

## ***Interactive comment on “Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean $p\text{CO}_2$ Mapping intercomparison (SOCOM)” by C. Rödenbeck et al.***

**C. Rödenbeck et al.**

christian.roedenbeck@bgc-jena.mpg.de

Received and published: 29 October 2015

Reply to Anonymous Referee 1

We would like to thank Anonymous Referee 1 for her/his thoughtful comments.

*Summary of paper. This paper explores the various new methods used to create full coverage open-ocean  $p\text{CO}_2$  maps for recent decades. Along with a general overview of the methods, including their similarities and differences, the authors delve into statistical comparisons involving mean annual values, interannual variability and seasonal amplitudes.*

C7147

*General Comments This is an excellent paper presenting worthy mapping methods with detailed analysis of seasonality and interannual variability within and between methods. The objective of fostering inter-method investigations and conversations is a valuable contribution as a whole. I think this work provides a valued resource to the ocean carbon community and has excellent scientific significance.*

Thank you very much for this rating.

*However, I feel it could improve on its aim to provide an uncertainty estimate or assessment of ocean carbon.*

We agree that the uncertainty assessment is still far from covered by our study. However, given that many similar intercomparison studies are also struggling with finding convincing ways to assess uncertainties, we feel that this topic needs a more detailed analysis, which will be covered in a subsequent study.

*One critique- it would be beneficial if the authors could provide a suggestion for an "optimal" method of those presented here. It is perhaps the case that the "optimal" method would vary for regional studies versus global studies, but guidance from the authors on this selection would prove valuable for the users of these maps who are less in-tune with their intricacies. I acknowledge that this work is not meant to rank methods but instead exploit the benefits of their complementarity, however some comments on optimal methods would be appreciated.*

We think that we cannot identify an "optimal" method here for the reasons given, but realize from this comment that there is a need for guidance to readers looking for sea-air  $\text{CO}_2$  flux products. In our opinion, analyses involving sea-air  $\text{CO}_2$  flux products should –if any possible– be done with several interpolation products, to test for robustness. However, the products should be selected according to suitable performance diagnostics. The presented "relative IAV mismatch" criterion represents a necessary condition for IAV applications, that could also be used in similar studies. Analogous "relative mismatch" criteria can also be defined and calculated for other time scales.

C7148

However, as discussed in the paper, it would be even better to use sufficient conditions, for example also incorporating synthetic-data reconstructions. Such sufficient conditions are not yet available for the SOCOM ensemble in a comparable way, but will be considered in forthcoming studies. Studies resulting in statistical quantities of the spatio-temporal fields, such as amplitudes of variation, correlation coefficients, etc., may also summarize the ensemble into averages of these quantities, weighted according to the above-mentioned performance diagnostics. We stress however that such weighted averages only make sense for scalar quantities. In contrast, any ensemble averaging (or medians, etc) of full spatio-temporal fields or time series would result in variations that are not self-consistent any more, and fit the data less good than individual products. Part of these remarks have been mentioned in the last paragraph of the conclusion section, but we realize that it has not been formulated clearly enough as a recommendation. We will rewrite and enlarge this paragraph accordingly.

*Specific Comments -The emphasis throughout is on the consistencies and differences between regressing and non-regressing methods (including the amplitude of the interannual variability). I think further information and discussion on the methods tying interannual variability to model simulations would be helpful.*

We agree that more information on model-based methods would have been interesting, but at present both of the model-based methods showed rather large relative IAV mismatches in most biomes, such that we cannot make more specific statements with regard to the features considered here.

*-SOCAT provides values of fCO<sub>2</sub> while LDEO provides values of pCO<sub>2</sub>. This paper discusses differences in pCO<sub>2</sub>, which leads me to believe that all values have been converted to pCO<sub>2</sub>. Was this transition done consistently between methods or is this an additional source of (albeit, I recognize small) variability?*

The comparison has been done in terms of pCO<sub>2</sub>. The conversion from fCO<sub>2</sub> to pCO<sub>2</sub> needed in SOCAT-based methods is considered part of the respective method. Some

C7149

methods use temperature and salinity dependent conversions, others use a constant factor. As already alluded to by the referee, the conversion is a constant ratio of (1-0.004) to very good approximation, such that any influences from this conversion are certainly only a very small contribution to the ensemble spread.

