

## **Answer to Anonymous Referee #2 comments**

We thank the Referee #2 for the valuable comments which we used to revise our manuscript. In the following we state each referee comments in italics and answer directly below to each of them also highlighting how and where we changed the manuscript accordingly. All changes in the manuscript are visible in red, except for Table 2 because here red signifies already “medium quality”.

Note:

- 1) During revision we noticed that 5 of our samples in the satellite data set have not been used by mistake in the model due to some annotation error in the input data. We now revised the input data description and by that increased the number of sample for the 1x1 (now 139 instead of 135), 3x3 (now 155 instead of 150) and 5x5 (now 160 instead of 155) collocations. Results are still very similar, but because of that all related Figures (Fig. 3c & f, 6, 7d-f and 8, according to new figure names) and Tables (Table 1, Table 2b, Table 3c, Table 4) had to be changed. However, Table 1 and Fig. 8 did not change at all.
- 2) As pointed out by Referee #2 (see point 11) we have used by mistake the wrong PPC (=Allo+Diadino+Diat+Viola+Lut+Zea+Caro) definition for Hooker et al. (2005) with and now revised it to the correct one (=Allo+Diadino+Diat+Zea +Caro), accordingly for all models the PPC results were recalculated and all related Tables (in addition to the above mentioned (Table 2a, 3a, 3b and Fig. 5c, 7a-c and 8f-according to new figure names) were revised.

*This study provides an EOF-based ocean color approach to estimate multiple phytoplankton pigments from the radiometric signals. It is an important issue in oceanography and well within the purview of OS. However, the manuscript should be constructed more logistically. I would recommend it acceptable for publication after the necessary moderate revisions made.*

### *General comments*

*1. It is better to use tables, rather than many confused sentences which are hard to follow, to describe the different data sets used for analysis. For example, in Section 2.1, create a table to clearly show what different data sets are used in developing the EOF based approach and for the validation. Also, include, at least, these columns of: cruise, period, sensor, parameter, and data point.*

We have revised (together with our U. S. American coauthor) carefully the whole manuscript in order to improve its language, style and logic. We have added a flowchart (new Fig. 2) to explain how the different data sets were used in our model. The flow chart gives an overview of the model development prediction and model validation. Regarding the source of pigment data Supplement-Table 1, before only containing the information on the pigment data collocated to the satellite reflectance data set (now presented in lower panel, regarding cruise name, time period and area sampled) was extended. Now the number of data points of each collocation is given and the same information also for the pigment data collocated to the field (upper panel) reflectance data set is given. Chapter 2.1.2 and 2.2 have been extended to give the details on the sensors used to obtain the measurements for the field and satellite, respectively, reflectance data sets.

*This is a paper in developing an EOF-based approach for estimating pigments from space. Thus, two independent data sets should be used in the development of the approach and in the validation, respectively.*

As pointed out to Referee #1 we have applied the cross validation technique to provide proof of prediction models. This technique allows the development of the models with different data than the validation and the resampling of the whole data set for 500 different subsets (i. e. 500 permutations) into testing and validation data sets we get a valid proof of the robustness of our models. To discuss limits and application of our method further we have added this explanation to the discussion (2<sup>nd</sup> last paragraph of section 3.5).

2. a) *It is unclear what HPLC pigments are used in the EOF analysis. It seems that the authors used the list of pigments or groups in EOF training, such as: TChl a, PSC, PPC, MVChl a, DVChl a, etc in Table 2.*

b) *Since, for example, TChl a is calculated from the sum of MVChl a, DVChl a and chlorophyllide a, is it a fair way to use TChl a, MVChl a and DVChl a together to develop an EOF-based approach? To my understanding, any one of the pigments used for the EOF training should not be so relevant or dependent to another one.*

a) From this comment, we suggest that it is not clear to the referee that we have developed separately for each pigment or pigment group its specific EOF model. Before we stated before the sum of all pigments and pigment groups (i.e. pigment columns) as “N” which was misleading because also the number of bands was named “N”. To clarify we changed now the pigment column number to “P”.

b) It is expected that models for TChl a and MVChl a are similar (as seen in Table 3), but quite different for Div a and Zea which do not vary as the overall phytoplankton biomass varies. We hope that now by providing the flowchart (new Fig. 2) this becomes clearer.

#### Specific comments

1. P2076, L20-22: *Are the functions of the photo-protective carotenoids only to protect PSC from photo-damage?*

We actually wrote this sentence differently, as that PPC also protects the chlorophylls and other sensitive pigments from photodamage: “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 are protecting Chl a and other sensitive pigments from photodamage (photoprotective carotenoids, PPC).”

