Ocean Sci. Discuss.,
https://doi.org/10.5194/os-2019-11-AC1, 2019

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: 14 May 2019

We thank the reviewer for his helpful comments. We answer point by point: We use the following typo : The reviewer comments are in italic Our answers are in normal typo

SPECIFIC COMMENTS 1 Line 50-52 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 (Sosik et al, IOCCG report n°15, 2014).

2-3-4-5 Pigments allow to estimate phytoplankton groups but not phytoplankton species. We withdrew this statement in the text. Lines -53-60 we rewrote this section: 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 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 (Sosik et al, IOCCG report n°15, 2014).

2-Lines 139-140 , we added in the revised version of the manuscript 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. In figure 3, we present the regression line between Chla- given by OC4V4 and in situ Chla. The data are given in mgm-3 but the scale in figure 3 are log scales. Please use more statistical metrics in addition to R-square and RMSE according to Brewin et al, 2015: Brewin et al (2015) give a large variety of statistical parameters due to the fact that 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. Besides the software we used to do the cross validations only computes the RMSE and the R2 parameters. Concerning the pigment concentrations, the statistical tests were done in mgm-3. We included this information in the text.

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) in order 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. We strongly think that the bottom right region in the 2S-SOM correspond to diatoms since high fucoxanthin is associated with high chlorophyll concentration and low peridinin. 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) whilst high peridinin geographical region 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 a good confidence since high fucoxanthin is associated with high chlorophyll concentration and low peridinin. This is endorsed in section 5 of the present paper by looking at the geographical location of the different pigment concentrations (figures 8, 10, 11).

6. Please spell MLP out in the Discussion section. MLP stands for Multi Layer Perceptron

7. Line 649-654: Can you summarize why SOM needs fewer data points than MLPs and other super- vised learning? Why MLP cannot be trained with total âĹij500 dapa 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 that 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. In 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 small database whose elements are very inhomogeneous'

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 SOM use has 162 neurons, the interval between the minimum and maximum values is divided in 162 discrete values corresponding to the values captured by the referent vector associated with each neuron. We get these dis-

crete values empirically only by looking at the different referent vector of the SOM. The discrete values are computed during the learning phase and depend on the density of the learning data set: sparse regions are les sampled than dense ones.

TECHNICAL CORRECTIONS

1 We homogenized the spelling of Sénégal in the revised version 2, 3 OK, we did the correction 4 lines 43-44 We put as a reference for the rate of CO2 captured by the OCEAN : Le Quéré et al, 2018 which is more appropriate 5 line 48 We changed the sentence as : and fisheries with a possible effect on fish grazing on phytoplankton via the marine food chain 6, 7 OK, we did the correction 8 line 86 We modified the sentence as : ‘ These algorithms try to establish a relationship between the chl-a concentration and the chl-a concentration fractions associated with each of the three PSC’ 9 We changed ‘red’ into ‘brown’ 10, 11, 12, 13: OK, we did the correction suggested 14 We modified the line 181 : ‘which is defined as the ratio of the diagnostic pigment (DP) versus the total chl-a’. 15 We used the definition of Alvain et al (2005), where Chl-a is part of Tchl-a (Tchl-a= Chl-a+ Divinyl chlorophyll-a). 16 We delete the sentence in lines 186-190 17 Rs stands for ïĄšw(ïĄň), we made the change in figures 4 and 5 in the revised version

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.

Lévy M. , 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. (2014) 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, 2014; pp. 21–38.

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