Ocean Science
Discussions

Interactive comment on "Estimation of phytoplankton pigments from ocean-color satellite observations in the Sénégalo-Mauritanian region by using an advanced neural classifier" *by* Khalil Yala et al.

Khalil Yala et al.

crepon@locean-ipsl.upmc.fr

Received and published: 20 December 2019

Answers to reviewer n°1

We first thank the reviewer for his helpful comments and suggestions that have helped us to improve the manuscript. In the following, we answer point by point using the following convention: The reviewer comments are in italic Our answers are in standard typo The changes we made according to the recommendation of reviewer 1 of are in yellow in the track document.

C1

1. The introduction can be written in a much more accurate way. I would check it phrase by phrase and sentence by sentence. We rewrote the introduction taking into account the remarks of the reviewer

1.1. Line 50-52 For example, one limitation of microscopy is the difficulty in indentifying picoplankton 1.2. The optical microscopy method is developing, for example the imaging flow cytometry (IFC). We rewrote these lines: "Microscopy is time-consuming and is unable to identify picoplankton. Imaging flow cytometry (IFC) has renewed microscopic methods, thanks to the speed at which they are able to characterize phytoplankton in a water sample (IOCCG report n°15, 2014)". (Lines 49-52 in the revised version).

1.3. Line 54-55: Mind the use of the terms PSC and PFT. PFT depends on how you define it. PSC is also a type of PFT definitions. Pigments allow estimating phytoplankton groups but not phytoplankton species. We withdrew this statement in the text.

1.4. Line 57-60: the conversion formula method is the so-called "Diagnostic Pigment Analysis". CHEMTAX uses matrix factorization to estimate PFT from pigments. We mentioned the so-called "Diagnostic Pigment Analysis" line 57 1.5. Line 60: I am not sure with just marker pigments themselves the identification of phytoplankton can be achieved in species level. We agree and we, therefore, modified the text of the revised version

1.6. In summary, please check IOCCG report 15 and related literature carefully. According to comments n°3, 4, 5, 6 we rewrote these lines which are now (Lines 52-61 in the revised version) taking into account the material in the IOCCG report 15: "An alternative method is the analysis of seawater samples by high-performance liquid chromatography (HPLC) which is widely used to categorize broad phytoplankton groups such as PFT or PSC (Jeffreys et al, 1997, Brewin et al, 2010, Hirata et al, 2011). HPLC enables identification of 25 to 50 pigments within a single analysis, which is much easier and faster to conduct than microscopic observations. Each phytoplankton group is associated with specific diagnostic pigments and a conversion formula can

C2

be derived to estimate the percentage of each group from the pigment measurements (Vidussi et al, 2001; Uitz et al, 2010). HPLC measurements are now recognized as the standard for calibrating and validating satellite-derived chlorophyll-a concentration and for mapping groups of phytoplankton (IOCCG report n°15, 2014)".

2-

2. Lines 139-140 Match-up procedure can be more detailed, for example, by adding the criteria of refusing data points and the reason why you choose 20km We rewrote these lines in the revised version of the manuscript (lines 138-151) "Matchup procedure between in situ and satellite observation is a crucial question to estimate remote sensing algorithms. If the parameters of the procedure are too severe, the number of collocated data is dramatically decreasing. If the parameters are too large, the accuracy of the matching is decreasing. We then chose some compromise. Usually, people use a matchup window of 3X3 pixels (Alvain et al, 2005) which corresponds to a distance somewhat less than 20km between the satellite pixel and in situ measurement since we deal with level 3 satellite observations whose pixel is of the order of 9X9km. This criterium refers to the typical length of ocean variability (Levy et al, 2012; Levy, 2003)"

3. Lines 150-160 and Figure 3. Please use more statistical metrics in addition to R-square and RMSE according to Brewin et al 2015. Please specify whether they are calculated in log scale or not. Brewin, Robert JW, et al. "The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms." Remote Sensing of Environment 162 (2015): 271-294 Brewin et al (2015) give a large variety of statistical parameters because they compare a large number of models whose performances are close together, which implies the use of several criteria to separate them. In the present study, we only need to estimate the quality of our model, which can be done by standard statistical parameters as usual. Concerning the pigment concentrations, the statistical tests were done in mg.m-3. We included this information in the text (lines 181-183).