*Zeaxanthin may also be found in a lot of phytoplankton groups other than cyanobacteria (Jeffrey et al., 1997, Phytoplankton pigments in oceanography, UNESCO Publishing).*

As suggested, we changed the sentence to peridinin which is only found in dinoflagellates.

2. P2078, L6: *Use the abbreviations consistent to all parts of the text. For example, RRS and Hex/But are used in P2079 and P2082 L17, but RRS and 19BF/19HF are used in P2080 and P2082 L24, respectively.*

We changed that accordingly.

3. P2079, L1: *Give out the full name of MERIS.*

We changed that accordingly.

4. P2079, L5: *Make sure the cruise names are consistently used in the whole text.*

We changed that accordingly.

5. P2079, Section 2.1.1: *Briefly describe what instrument and column are used for HPLC system.*

This information was added accordingly: “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 (a Waters 600 (Waters, USA) controller with, a Waters 2998 photodiode array detector, a Water717plus auto sampler and a LC Microsorb C8 HPLC column) as detailed in Taylor et al. (2011).”

6. P2079, L22: *Are HPLC samples not available for other cruises?*

The pigment data of the other two cruises are now incorporated in a manuscript for the first time. To make that clear we changed the sentence to “HPLC data for ANTXV/1, as opposed to the other two cruises, were already published in Taylor et al. (2011) and are available from PANGAEA (doi.pangaea.de/10.1594/PANGAEA.819070).”

7. P2080, Section 2.1.2: *It is better to briefly describe how to make the AOP measurements here.*

We briefly now describe the measurement but for the details still refer to Taylor et al. (2011) where those measurements have been detailed in chapter 2.1.2:

“For all three cruises as AOP input data, we used  $R_{RS}(\lambda)$  data obtained from profiles of radiance and irradiance from 320 nm to 950 nm with an optical resolution of 3.3 nm and a spectral accuracy of 0.3 nm measured with hyperspectral radiometers (RAMSES, TriOS GmbH, Germany) measuring at the same time and place when pigment data of section 2.1.1 were sampled.”

*What is the spatial resolution of the radiometric sensors?*

We are not sure what is meant by this question. The radiometric measurements were obtained directly in the water column and probably reflect the optical environment within a few meters of the instrument placed. (We do not know the sensors spatial resolution and this probably can only be obtained from radiative transfer calculations since it will vary for the different wavelengths in dependence to the specific IOPs which we have not measured).

8. P2082, L1-7: *What does MERMAID represent? What is the time window between the water sampling time and satellite overpass time?*

We changed this sentence accordingly to “Matchups between pigment data and MERIS Polymer  $\rho_{wN}(\lambda)$  and TChla products were determined according to the MERMAID (MERIS MATCHUP *In-situ* Database) data base as 1x1 (within the MERIS pixel), as 3x3 and 5x5 pixels measured at the same day around the field observation (see Barker et al. 2008).”

9. P2082, L10: *“For the field. . .”. This sentence is unclear.*

We changed actually the whole paragraph and added a new figure (flowchart- see new Fig. 2) outlining out the model development and prediction. Accordingly we also added a last paragraph referring to this figure.

10. P2082, L16: *MVChl b*.

Was changed.

11. P2082: *The definitions in PSC and PPC seem not same to those in Hooker et al. (2005). Are all these pigments/groups used for the EOF-based analysis. Please identify those used in the paper only.*

The referee is correct for the PPC calculation, the PSC calculation was done as suggested in Hooker et al. (2005). We therefore corrected this statement (“and according to Hooker et al. 2005 and Roy et al. (2011) the photosynthetically active carotenoids (PSC: Fuco, **But**, **Hex**, Peri), and the photoprotective carotenoids (PPC: Allo, Diadino, Diato, ~~Viola~~, ~~Lut~~, Zea, Caro).”) and redid all calculations for PPC (as outlined at the beginning of our final response).

12. P2083, L10: *“where done”?*

Was changed to “were done”.

13. P2083, L14: *“N pigment columns”. Please list the pigments used.*

We have listed the pigment and pigment groups in the beginning of section 2.3 (now the last paragraph), but to clarify we now introduced here a reference “Spectral samples were collocated to the respective pigment data set  $\mathbf{Y}$ , of dimensions  $M$  sample rows by  $P$  pigment columns (pigments and pigment groups included are outlined above).” (See also our answer to General comment 2b).

14. P2084, L19: *What do “e1,2,...” represent?*

Thank you for that comment, this was actually erroneously changed during type-setting the manuscript. We will carefully check the proofs this time!

