Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-304-RC1, 2021

Review of Rodenbeck et al „Data-based estimates of interannual sea-air CO2 flux variations 1957-2020 and their relation to environmental drivers

Anonymous Referee #1

Referee comment on "Data-based estimates of interannual sea–air CO_2 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-RC1, 2021

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.

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. 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 Landschutzer et al 2016 – supporting information, and should be cited). 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)? 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? Furthermore, none of the line plots in the presented figures include error estimates, which causes the impression that there are no uncertainties. The only exception is figure 10, however, error bars only relate to the

linear slope uncertainty and not the method uncertainty. Furthermore, what is now actually the difference (if any) between this method and the GCB models over the full historical period? As it was mentioned in the introduction, I got curious but such an analysis was never presented (this could also serve as validation).

- 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 (probaly even before 2000) that is based on SOCAT should be viewed with caution. 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. 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), or by subsampling and reconstructing a hindcast model run (similar to Gloege et al 2021), where a known truth exists.
- 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.
- 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?
- 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?
- 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-dimensionsional 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.

Minor points:

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

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

.) 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

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

.) 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.

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