In figure 3, we present the regression line between Chla- given by OC4V4 and in situ chl-a. The data are given in mg.m-3 and the statistical estimators were computed in mg.m-3 but the scale in figure 3 is log scales.

4. Lines 288-289: you have said the same as Line 264-265. We insist on that point because it constitutes the original component of 2S-SOM.

4. Table 2: often these statistics are done on log(pigments) - given their distribution and expected errors. Our strategy is to compute the statistical parameters in the physical space as most statisticians do and as did Brewin et al (2015) to facilitate the interpretation. The concentration values are normalized during the learning procedure of the SOM.

5. Line 402: Unfortunately, it cannot be concluded that diatoms dominated because of high Fuco ratio and chl-a, without additional information on phytoplankton groups using e.g. microscopy. We do not have concomitant microscopy measurements. When analyzing the referent vectors presented in Fig 6, we strongly think that the bottom right region representing the neurons of the 2S-SOM may correspond to diatoms since high fucoxanthin is associated with high chlorophyll concentration and low peridinin. Besides, it is seen in Figures 8, 10 and 11 that high fucoxanthin geographical regions are situated near the coast where diatoms were observed in previous studies (Farikou et al., 2015; Blasco et al., 1980) while high peridinin geographical regions are situated in offshore regions. We changed our previous sentence in: 'Moreover, the bottom right region in the 2S-SOM may correspond to the diatoms with good confidence since high fucoxanthin is associated with high chlorophyll concentration and low peridinin. This is endorsed in section 5 by looking at the geographical location of the different pigment concentrations (figures 8, 10, 11)'. (Lines 352-356 of the revised version)

6. Please spell MLP out in the Discussion section. MLP stands for Multi LayerPerceptron, it has been added on line 596

7. Line 649-654: Can you summarize why SOM needs fewer data points than MLPs

and other supervised learning? Why MLP cannot be trained with a total of âĹij500 data points? This is a well-known property of SOM versus MLP. The main difference between MLP and SOM is in the learning process: MLP is a supervised algorithm while SOM uses unsupervised learning. Both have to estimate a large number of weights during a learning phase; the accuracy of the methods depends on the dimension of the input and output spaces, the number of data available and the number of weights to estimate. In SOM the weights are highly regularized by the neighborhood function, so the number of data needed for learning is less than for the MLP. In the present application, the MLP would have to approximate a highly non-linear function from the R11 input space (the remote sensing parameters) to the R6 output space that represents the pigments. Due to the highly non-linearity of the function, the 515 data available for the learning is too small to adequately sample the R11 space of the function. On the other hand, SOM is not a regressor but uses automatic clustering methods and provides more robust values. Moreover, the topological order prevents to make errors in interpolating between two clusters. We think this explanation is too long to be included in the present text and out of the scope of the present study. It would be relevant in a Text Book or a review paper dedicated to NN. We propose to escape this question and to withdraw the sentence at line 650: 'which makes MLPs and classical supervised learning methods unusable' The sentence is now: 'We used an unsupervised neural network classification method which is an extension of the SOM method well adapted to deal with a small database whose elements are very inhomogeneous'(lines 605-607 of the revised version)

8. Is it possible to clarify the minimum threshold of pigment concentration of the applicability of 2S-SOM? The minimum and maximum values of a parameter are those of the learning data base. As the 2S-SOM has 162 neurons, the interval between the minimum and maximum values is divided into 162 discrete values corresponding to the values captured by the referent vector associated with each neuron. Classification acts as a piecewise continuous model permitting the achievement of complex tasks. We get these discrete values empirically only by looking at the different referent vectors of the

SOM.

