

Earth Syst. Sci. Data Discuss., referee comment RC4
<https://doi.org/10.5194/essd-2022-71-RC4>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-71**

Anonymous Referee #3

Referee comment on "A new estimate of oceanic CO₂ fluxes by machine learning reveals the impact of CO₂ trends in different methods" by Jiye Zeng et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-71-RC4>, 2022

Review for 'A new estimate of oceanic CO₂ fluxes by machine learning reveals the

impact of CO₂ trends in different methods' by Jiye Zeng *et al*

The paper aims to investigate issues related to interpolating sparse in situ ocean pCO₂ data to allow complete fields to be generated and then global assessments of air-sea gas fluxes and the net oceanic sink to be assessed. The study investigates the impact of assumed or specified trends that exist within some of these interpolation methods towards understanding what impact these assumed trends can have on the resultant oceanic net sink estimate (and in particular the results at the ends of time series). The analysis includes producing a novel reconstruction method which produces an integrated net sink value which is markedly different in magnitude to all of the other datasets analysed.

I would recommend that the paper is significantly revised, including a modified title before then being reviewed again. The reasons for this recommendation are given below.

My main comments are:

- The paper seems to be focussed on a trend analysis, and identifying the impact of assumed trends within different datasets, rather than providing 'a new estimate of oceanic CO₂ fluxes' and determining the absolute oceanic sink. The methods identify how the rate of ocean absorption has changes and how each of the methods that are also compared (within the inter-comparison) does or does not capture similar trends. As a result the paper title seems a bit misleading as the title should really explain that's its focussed on the analysis of the trends and the rate of ocean absorption of carbon, and not the oceanic sink strength and magnitude (and not on presenting a new estimate of the oceanic sink). The focus in the title and abstract also seems to be a bit out of scope for the journal ESSD as I thought that the journal is focussed on providing

datasets for wide use, where the focus here seems to be more focussed on novel science or result-focussed work.

- A detailed methods section seems to be missing from the paper. This lack of information makes it hard to evaluate the approach presented and the results. The limited amount of information on methods (which appear in the section called 'data') lack detail and need significantly expanding. In particular the methods used to calculate the air-sea fluxes and the derivation of each of the parameters needed as input to the flux calculation are missing (eg do the NOAAs Marine Boundary Layer Reference provide $p\text{CO}_{2\text{atmosphere}}$ data, or do they provide $X\text{CO}_2$ data that you use to calculate $p\text{CO}_{2\text{atmosphere}}$?) The calculation of the net integrated air-sea fluxes is also missing (eg how were the values in Figure 6 calculated, what land and ocean masks were used, how were mixed pixels containing land and ocean handled, was surface area calculated using assuming a sphere or an ellipsoid?). Providing a separate section on methods that detail all of the methods used would help the reader. I could not follow the methods in section 2.3 Rate extraction. A more detailed overview is really needed in the paper for this part of the methods (The authors rely on the reviewer/reader reading another paper to gain a basic understanding).
- The difference between air-sea flux data products is likely to be at least partly (maybe mostly) driven by the differences in input data for calculating the air-sea gas fluxes (eg differences in wind speed data, or sea surface temperature, salinity etc) and this could also be one of the causes for the different identified trends. This issue has been overlooked and should at least be discussed and/or tested for. It may be possible to partially remove it, eg by using a common method to calculate the gas fluxes, a common set of wind data, common atmospheric data and then just the $p\text{CO}_{2\text{water}}$ fields from each of the datasets within the inter-comparison. This would allow the authors to isolate the impact of different wind and atmospheric forcing data.
- All of the spatially complete datasets used within the inter-comparison form submissions to the Global Carbon Budget assessments. Within the GCB assessments these datasets are all assessed using SOCAT data. If the authors want to keep the 'new estimate of oceanic CO_2 fluxes' as a key finding and focus of the paper, then their new $p\text{CO}_2$ fields should be evaluated using the community SOCAT database eg train on a subset of the dataset, test against the remainder (eg RMSD and bias), or at the very least evaluate their complete $p\text{CO}_2$ dataset using the data that were used to train the approach. At the moment the paper presents an inter-comparison where the new approach is vastly different from all other data included in the inter-comparison but no justification or evaluation as to why these new data are useful is given; only that that they are different and new. Can the authors provide some sort of statistical assessment or evaluation of the output data ($p\text{CO}_2$ fields and/or gas fluxes) to provide the reader with some confidence that the new results are valid?

General comments

The paper would benefit from the help of a copy editor help improve the main text.