15. P2085: *Why not to use the log-transformed pigments for calculating RMSE, etc.? It seems the statistical results shown in Figures 4-5 are based on log-transformed data. Is the definition in “R2” correct?*

As suggested (also by Referee #1), we have now modified the RMSE and RMSE<sub>cv</sub> calculation based on the log-transformed comparison. Accordingly Table 2, Table 4, Fig. 5, Fig. 6, Fig. 7c and Fig. 7f (according to new figure names) were modified. For Fig. 7c and 7f we present now the jack-knifing results for RMSE<sub>cv</sub> as ratio to the RMSE (as it was done for  $R^2$  and MPD). The recalculation of RMSE and RMSE<sub>cv</sub> on log-transformed comparisons still proves the validity of our method.

16. P2086, L15: *steps 2-8.*

Was corrected.

17. P2087: *Similarly to comment 15, why not to use the log-transformed data?*

See our response to Point 15.

18. P2088, L2-6: *This sentence is confused.*

To clarify we have changed the sentence to: “In order to predict pigment concentration from MERIS  $\rho_{wN}(\lambda)$ , for a whole month of data in November 2008, where we did not have corresponding pigment measurements, the following method was applied: 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. 11, Fig. 2, right panel), where  $b_{1,2,\dots,n}$  are taken from the EOF model developed with the 1x1 MERIS Polymer  $\rho_{wN}(\lambda)$  matchups (following Fig. 2, left panel).”

19. P2088, L17: *Why are there only two seasons? There are three cruises.*

We actually now moved this on request of Referee #1 to section 2.3 and removed the statement on the two seasons (which is mentioned in Sect. 2.1.2). However, one season was sampled by cruises ANTXXIV/4 and ANTXXVI/4, and the other season by the ANTXXV/1 cruise.

20. P2089, second paragraph: *These sentences are not well logistical.*

To clarify we have changed these sentences to “The composition and range of pigments (as detailed with maximum, minimum, mean and standard deviation in Supplement -Table 2) shows for all pigments (except for Fuco for which it is equal, and for Zea for which it is inversed) that the collocations to the field data set contain higher maxima and minima than the collocations to the satellite-based data set. Mean values are for most pigments very similar among both data sets, but standard deviation is rather 2-3 times higher than the mean value for all pigments (except for PPC and Zea) in the field data set as opposed to the standard deviation being of a magnitude similar to the mean value in pigment data collocated to the satellite data set.”

21. P2090, L1: *“3.3 nm. the hyper00?”*

Was removed.

22. P2090, first paragraph: *Are the discussions on the higher-order EOF functions necessary?*

Yes it is, because those are relevant the loading of these EOF modes are necessary components of the linear models for all other pigments predictions besides TChla, MVChla and PE.

23. P2091, first paragraph: *The interpretations are not well clear.*

We have reorganized the whole subchapter 3.2 (also in response to the comments by Referee #1) and also added a summary at the end. We hope it is now much clearer.

24. P2095, second paragraph: *Confused sentences.*

The sentence was corrected and moved to the end of the first paragraph 3.3.3 “Nearly all linear models using the hyper\_ $R_{RS}$  data set to predict pigment concentrations incorporated the loadings of three to five EOF modes. By contrast, predictive models for DVChla, Zea and PPC incorporated nine, eight and six, respectively.” (now in first paragraph of section).

25. P2099, Section 3.6: *Is DVChl a not directly calculated from the EOF-based approach?*

*If so, why to use the method to subtract MVChl a from TChl a?*

DVChla cannot be calculated from the MERIS  $\rho_{wN}(\lambda)$  data models because it delivers poor predictions. Only when collocation with DVChla=0 mg m<sup>-3</sup> are removed predictions are possible. But these models are not appropriate to be used for this region where we want to extract the predictions because as it has been shown by the in-situ HPLC data also here values of DVChla=0 mg m<sup>-3</sup> occur. Therefore, DVChla had to be calculated from the subtraction of well predicted MVChla from well predicted TChla.

26. *Figure 1: The knowledge on the sampling locations at different seasons/years may be more important. Since the match-up locations based on the 1x1, 3x3 and 5x5 pixel criteria agree more or less to each other, only one set of location (e.g. based on 5x5 criterion) is needed to show here.*

We added as wished by Referee #1 the lat/lon and the legend on the plot. Besides this, we prefer to leave the figure like it is, because the information of sampling locations at different seasons/years is provided in Supplement Table 1 and we now discuss and show more the different results of the models based on 1x1 versus 3x3 versus 5x5 MERIS  $\rho_{wN}(\lambda)$  reflectance data.

27. *Figure 2: “standardized (subtracted mean). . .”, but not divided by the SD?*  
Was corrected (now Fig. 3).

28. *Figure 3: add “dotted lines”.*  
Was corrected (now Fig. 4).

29. *Figure 7: Show latitudes/longitudes.*  
Was corrected (now Fig. 8).