

Biogeosciences Discuss., author comment AC2
<https://doi.org/10.5194/bg-2021-304-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Christian Rödenbeck et al.

Author comment on "Data-based estimates of interannual sea–air CO₂ flux variations 1957–2020 and their relation to environmental drivers" by Christian Rödenbeck et al., Biogeosciences Discuss., <https://doi.org/10.5194/bg-2021-304-AC2>, 2021

We thank the anonymous reviewer for her/his detailed comments.
In the following, original comments are quoted in bold italics.

The authors use a well-established method (Rodenbeck et al 2013) to estimate the air-sea CO₂ flux from observations and expand it to cover the period 1957-2020 by adding a multi-linear regression approach to form a novel hybrid method. The manuscript is well structured and well written, but the methods (although very well explained) are difficult to grasp. This is a novel approach and I think this study presents a major step forward in estimating the marine CO₂ sink.

Thank you for your positive rating.

We will try to add a non-technical summary to the methods section to make it easier to understand.

For reference below, we would like to point out here that the reviewer's summary is only mentioning 1 out of our 2 equal-ranked topics: We do not only present a spatio-temporal flux product, but also interannual sensitivities (γ_i) and discuss them in terms of processes (see the title as well as abstract, Introduction, Results, and Conclusion, all mentioning both these topics side by side).

I do however have one significant issue with the current study and that is the representation of uncertainties. Many recent studies (Bushinsky et al 2019, Gloege et al 2021, Hauck et al 2020, Fay et al 2021, and others) have focused on data limitation and uncertainties, and this should be the standard.

In addition to our more detailed responses below, we would like to make 2 general remarks about this concern by the reviewer:

First, the reviewer's comment only addresses the hybrid mapping seen as a numerical flux product, even though that is just part of the focus of the study (see above).

Second, an assessment of the impact of data density on the 3-dimensional flux field has already been presented for the original CarboScope product in the cited reference Rödenbeck et al. (2014) using the "Reduction of Uncertainty" metric. Regarding the recent decades (SOCAT data period) we do not claim anywhere in the manuscript that the hybrid

mapping would be able to improve upon these previous estimates in spatial areas without pCO₂ data coverage. Concerning the extrapolation into the early decades, the manuscript clearly states that the interannual variations are likely underestimated and that the secular trend is just coming straight from the OCIM prior.

We will add a note into the revised manuscript pointing to the "Reduction of Uncertainty" assessment in Rödenbeck et al. (2014).

While I do believe the authors have done a great job in testing their method using a large variety of sensitivity runs as well as a tough test where 5-year periods are excluded (btw. the same test has been conducted in Landschützer et al 2016 – supporting information, and should be cited).

Historically, this test was originally suggested by C.R. in an e-mail to the SOCOM community on 2016-01-08 (back then in a variant "CrossVal5yr0"/"CrossVal5yr1" using alternating 5-year periods with and without pCO₂ data). Peter Landschützer kindly did the suggested test runs with the SOM-FFN mapping method. Lacking further participation, the envisaged SOCOM community paper was never written, but Peter Landschützer made at least use of his runs in his 2016 paper. Given this historical background, the citation suggested by the reviewer does not seem appropriate.

My concerns however are the following:

Although sensitivity runs are performed and data omission tests are performed, there is no serious indication of uncertainty of the product. As stated by the authors, one application of this product is to add it to the GCB estimate of the full historical period (page 7 lines 17-20), but how can we have enough confidence in such a reconstruction without thoroughly estimating the uncertainties in the annual mean flux and the presented trends (see also points below)?

On the one hand (in addition to the general response given above), the study does give uncertainty ranges for GCB-relevant traits (mean, variability, trend) based on the spread across our suite of uncertainty runs (where we do state that this may not comprise the complete uncertainty in the result). On the other hand, the GCB study does not make use of any uncertainty ranges around individual products anyway, but rather does its own uncertainty assessment based on the spread across the ensemble of products from the various groups.

To provide a more direct example: On page 13 line 22, the authors report a trend of 0.002 – 0.005 PgC/yr/yr – how can they be confident that such a trend is significant?

