Biogeosciences Discuss., 11, C4605–C4609, 2014
www.biogeosciences-discuss.net/11/C4605/2014/


Biogeosciences Discussions

BGD

11, C4605–C4609, 2014

Interactive Comment

# *Interactive comment on* "Remote sensing algorithm for sea surface $CO_2$ in the Baltic Sea" *by* G. Parard et al.

**Anonymous Referee #1**

Received and published: 25 August 2014

Review of Parard et al 2014; Remote sensing algorithm for sea surface CO2 in the Baltic Sea

Parard et al present a combined self organizing map and linear regression algorithm to create monthly temporal and high resolution (4 km) spatial maps of the sea surface pCO2 in the Baltic Sea. The algorithm is based on relationships of pCO2 observations with satellite derived predictor data. Furthermore, data gaps arising from cloud cover, etc. are filled using a self organizing map. The authors set their main focus on the methods part, but also present a validation section and some results in the end of the manuscript.

I do think that the part where the authors use a SOM to fill satellite data gaps is very interesting. Unfortunately, particularly this part is subject to a forthcoming article and

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

is not much discussed in the current manuscript. Besides that, the combination of a SOM and a regression technique as such is not unique (e.g. Landschützer et al. 2013 or Sasse et al. 2013) and hardly qualifies to be published as a stand-alone method paper – at least not without discussing the differences to existing approaches. Having said that, the manuscript would further greatly benefit from a much longer and stronger results section.

There are other major and minor concerns regarding this manuscript which are listed below. Considering all these points, I can not recommend the manuscript as it is currently presented to be published in BG.

Specific comments:

.) General:

- The manuscript needs substantial language editing. There are many grammar and spelling mistakes - too many to be all listed here.

- As mentioned above, the combination between a SOM technique with some sort of regression technique is not unique. What is different to previous studies combining two techniques? Unfortunately, the gap-filling technique (section 3.1) is not discussed in more detail. This would be a very interesting and useful application. Potentially, this could e.g. lead to satellite derived high resolution global ocean pCO2 datasets.

.) The methods (section 2):

- particularly this section is very difficult to follow and for a non expert it is almost impossible to understand how the method works. One example from the text: "The topological aspect of the maps can be justified by considering the map as an undirected graph on a two-dimensional lattice whose vertices are the N classes" - I really struggle to understand this, and many other descriptions in the text. It would be much simpler to first in "easier language" describe how the method works, i.e. provide a summary first, and only afterwards go into more technical details. Alternatively, as both SOM and multi-linear regression techniques have been published before, only refer to those settings that are unique in your approach.

- I appreciate the effort to combine several products to create their pCO2 and remote sensing data sets. There are however several questions arising, e.g. how well does the combination of the remote sensing datasets work, i.e., are there discontinuities or steps in years where you switch from one dataset to the next? All the authors state is on page 12262 lines 24-25 that "Some corrections were applied for each parameter of the data to render the different products coherent between themselves".

- Nowhere the authors discuss why explicitly these predictors listed in section 2.3 are used. The authors need to better link the parameter choice to the introduction, where the Baltic Sea system is discussed. Or have the parameters been chosen because of their successful application in the open ocean?

- on page 12263 line 19 the authors state: "In our case, the each data point is characterised by . . . but also an information on the date the measurements were taken". Why do you need extra information about time? Would you not expect this information being present in the other predictor data (SST, etc.)?

- On several occasions you mention that the relationships between pCO2 and remote sensing data are non-linear and you claim that using the SOM entitles you to the assumption that a linear regression is sufficient, but you provide no evidence for that assumption.

.) The validation (section 3.2.2):

- You create monthly estimates for more than a decade, but you only provide one number for validation purposes (though for each neuron), i.e., the root-mean squared error and the correlation coefficient. What about spatial and temporal biases. The current validation section provides no evidence that the method can be trusted over the full time period or spatial extend.

- How do you choose the independent validation data? Randomly? If so, would it not be possible that some of these data are not really independent in the presence of autocorrelation?

- What about other potential data sources for validation? Have you considered the SOCAT or LDEO databases?

.) The results:

- One main achievement of this study (the way I understand it) is that you create monthly maps on a very fine spatial resolution. However, this is hardly outlined in the manuscript and provides strong improvement towards the much coarser open ocean studies.

- The paper could potentially be improved by showing more results, e.g. seasonal cycle or trend.

.) The figures:

- I am missing a proper figure description in all figure captions. It is not straight forward to understand the figures this way. Additionally, figures 1, 13 and 14 are very difficult to read.

- Many figures presented are not essential and can easily be removed. For example figure 8 shows the number of observations for each neuron of the SOM, however, the same information is presented in figure 11 with the number of observations printed in each hexagonal field.

- In the text (page 12267 lines 16-17) the authors state: "After this imputation of the missing data through the iterative training give a good representation of the data as presented in Fig. 7" This is absolutely not clear looking at figure 7. It would be more clear to e.g. first show the distribution without the missing data and then compare the distribution to the full set.



BGD

11, C4605–C4609, 2014

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

References:

Landschützer, P.; Gruber, N.; Bakker, D. C. E.; Schuster, U.; Nakaoka, S.; Payne, M. R.; Sasse, T. & Zeng, J. A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink Biogeosciences, 2013, 10, 7793-781

Sasse T. P.; McNeil B. I. & Abramowitz, G. A new constraint on global air-sea CO2 fluxes using bottle carbon data Geophysical Research Letters, 2013, 40, 1594-1599

---

Interactive comment on Biogeosciences Discuss., 11, 12255, 2014.



BGD

11, C4605–C4609, 2014

Interactive Comment