*-I appreciated the attention to detail with regard to spatial gap filling and other considerations made before intercomparisons between methods were discussed. I did wonder if a monthly climatological value was used for this pixel filling or if it was just an annual climatological value regardless of month.*

We used the values of the respective months from Takahashi et al climatology. We will add this clarification to Sect. 3.2.

*- What is the amount of data is shared between the LDEO and SOCAT datasets? Is this a large percentage, leading to possibly smaller differences than expected in the methods utilizing these separate databases?*

Both data sets indeed share a sizable fraction of data points. This may indeed be the reason why we do not find any systematic differences between methods with respect to using SOCAT or LDEO. This is confirmed by test runs with the Jena-MLS which is available both for SOCAT and LDEOv2013.

Most of the methods currently included in the project use SOCATv2 which holds a total of roughly 10.1 million data points, whereas those relying on the database of Takahashi et al mainly use data from version LDEOv1.0 or LDEOv2010 which includes roughly 3 or 5 million datapoints, respectively. Another difference between the databases is the extensive secondary quality control of SOCAT.

*Additionally, for those methods using the SOCAT dataset, was there consistency between methods in using the gridded product (either weighted or unweighted) versus using the individual observations to create the maps?*

Some methods use the original data, whereas e.g. all the regression methods rely on

C7150

the gridded data product. We did not notice any systematic influence with respect to this either. For all methods using grids equal as or coarser than the gridded SOCAT dataset (1 x 1 degree), we should actually not expect any such differences.

- *Figure 3 is exceptional and a great way to display the discussion.*

Thank you.

- *Figure 1A: "Monthly pCO<sub>2</sub> variations over 3 arbitrary years as estimated" could maybe be changed to Monthly pCO<sub>2</sub> variation over years 2003-2005 (chosen arbitrarily) as estimated ..."* Very valuable figure.

We will change the caption as suggested.

- *Explain figure A3 more in the caption. Is this the difference between the observations and what is shown in Fig A2? Have you accounted for "sampling" the maps at month and pixel that data exists for this comparison?*

Yes. We will add this information into the caption.

- *Figure A5- could you possibly improve by showing an area-weighted mean coverage or maybe the number of pixels (1x1 gridcells) with at least one observation per year vs the total number of gridcells in the biome?*

We agree that the chosen quantity has drawbacks, and had considered alternative options, but didn't see a clear advantage of any of them. Most importantly, the strength of the data constraint not only depends on the number of data points anyway but also on the way to use them (especially on coherence scales), such that there is no single way appropriate for all situations. However, the main purpose of Fig A5 is to show the temporal distribution of the data anyway, which is largely independent on what is shown in particular.

- *Footnote 5 gives a contradictory message than that stated earlier on the page (14063) in Lines 1-4 concerning the data density in the NA SPSS. This biome in fact has the*

C7151

*highest data density (especially in the 2000s). Sparse data cannot simply be blamed for methods that strongly differ in their seasonal cycle amplitude.*

Thanks for spotting this. Indeed the data coverage is quite good throughout NA SPSS. We will correct the footnote accordingly.

- *The sentence on pg 14064, line 18-19 stating "sampling biases pose the most prominent challenge to all the mapping methods" is a strong conclusion from this research and should be highlighted further throughout the paper and in the conclusions.*

More details concerning sampling biases have been addressed in the text on page 14062 lines 16-18. We will further add a remark into the Conclusions.

- *Appendix A gives a great concise description of the various methods, allowing a reader to not have to reference each paper individually- very nice inclusion.*

Thank you.

- *Technical Correction Page 14069, Line 25. Remove ", however," for clarity.*

We agree that "however" is misplaced, and will shift it to the beginning of the sentence.

---

Interactive comment on Biogeosciences Discuss., 12, 14049, 2015.

C7152