The study quantifies a range of the wind-related trend (with the range indicating uncertainties!) and compares it to the total trend. We do not actually make any statements that would depend on the wind-related trend being different from zero. Thus, what type of significance does the reviewer have in mind?

Furthermore, none of the line plots in the presented figures include error estimates, which causes the impression that there are no uncertainties.

During manuscript preparation, we had tested presenting the range of uncertainty estimates also in the line plots. The problem, however, is that some of the tests (e.g., GasexLow, GasexHigh) strongly affect the mean flux, thereby shifting the lines vertically and thus disguising any information about the spread in interannual variations anyway (remember that IAV, not the mean, is the actual focus of the study). We had therefore decided to defer the presentation of the range to the bar figures (Figs 8-10) where the

effects on mean and variability are separated from each other.

In the revised manuscript, we will additionally present line plots analogous to Fig 3 but for interannual anomalies in the supplement. Plotting anomalies allows to present the range of uncertainty estimates without the above-mentioned problem.

The only exception is figure 10, however, error bars only relate to the linear slope uncertainty and not the method uncertainty.

The information on uncertainty is provided by the range of results from the various uncertainty cases.

Furthermore, what is now actually the difference (if any) between this method and the GCB models over the full historical period?

If "the GCB models" refers to the global ocean biogeochemical models (GOBMs) collated in Friedlingstein et al (2020) and used here for comparison in Figs 8-10 (mint green), there is quite a fundamental difference: GOBMs comprehensively simulate the time evolution of the state of 3D oceanic biogeochemical variables based on natural laws and detailed process parameterizations. The present mapping scheme is primarily driven by the pCO₂ observations, making use of some explanatory variables as well as some simple parameterizations of mixed-layer dynamics as needed to relate pCO₂ observations and flux fields.

As it was mentioned in the introduction, I got curious but such an analysis was never presented (this could also serve as validation).

No, it would not be appropriate to consider GOBM simulations as a validation of observation-based products. Validation can only be made against independent observational data. Unfortunately, this is not an available option here (see the Introduction).

My biggest concern stems from the lack of historical data (see e.g. Bakker et al 2016 – figure 2). Any estimate before the 1990's (probably even before 2000) that is based on SOCAT should be viewed with caution.

We fully agree that the temporal extrapolation needs to be viewed at with caution. However we feel that this has been mentioned clearly at various places in the manuscript.

There is no serious attempt here to quantify or at least thoroughly discuss such missing data, maybe with the exception of the Southern Ocean where this is explicitly mentioned. The authors state that the analysis of another method in the Southern Ocean (page 17 lines 3-5) revealed an overestimation of the decadal variability amplitude, but what about this study? Is it in a similar range? At least by adding more regions in Figure 8 one could get an impression.

See below about assessing the method with synthetic-data tests.

To provide an example how one could test the results, the authors could use SOCAT data of lower quality flag (assuming that the measurement error may be small compared to the interpolation error),

Unfortunately, SOCAT data of lower quality flag are far from covering the entire ocean as well. In the data-void ocean areas accounting for much of the interpolation error, such data are thus not available for validation either.

or by subsampling and reconstructing a hindcast model run (similar to Gloege et al 2021), where a known truth exists.

Indeed, we have already run tests with synthetic data very analogous to those presented in the cited study by Gloege et al. (2021). Analysing and presenting these synthetic-data runs, however, needs a separate paper and cannot be added in passing into the present paper.

Furthermore, a study by Bushinsky et al 2019, using the original version of this method revealed that additional data (in the Southern Ocean) caused a substantial change in the air-sea flux. There is quite some debate about the reliability of the added float data in this study, nevertheless, when considering the historical period, it shows that additional data have the power to substantially change the flux estimate in data sparse regions, which should at least be further discussed.

We can only repeat that we agree to the concerns of the reviewer, but that we do mention this problem already, while a more quantitative assessment is either impossible due to the lack of independent data or needing a separate paper.

