

Review of the revised version of

*Parard et al. [2014]*

**Overall evaluation:**

The manuscript has improved significantly and I am confident that it can be published after some revisions. However, some of my major points of criticism regarding the validation of the method were not addressed correctly in the revision. Please see below.

**Accuracy of pCO<sub>2</sub> map and errors in remotely sensed data:**

In response to my initial review the authors state:

*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.*

This is exactly the problem: The errors are not recurrent and constant in nature! The errors are random in nature. The validation provided for the individual parameters in section 2.3 clearly shows that we are not just dealing with a simple offset (otherwise the correlations would all be 1.0). In the best case scenario, the random errors in the remote sensing data translate into random errors in the pCO<sub>2</sub> that cancel each other out when averaged (e.g. as done for Figure 14). But not even this is guaranteed. Definitely do these random errors affect the accuracy of the pCO<sub>2</sub>(x,y,t) estimate and we need to know to what degree. A statement like “*As for the chlorophyll, the bias is applied to all years so it does not affect the estimation of pCO<sub>2</sub>*” (line 165) completely misses the point and does not provide the reader with confidence that the authors were seeking a thorough error estimate.

The validation presented in section 2.3 provides some estimate of the magnitude of the error for each parameter. A random noise with the standard deviation given by the validation can be added to each parameter and the SOMLO method can then be repeated with noisy parameters. This would provide at least a first order estimate of the additional error coming from the remote sensing of parameters.

In a similar manner my questions regarding the effect of filling in data gaps was not addressed by the authors. They state:

*The completion of the missing data was a very important step in the reconstruction of the pCO<sub>2</sub>, 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 pCO<sub>2</sub> value of the class to which the data would have been projected without being completed.*

I understand that it was necessary. But my questions is how does it affect the accuracy of the method?

The validation presented in Figures 12+13 simply ignores errors implicitly introduced by the data and the data handling.

Repeating the statement from my first review:

Presenting a thorough error analysis for the pCO<sub>2</sub> maps is as important as presenting the actual maps!

The random 90% / 5% / 5% separation of the data set for the purpose of validation needs to be repeated an appropriate number of times to make sure that the error estimate is robust and not just the result of a lucky or unlucky splitting.

### **Figure quality:**

The quality of the figures is still not acceptable! At the normal print-out size I am able to distinguish between individual pixels with my naked eye. Given the simplicity of the plots it should be no problem to view them in great resolution on 300% zoom if they were true vector graphics.

For example, I assume there is text in the upper left corner of the lower panel of Figure 5 but it is absolutely not recognizable.

Since the readability of figures was claimed by both reviewers I do not understand why the authors do such a poor job in improving them.

### **Miscellaneous:**

line 41-42:

Please explain what you mean by "*the non-linearity of the emission and absorption system*"

line 90-91:

Please provide a reference.

Figure 1:

Do we need the black pixels over land?

line 99:

Maybe the location can be indicated in Figure 1.

line 126:

remove "from"

line 209-210

"density measurements" ?

line 248-257 and corresponding Figure 3:

The explanation provided is unclear. What is the distribution of the parameters in Figure 3 telling us?

line 291:

The explanation is very good but what is “*argmax(i)*”?

line 451:

remove one “)”

Figure 15:

aspect ratio is off (not to mention the resolution)

line 503:

“*This led to our having to complete ...*” ?

line 511-513:

I assume that spatial pCO<sub>2</sub> variability might also be larger in the Baltic Sea compared to open-ocean. This would also lead to a larger error.

References:

The correct spelling is “Lefèvre”

Please remove dispensable dots and letters in reference for Takahashi et al. [2009] and some other references.