

Biogeosciences Discuss., referee comment RC3  
<https://doi.org/10.5194/bg-2021-207-RC3>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on bg-2021-207

Meike Becker (Referee)

---

Referee comment on "A seamless ensemble-based reconstruction of surface ocean  $p\text{CO}_2$  and air-sea  $\text{CO}_2$  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-RC3>, 2021

---

The authors present an estimate of global air-sea  $\text{CO}_2$  fluxes based on interpolating gridded SOCAT  $p\text{CO}_2$  data. They use an ensemble of 100 feed-forward neural network models (FFNN) and sea surface height, sea surface temperature, sea surface salinity, mixed layer depth, chlorophyll-a, atmospheric mole fractions, a  $p\text{CO}_2$  climatology and position data as drivers. They present an uncertainty analysis based on their ensemble spread as a semi-independent parameter, which is better than many available air-sea flux products. However, there are a few things that should be improved.

One point that needs improvement is the description of the driving data, where some important information is missing. The driving data that were used are not available for the full period for which the authors present flux maps. How did you deal with that? Did you use a climatology for CHL, SST, MLD etc. before the early/mid 90s? If so, what was this based on? This information is crucial for interpreting interannual variability prior to the mid-90s.

Another thing that I want to point out, is the inconsistent and partly misleading use of the terms 'observations', 'sample' and 'data'. The authors base their product on a gridded version of the SOCAT data set (monthly,  $1^\circ \times 1^\circ$ ). In order to avoid confusion, the term 'observations' should be reserved for data that has been retrieved from field work, in the case the original  $p\text{CO}_2$  measurements in the SOCAT database. The gridded version contains monthly,  $1^\circ \times 1^\circ$  averages of these  $p\text{CO}_2$  measurements. When the authors write about 'X observations' in a certain region, they actually mean 'grid boxes with observations'. Please make sure that this becomes clearer. In line 195 for example, the authors write '50 to 220 samples per year'. Here the authors should specify that they mean 'grid boxes with data' as the reader easily can assume that there were only 50-220  $p\text{CO}_2$  observations every year.

I also want to comment on Figure S1. Here the authors show the coverage of the gridded SOCAT product and its variability where they mention ' $p\text{CO}_2$  individuals'. I don't understand if this means the original SOCAT  $p\text{CO}_2$  observations (i.e. a measure of how well the grid box mean represents the actual conditions), or the  $p\text{CO}_2$  of the gridded

version (showing the variability within the gridded product).

Please go through your manuscript with these comments in mind and make sure, that different terms are consistently used and that it is clearly stated what you mean.

I understand that the authors used the subocean divisions from RECCAP 1. This of course increases the comparability to the results of RECCAP 1, but also this makes the results difficult to interpret. Using a biome scheme such as used in RECCAP 2 (e.g. after Fay and McKinley (2014)) would have led to a clearer separation of regions with similar characteristics, and thus increased the interpretability. I also miss a discussion of how this product performs in comparison to other global air-sea CO<sub>2</sub> flux products.

Minor comments:

L 40: pCO<sub>2</sub> was not introduced as an abbreviation

L 59: Tr is not described

L 86: change to: p is the number of grid cells with observations

Figure 3: The yellow bars in panel c) are very difficult to read, especially the first one. Additionally change STD to s. Go through the manuscript and make sure, that you use consistent terminology.

L 140: add: temporal offset from the cell center. In many regions this will be the dominant one, especially during the productive season.

Figure 4a/b: You show the number of observation (or grid cells with observations) per year. Please change that.

L 182: Be aware that Laruelle et al. 2107 and Landschützer et al. 2020 are climatologies.

L 307: Please round the uncertainties to 2 significant digits (or less if it seems unrealistically low) and the measured value to the same number of digits, for example 2.336 +/- 0.104 to 2.34 +/- 0.10. Please do so for all uncertainties in the manuscript.

L 328/330: Please correct this. Primary production and respiration have usually only a very small influence on alkalinity (if we neglect anaerobic remineralization processes for the moment): primary production increases alkalinity, while remineralization processes reduce alkalinity

L 332: Another important influence factor in coastal regions is the inflow of terrestrial POC, e.g. in the southern North Sea, leading to the release of CO<sub>2</sub> to the atmosphere.

L 376-377: Are these really the dominant factors? After your argumentation for why the open ocean region is neutral (vertical convection brings up old, DIC rich water which balances the influx during summer) I would expect the absence of this deep mixing in coastal, shallow regions to be one of the major reasons why the coastal regions are a larger sink than the open ocean.

Figure 9: To be honest, I can't really see from this figure that it covaries with the ENSO mode. As I see it the flux increases equally often during La Niña as during El Niño. It would be more interesting to see a comparison of the interannual variability with other air-sea flux products.