I am also puzzled by Figure 5 (but this may be a misunderstanding on my side), but does this figure not suggest, that the multi linear regression is more robust in reconstruction periods without any data, whereas the hybrid mapping is not as robust?. Would a multi-linear regression not be more robust then, considering that only a tiny fraction of the ocean is actually covered by observations?

Indeed there seems to be a misunderstanding. Regression and hybrid mapping are not "alternative methods" we could choose between. Rather, they have different purposes (sensitivities vs. flux variations), and the regression is a part of the hybrid mapping. The discrepancy seen in the right panel of Fig 5 does not indicate a lack of robustness of the the hybrid mapping, but illustrates that the regression is not able to represent the full amplitude of variations (as discussed in the manuscript).

I was quite surprised of the authors statement in the main text that chlorophyll did not make a big difference. Figure S7 suggests that chlorophyll makes a substantial difference in the Southern Ocean, maybe in line with Hauck et al 2013?

As mentioned in the text, we expect the additional spike to be an artifact, not a meaningful signal. By the way, also other pCO₂ regression methods in the literature indicate that chlorophyll is not a very helpful explanatory variable.

While trends and temporal changes are investigated, there is little discussion about spatial features and (again) the uncertainty spatially. A strong focus is set (understandably) on the tropical Pacific Ocean, but there are other regions that are well observed (like the North Atlantic or the North Pacific Oceans) that could serve as a benchmark test how well the method reconstructs the air-sea CO₂ flux in space. In the end, the authors present a 3-dimensional product (with increased resolution), hence a comparison in space, e.g. with other methods or direct observations from SOCAT or model estimates should be considered to increase the confidence.

We explicitly state from the beginning (even already in the title) that this is a paper on interannual variability. While spatial variability may indeed be interesting as well, a single paper cannot consider every aspect of a 3D field.

The reviewer is very welcome to analyse spatial signals in the estimated flux field (which is openly available as given in the manuscript). We would be happy to collaborate on that.

Minor points:

.) page 2 lines 14-17: What about ocean inverse estimates that rely on repeat hydrography measurements and an ocean circulation model?

Indeed, ocean-interior DIC data have been used to estimate sea-air CO₂ fluxes. However, while DIC in the ocean interior can constrain the mean flux, it cannot constrain the year-to-year variability focused on here.

.) page 2 line 26: SOCAT provides fCO₂ not pCO₂

We agree and will clarify this in the revised manuscript.

.) page 3 line 11: I disagree – a response function analysis very well reveals the individual relationships, even in neural networks with many layers and many degrees of freedom

Of course, by performing additional retrospective analyses on an already-trained neural network, it would be possible to reconstruct input-output relationships. However, linear regression yields this information in the first place (provided the relationships are sufficiently linear).

By the way, we are not aware of any study in the literature that explicitly presented relationships between pCO₂ and the drivers of a neural network. For example, Landschützer et al., Science 349, 1221-1224, doi:10.1126/science.aab2620 (2015) rather used parameterized pCO₂=pCO₂(SST) relationships to discuss process contributions.

.) page 6 lines 19-21: Salinity may be an important regression variable, particularly in the polar regions

Thank you for pointing this out. In past tests, we had indeed used Sea Surface Salinity (SSS) as additional explanatory variable. We have re-done such runs now (using SSS and dSSS/dt) and will add them into the supplement of the revised manuscript. They do not change the result much, however. A hesitation about these runs comes from potential measurement problems (e.g., by fouling).

.) page 7 and following: I am not so sure how much these experiments add to exploring the robustness. In Figure 9 it seems that only the gas exchange experiments make a notable difference when it comes to the mean flux analysis.

We are surprised by this comment as the reviewer expressed before the opinion that uncertainties had not been sufficiently explored. The various test cases do impact specific aspects of the result, especially the variability. Moreover, even if a given uncertain set-up element turns out to have little impact, we consider this an interesting piece of information, too.

.) Figure 4: the black dots are difficult to see

We agree that this is not optimal. However, for consistency with previous papers we had decided to keep blue for the base result and black for data. As the blue and black dots are mostly co-located anyway, there isn't actually any information loss from the poor visibility.