Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-207-RC1, 2021

Comment on bg-2021-207

Anonymous Referee #1

Referee comment on "A seamless ensemble-based reconstruction of surface ocean pCO_2 and air–sea CO_2 fluxes over the global coastal and open oceans" by Thi Tuyet Trang Chau et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-207-RC1, 2021

The authors use a neural network model to generate a pCO_2 product for the global ocean using the SOCAT data, and combine these pCO_2 estimates with a wind speed product to compute the CO_2 flux. The ensemble model results compare well overall to the observations, and the carbon flux estimates are in-line with the literature. My main comments concern how novel these results are compared to the extensive literature on the topic, and the interpretation of some of the model statistics.

There is a lot of previous literature using spatially and temporally sparse observations of surface pCO_2 to generate global data products and provide estimates of ocean carbon uptake, some of which use very similar methods to those in this present manuscript. The authors cite this previous literature, but there's very little discussion of it. Consequently, I found it difficult to interpret how the present authors' methods and results are novel and differed from these previous studies. The motivation appears to be in lines 41-46, however, I don't follow how the previous literature did not incorporate "space-time varying uncertainty estimates"? It would also appear that the incorporation of the coasts is relatively new, though the authors then cite a few recent studies and declare that it's a closed gap? I suggest that the introduction needs to contain a much clearer description of how the methods used here compare to previous studies, and what is new about this analysis.

I think the methods section is missing a few key details that will help support this manuscript. First, it would be helpful for the authors to explain how to interpret and compare the RMSD and r^2 values for each region. The reason being, that these values are listed for each ocean region, but it's a little unclear what differences in these values between regions is saying about the model estimate. For example, I was surprised by how low the RMSD value for the Southern Ocean is (slightly lower than the global mean), despite the somewhat limited observational data and well documented, substantial inter-annual variability. However, the Southern Ocean does have a lower r^2 value, which the authors seem to attach a greater weight to in their interpretation. Second, I'm a little confused by equation (2). Why is the equation for the mean squared deviation (MSD) shown when it's the root mean squared deviation (RMSD) which is calculated throughout the manuscript? Also, the text refers back to this equation for the definition of the σ misfit, but this definition is itself within the MSD equation and is not clearly labeled on its

own. Lastly, I think the description of the wind speed product used should be included in the main text rather than the supplementary, considering that this will have a large impact on the overall flux numbers (which the authors do highlight in the results).

I suggest re-working the 2nd paragraph of the abstract. This paragraph currently reads like a laundry list of different regions and where they fall in terms of largest total source/sink, largest flux density source/sink, along with coastal and open ocean qualifiers. This many iterations of "X is the greatest …" makes it difficult to follow-along and is not particularly interesting (e.g. the equatorial Pacific as the strongest source of carbon to the atmosphere is not a surprising result). Instead, highlight some of the other key findings, like the increase in ocean carbon uptake over the 1985-2019 timeframe (right now the mean is just listed, but the change is highlighted in the conclusion).

Other comments

Lines 37-40: Should all the manuscripts be separated with a comma rather than a semicolon? And why is the Rödenbeck et al., 2015 manuscript specifically highlighted as "other mapping methods"?

Line 139: How is the variability in "analytical equipment" accounted for here?

Figure 2: I suggest directly labeling each region in the figure with the abbreviated label (i.e. SpA for subpolar Atlantic) for clarity.

Figure 5 and 8: The tick marks in the colorbar for these figures are relatively large and look like a negative sign, I'd suggest making them much smaller.