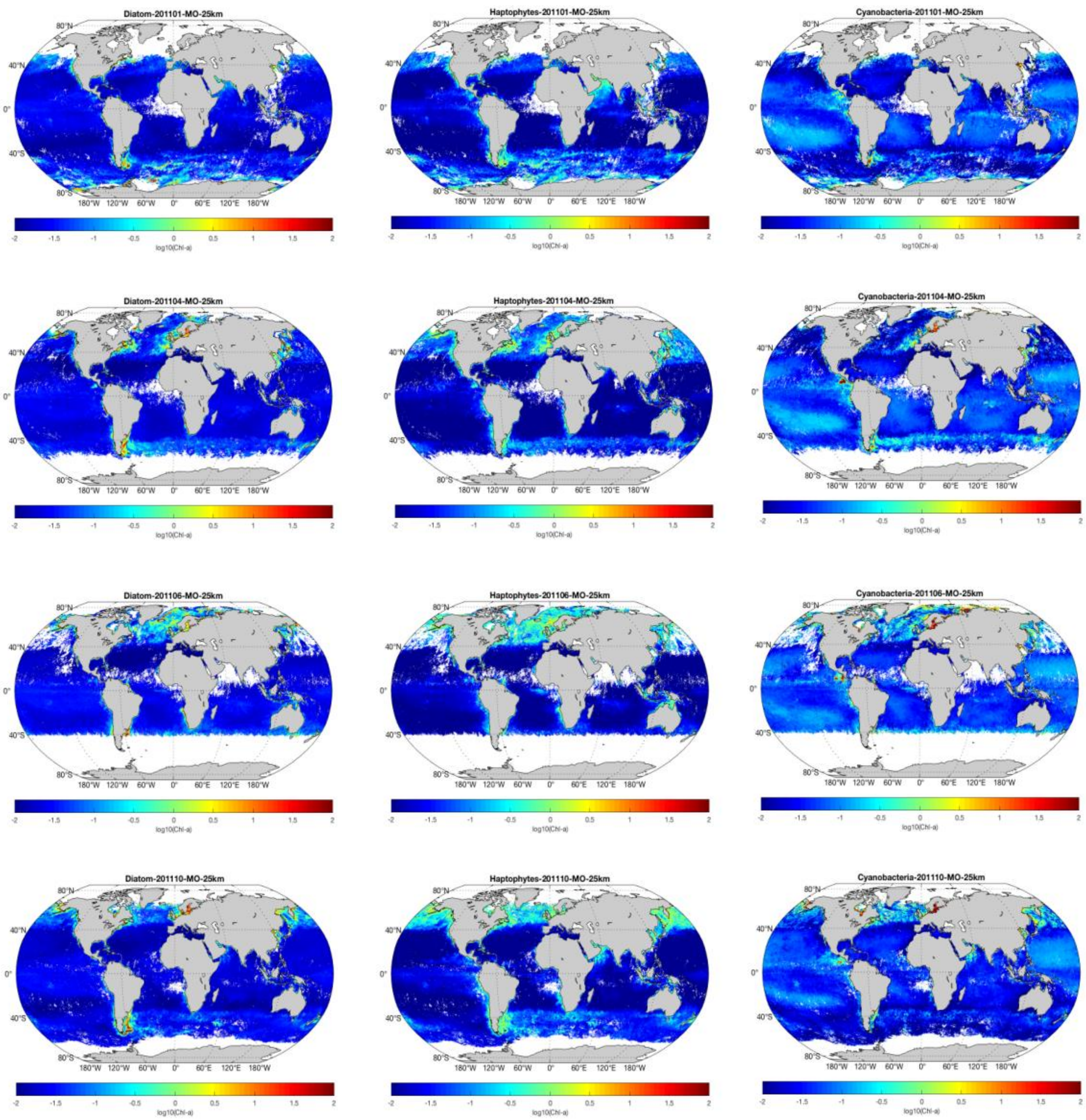

Figure 4: Global PFTs retrieved by EOF-based algorithm (9 bands) applied to GlobColour merged monthly R_{rs} products in January, April, June and October 2011. Left panel: diatom, middle panel: hyptophytes, right panel: cyanobacteria.