

Specific comments:

Line 41-43:

The reference Le Quéré et al, 2018 does not provide this information. The oceans presently take up about 26% of the anthropogenic CO₂, and phytonplankton contribute to this via the biological pump. However, the physical pump also plays an important role in the ocean-atmosphere CO₂ exchange. The role of phytoplankton on regulating the climate should be described in a much more precise way.

Line 49-50:

- 1) "microscopy" should be revised as "light microscopy". Microscopy for identifying phytoplankton includes light, electron and epifluorescence microscopy. Epifluorescence microscopy and electron microscopy enable the identification of picophytoplankton.
- 2) Replace "unable" with "very difficult".
- 3) It is not appropriate to say IFC "renewed" microscopy. Faster sampling speed is an advantage of IFC over microscopy. However, both methods have their own limitations.

More information can be found in IOCCG report 2014 and Nair et al 2008.

Nair, A., et al., (2008). Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, 112(8), 3366-3375.

Line 53-54, 84, 91:

full names and abbreviations of PFT and PSC.

Line 58-59:

References Vidussi et al, 2001 and Uitz et al, 2010 are 2 examples of DPA.

Line 60:

IOCCG report 2014 does not provide the information that HPLC measurements are recognized as STANDARD for mapping phytoplankton groups. It originally says "HPLC measurements are now recognized as the standard for calibrating and validating satellite-derived chlorophyll-a concentration. Pigment analysis also has a useful role to play in validating satellite algorithms for mapping functional types of phytoplankton."

Line 48-61:

As far as I understand, this paragraph was meant to address the progress of methods for identifying phytoplankton, as mentioned in Line 48. However, instead of providing a relatively comprehensive review of the methods, the current manuscript only picked up 3 methods (i.e. light microscopy, IFC, HPLC) as the whole story "during the last two decades". I

would refer to the IOCCG report 2014 to the readers for the details and briefly summarize them in this paragraph.

Line 63-68:

It does not hurt if these lines are omitted.

Line 65-66:

light "transmitted" to satellite? "reflected back"?

Line 67:

- 1) water and non-algal particles also influence backscattering and absorption.
- 2) add "and" before the last comma.

Line 68&76:

full name and abbreviation of "CDOM".

Line 69:

"This upwelling radiation" is confusing. Remote sensing reflectance is a ratio. Also, it is more commonly symbolized as "R_{rs}". "rho_w" more commonly stands for "water-leaving reflectance".

Line 76-78:

- 1) replace "detrital" to "non-algal".
- 2) b_b includes the contribution of water and particles (not just phytoplankton).
- 3) please describe exactly what G is related to.

Line 96:

Please spell out 2S-SOM or at least SOM.

Line 99:

- 1) replace "sensed" to "sensing".
- 2) add "in situ" before "measurements".

Figure 1:

- 1) UPSEN campaign is not described (also in Line 122).
- 2) typo in the legend.
- 3) upper right edge is cutted.

Section 2:

The method of measuring pigments is missing. The full names (Line 255-256) and abbreviations (Table 1) of pigments should also be mentioned.

Line 132-133 & 134-135:

repetitive.

Line 136, 140:

I may be wrong but I do not understand the word "severe".

Line 139, 151:

I have difficulty in understanding these sentences.

Line 147:

What do you mean by "parameters too large"? Does it mean the matchup window is too large?

Line 148:

- 1) replace "people" with "studies".
- 2) specify the spatial resolution of Seawifs, i.e. how big one pixel is.

Figure 3:

- 1) coefficient of vraisemblance?
- 2) delete one "in" at the end.

Line 164:

- 1) replace "having" with "with".
- 2) rho_w and R_a should not be called as "ocean reflectance".

Line 165:

add "respectively" for rho_w and R_a.

Line 167:

missed half bracket.

Line 170:

Please briefly describe how rho_wref is computed.

Line 191-192:

five rho_w and five R_a at five wavelengths?

Line 237-238:

I have difficulty in understanding these sentences.

Line 240:
Delete "frequently".

Line 243&282:
Please use the "distance" term consistently.

Section 3-2:

1) The 2S-SOM algorithm is the central part of the paper, and should then be addressed in detail directly in this section rather than in the appendix. The step-by-step procedures of applying this method to the dataset can be put in the appendix.

2) the currently description of the 2S-SOM method is not easy for me to understand, e.g would you possibly describe clearly how the dimensionality is reduced and how the weights are estimated? Also, adding an architecture of the 2S-SOM map would help.

Line 274:
How many classes in the case of 9x18 grids?

Line 300:
Again, the readers should not be sent to the appendix to check what mu and eta stand for.

Line 322:
How are the p-values of cross validation calculated?

Line 696:
concentrations.

Line 938:
k-means

Section 3-4.1:
please check again language and grammer mistakes.

I hope you'll find the comments helpful.

Best Regards,
Yangyang Liu (yangyang.liu@awi.de)