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

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 2 of are in turquoise in the track document

There is a lack of comparison with controls for the reader to appreciate the advantage

of using this new model. At the minimum, there should be more comparison between-scewise the new 2S-SOM model performance scores versus the standard SOM model scores. [The paper would be more interesting if the performance of 2S-SOM is also compared against standard supervised learning models such as multi-layer perceptrons or random forests.]

We comment on the advantages/disadvantages of the different methods in the discussion section (line 594-609). An objective comparison of the different methods is out of the scope of the present paper as it would considerably increase the length of the present paper. In fact, it would deserve a full paper (see the paper of Brewin et all (2011) dedicated to a comparison of the different methods and also the paper of Bracher et al, 2017, Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development. Front. Mar. Sci. 4:55.). Besides to be conclusive, such a comparison should be done on a specific region where in situ measurements are more numerous than in the present region. We first used a SOM and then decided to use a 2S-SOM mainly by the information provided by the 2S-SOM on the role of the different variables in the classification process. The major advantage of the 2S-SOM compared with the SOM and other classification methods is to partition the different variables of the dataset under study into blocks and to affect weights to these blocks. The block weighting facilitates the clustering procedure by favoring the taking into account of the most pertinent variables. This method is related to the research area developed in statistics under the designation of clusterwise method (Parson et all 2004; Kriegel et all 2009)

Parsons L, Haque E et Liu H : Subspace clustering for high dimensional data : a review. SIGKDD Explor. Newsl., pages 90105, 2004. ISSN 1931-0145. 73, 74, 80 Kriegel H.-P, Kröger P et Zimek A : Clustering high-dimensional data : A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans. Knowl. Discov. Data, 3(1):1 :11 :58, mars 2009. ISSN 1556-4681. 37, 73, 74, 80

A high weight affected to a block means that the associated variables play a major role

in the classification process; a small value means that the associated variable plays a minor role: this information is of importance to identify the variables which control the process under study. Besides the block weighting provides useful information on the functioning of the classification by permitting to identify the variables which control it and allows us to better understand the dynamics of the phytoplankton communities. This is discussed in lines 371-376 of section 4-2 Analysis of the topology of the 2S-SOM corresponding to the analysis of figure 7 showing the different weights affected to the neurons of the 2S-SOM also in lines 494-509 of section 5 corresponding to the analysis of figure 13, and in line 622-627 of the discussion section. Moreover, we added the block weights α as an output of 2S-SOM in figures 4 and 5

On line 321, the choice of the elongated 2-dimensional grid of 9x18 is not obvious. Why is a more square grid (e.g. 12x12, 12x13 or 13x13) not used?

The size of the map has been determined (line 275-276, added in the new version), by using the SOM software http://www.cis.hut.fi/projects/somtoolbox/download/, assuming that the size of SOM and 2S-SOM depend on the same criteria. We also checked other grid configuration and found that the most efficient is the 9x18 neurons

The paper is very hard to read as there is a tendency to present many undefined symbols all at once, with the symbols remaining undefined until much later in the paper. For instance, Eq.(5) introduces a large number of symbols and terms all at once. The "block" is not explained in a concrete way until the next section (Sect. 3.3), so I had a misconception on how the data were blocked when reading Sect. 3.2. A much more logical order of presentation is to present the concept of blocking variables first, and try to explain as many of the symbols coming up in Eq.(5) before actually presenting the equation. Also around Eq.(5), there are numerous typos and inconsistent fonts (as listed later in this review).

We have rewritten the sections 3.1 and 3.2 describing the functioning of the SOM and 2S-SOM. We put in Annex the mathematical description of that functioning. In the main

text, we only describe the principle of the functioning of the 2S-SOM. We now explain all the symbols we used. The blocks are described in the main text (lines 255-260) before the explanation of the functioning of the SOM and the 2S-SOM, which is in Annex. We also focused attention on the typos

Line 22-23 Thanks to . . . new method. It primarily consists in. . ." is verbose. Simplify to "Our new method consists of .

done (line 22)

Line 25 "carried using" should be "carried out using".

done (line 24)

Line 69 and throughout the manuscript: Bold fonts are for vectors and matrices (see the journal's manuscript preparation guidelines), but here they are often used for scalars and units. There are many places where the font switches back and forth between bold and Roman and italics (e.g. lines 248-251 and line 272).

We carefully read the manuscript and corrected the font errors

Line 151: Need a reference for the OC4V4 algorithm.

We give a reference for the OC4V4 algorithm (O'Reilly et al, 2001) – (line 154 in the revised version)

Line 162 The last sentence of the paragraph and Table 1 need to be moved to after line 183. The table is currently placed before the terms in it are defined.

Done

Line 174: How can Ra be independent of chl-a if it is divided by rho_wref which is dependent on chl-a?

Ra, which is defined as rhow(lambda)/rhow wref(Lambda, chl-a) is the key parameter of the Physat method (Alvain et al, 2005, 2012). rhow(lambda)depends on sec-

ondary phytoplankton pigments + chl-a, while)/rhow wref(Lambda, chl-a) depends on chl-a only. The reasoning of Alvain et al (2005) is that the ratio rhow(lambda)/rhow wref(Lambda, chl-a) depends on secondary phytoplankton only since both depend on chl-a.

Line 248: W is undefined.

W is now defined in line 230

Line 254: should give a specific reference on the kernel and temperature.

References are given in lines 930, 931 of Annex (Kohonen, 2001, Niang et al, 2003)

Line 276: How were B and Pb chosen?

These variables are defined in lines 941 and 944: B is the number of blocks (B=4) and Pb is the number of variables in block b. According to the definition of blocks (lines 257-262), P1= 5, P2= 5, P3= 5, P4= 2.

Line 278: "a" should be alpha.

Corrected (line 953 in the new version)

Line 282: Eta should be beta.

Corrected (line 953 in the new version)

Figure 4: For 2S-SOM, I can see long dash, short dash, space and no space variants.

Figure 4. We check the pdf output corresponding the figure 4. It seems ok in the modified version. Perhaps there was a software problem in the conversion of the original text written in Word into pdf.

Line 420: Last sentence of paragraph: I have trouble understanding this sentence.

We changed this sentence and gave more explanation on the description of the 2S-SOM neurons. The sentence is now (Line 381-384 in the new version): "These neu-

rons correspond to very small chl-a concentrations, which are estimated with large errors. Besides, we remark that high ïĄą values for chl-a correspond to high chl-a concentration values (bottom right of the chl-a panel in figure 7 and figure 6 respectively). For these cases, the clustering assembled data that mainly depend on chl-a concentration".

Fig.13: Top right corner is slightly chopped off.

Done.

Line 542: "a" should be alpha.

Done. We replaced 'a weight' by 'a weight ïĄą' which is clearer (line 497 of the new version)

Fig.16: I don't understand why the black curve tends to lie closer to the blue curve than the red curve is to the blue curve. I would have expected the red curve to lie closer to the blue curve. I might have misunderstood what the curves represent – please give more detailed explanation.

A VIIRS sensor observation is captured by a neuron of the 2S-SOM whose associated referent spectrum is the red curve in figure 16. This referent spectrum is the mean of the different spectra captured by that neuron during the learning phase. Among these different spectra, there is one (black curve in figure 16) which is the closest to the UPSEN spectrum (blue curve in figure 16). It is expected that the black curve is closer to the blue curve than the red curve which is flattened due to the averaging process. We reformulated this description in the text which was not clear in the first version. (line 539-546).

Line 639: Replace "people" with "studies".

Done (line 596 of the revised version).