TECHNICAL CORRECTIONS

1. The country Senegal has three versions of names in the manuscript, i.e. SeÌ ËŻAneÌ ËŻAgalo (title), Senegalo (context) and Senegal (Figure 1). Please keep the consistency. We homogenized the spelling of Senegal in the revised version

2. line 41 The word "phytoplankton" is more often used as a plural modified (line40, 41, 49 of the revised version)

3. Line 42-44: mind the subscript of CO2 modified

4. lines 43-44: I have not found the information of 30% in Behrenfield et al, 2005 We put a more appropriate reference for the rate of CO2 captured by the ocean: "Le Quéré et al, 2018" (line 43)

5. line 48: The description "fish grazing on phytoplankton" is not accurate. The effect of phytoplankton on fisheries is via marine food chain, i.e. zooplankton grazing on phytoplankton provide food source for some fish. We changed the sentence as: "and fisheries with a possible effect on fish grazing on phytoplankton via the marine food chain" (line 46-47 of the revised version)

6. Line 56: Please add the citation: Sosik, H.M.; Sathyendranath, S.; Uitz, J.; Bouman, H.; Nair, A. In situ methods of measuring phytoplankton functional types. In Phytoplankton Functional Types from Space. Reports of the International Ocean-Colour Coordinating Group (IOCCG), No. 15; Sathyendranath, S., Ed.; IOCCG: Dartmouth, NS, Canada, 2014; pp. 21–38 Done (line 56 in the revised version)

7. Line 84: use the abbreviation of "PSC". Full name is not needed Done

8. line 86: the term "PSC percentage" is inaccurate. It is the contributions of Chla from different phytoplankton size classes to total Chla concentration We modified the sentence as: ' These algorithms try to establish a relationship between the chl-a con-

centration and the chl-a concentration fractions associated with each of the three PSC' (lines 86-88 of the revised version)

9. Line 105: the colour of the land is not red. We changed 'red' into 'brown

10. Line 111: delete "a". 11. Line 112: "systems". 12. Line 161: "wavelengths". 13. Please define the abbreviation of a variable before using it (e.g. Table 1 and a lot of places). We implemented the suggested corrections.

14. lines 181-182: this not a sentence We modified this line which is now 'which is defined as the ratio of the diagnostic pigment (DP) versus the total chl-a'.(lines 178-179 of the revised version)

15. Line 182: typo: divinyl chl-a. Did you consider chlorophyllide-a as part of Tchl-a? We used the definition of Alvain et al (2005), where Chl-a is part of Tchl-a (Tchl-a= Chl-a+ Divinyl chlorophyll-a). (line 179)

16. Line 186-190: you have mentioned these in Line 113-117 We delete the sentence in lines 186-190

17. Figure 4&5: Rrs is not defined. Rrs stands for ïĄšw(ïĄň), we made the change in figures 4 and 5 in the revised version

The manuscript has been read and corrected by a native English-speaking person

Added references Levy, M., Mesoscale variability of phytoplankton and of new production: Impact of the large-scale nutrient distribution, J. Geophys. Res., 108(C11), 3358, doi:10.1029/2002JC001577, 2003.

M. Lévy, D. Iovino, L. Resplandy, P. Klein, G. Madec, A.-M. Tréguier, S. Masson, K. Takahashi, (2012) Large-scale impacts of submesoscale dynamics on phytoplankton: Local and remote effects, Ocean Modelling,77–93

Le Quéré et al, (2018) Global Carbon Budget 2018, Earth Syst. Sci. Data, 10, 2141–2194, 2018 ; https://doi.org/10.5194/essd-10-2141-2018

Sosik, H.M.; Sathyendranath, S.; Uitz, J.; Bouman, H.; Nair, A. In situ methods of measuring phytoplankton functional types. In Phytoplankton Functional Types from Space. IOCCG report, No. 15; Sathyendranath, S., Ed.; IOCCG: Dartmouth, NS, Canada, pp. 21–38, 2014.

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11, 2019.