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 is not much discussed in the current manuscript.

This part will be expanded into a more general methodological article.

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 cur- rently 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.

For the grammar and spelling mistakes the paper was reviewed by a native english speaker.

- 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.

In our initial bibliographic research we did not happen upon the publication of Sasse et al. 2013, since our initial approach was more driven by the work of Diouf et al. 2013 where we attempted to use SOMs and give a reconstruction based on a more complex method. That method however was less applicable to the present study due to a significant lack of data and the quasi-equivalent discriminating factor of each variable, as seen in the ACP analysis we added to the manuscript (p 9 L 248-258). We therefore progressively developed on our own a methodology that is very similar to the one presented in the works of Sasse et al. 2013. We have, since, chosen to rename our paper to better reflect that this methodology already existed. We deem that the application of the method to this particular set of data, however, remains an important study with noteworthy results.

.) 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.

We have taken into account the complexity of the explanations provided, and simplified them (Section 2.4 p10-13).

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".

We agree with the reviewer, we add some details in the description of each parameter about the intercomparison between the product and with some in-situ data when these data are available (Section 2.3 p5-8).

-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?

We add a discussion about the parameters choice (p9 L 248-258).

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.)?

We added an explanation about the time predictor in the text (p8-9 L 238-247).

- 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 as- sumption that a linear regression is sufficient, but you provide no evidence for that assumption.

This comes from the theoretical part of the self-organizing maps (p10 L 270-273). In Dreyfus et al. 2005 in chapter 7, it is stated that, in the ending stages of the SOM training, when the interaction between neurons due to neighborhood constraints approaches zero, the algorithm can be assimilated to a K-means algorithm. In the same chapter it is stated that the K-means algorithm, in turn, can be considered to disperse its neurons to approximate the centers of multidimensional Gaussian distributions of equal standard deviation that generated the data that has been associated to them. Given a high discretization, i.e. enough neurons, we can assume a very localized variance centered around each neuron, that can be approximated by a MLR.

.) 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.

We added some comparison with a model which take place in the Baltic Sea (p 20, 21 L 449-470).

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?

We used a cross-validating and we details the results in the text (Section 3.2.3 p18 L 421-430).

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

The data used in this paper was found on CDIAC but the exactly same data are available in SOCAT no more data are available. LDEO database is the VOS ship, which are already used in our data.

.) 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.

We rewrote some parts of the paper to make the message clearer.

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

We agree with the reviewer, we added some results (section 3.2.3 p18-22).

.) 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.

We removed figure 8 and we kept the figure 11 which became figure 9.

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.

We added a "before" state for the figure 7 (which was expanded into figures 7 and 8) , we discussed this figure in the text (p14 L 340-350).

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

References:

Diouf, D., A. Niang, J. Brajard, M. Crepon and S. Thiria:Retrieving aerosol characteristics from satellite ocean color multi-spectral sensors using a neural-variational method. In Remote Sensing of Environment, Volume 130, 15 March 2013, Pages 74-86. 2013.

Dreyfus, G.: Neural networks : methodology and applications, Springer, Berlin ; New York, 2005.

Anonymous Referee #2

Received and published: 10 September 2014

Review of Parard et al. [2014]

"Remote sensing algorithm for sea surface CO2 in the Baltic Sea"

The study by Parard et al. presents and assesses two different methods of mapping pCO2 in the Baltic Sea exploiting in-situ and remote sensing data.

Overall evaluation: The study has a good potential to be published but at this point I can only recommend it for a re-submission. On the one hand the topic of the manuscript - Developing and testing methods of how to produce monthly pCO2 maps for the Baltic Sea - is very relevant for the Biogeosciences community. Furthermore, the presented methods are sophisticated and the overall quality of their assessment is good. On the other hand, the number of grammar and spelling mistakes contained in the manuscript makes a thorough evaluation of many critical paragraphs impossible. In numerous cases I cannot tell whether the explanations provided by authors are wrong or if it is simply because of the language that is being used. And quite frankly, a simple spell check on top of a thorough proof-reading by the authors would have caught almost all of the spelling mistakes which makes me wonder with how much care the manuscript was written and checked.

For the grammar and spelling mistakes the paper was review by a native english speaker.

A number of suggestions for a re-submitted version:

Validation of the method: All the parameters used (except for time) are subject to errors, e.g. remote sensing errors, errors and biases associated with the algorithms used to calculate e.g. NPP from remote sensing data, biases in the MLD model etc. All these errors affect the accuracy of the pCO2 maps. In fact, they affect the accuracy in the training process as well as in the application. So the overall mapping error will be larger than the number provided in the manuscript and the authors need to present an estimate for the additional error. (e.g. Friedrich et al. [2009, JGR] provide an example of how to estimate the contribution of remote sensing errors)

The method, as presented compensates for these types of errors since, if the errors are recurrent and constant in nature, the method will "learn" them and produce results based on these recurring types of errors.

The same is true for dealing with missing data. What is the effect of filling data gaps on the accuracy of the maps?

The completion of the missing data was a very important step in the reconstruction of the pCO2, since without it we would not have been able to apply an MLR and would have been forced to replace the missing value based on the average pCO2 value of the class to which the data would have been projected without being completed.

Flagging the maps is a good idea. The method used to derive the confidence level, however, assigns a lot of weight to areas where a lot of data are available. Figure 2 clearly shows that due to the scarcity of measurements the authors are forced to extrapolate from basically two lines of observations to the entire basin. Thus, in addition to the random 90%/5%/5% splitting of the data set we need to find out how large the mapping errors in the remote regions truly are. Presenting a thorough error-analysis for the pCO2 maps is as important as presenting the actual maps!

We added some information about this in the paper, with compare our result with a model and we divided the Baltic Sea in three main basin to make our comparisons (p21 L 449-470).

Methodology: The description of a SOM needs to be improved. The odd language and the grammar and spelling mistakes definitely contribute to the confusion but I also recommend checking how other studies have done it. E.g. Telszewski et al. [2009, BG] is a good example.

We checked other studies which are cited now in the paper. As noted before, in regards to the grammar and spelling mistakes, the paper was review by a native english speaker.

Figures: The figures need to be in substantially higher quality and resolution.Again, the study is very relevant for a wide range of readers and has good potential for publication. But in the current version a careful and sophisticated evaluation is impossible.

We have provided higher resolution figures.