

Biogeosciences Discuss., referee comment RC2  
<https://doi.org/10.5194/bg-2021-207-RC2>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on bg-2021-207

Anonymous Referee #2

---

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-RC2>, 2021

---

At present, there are many data products in marine physics, such as temperature and salinity products, but there are few data products in marine chemistry. I support the publication of more marine chemistry data products.

The author reconstructed surface ocean  $p\text{CO}_2$  based on FFNN with region divided by latitudes and similar predictors with previous researches was used, which is not novel. The reconstruction of  $p\text{CO}_2$  and sea-air  $\text{CO}_2$  flux over global coastal oceans are interesting works but the author needs to do much more works on the validation of coastal results. Because a standard deviation of  $41.79 \mu\text{atm}$  between  $p\text{CO}_2$  results and SOCAT observations possibly leads to opposite results in the estimate of coastal  $\text{CO}_2$  flux. In general, this manuscript needs to be further improved for the following reasons:

1) The validation of  $p\text{CO}_2$  reconstruction was completely based on SOCAT dataset itself. The author did not compare the result with any independent observations from time series stations, such as HOT, BAT and ESTOC station. Especially in the coastal ocean, the author will find their  $p\text{CO}_2$  product surprisingly deviated from the independent observations from time series stations if a comparison was carried out. In addition, reconstruction of  $p\text{CO}_2$  based on machine learning methods has been attempt in several previous researches, such as Landschützer 2016, Denvil-Sommer et al. 2019, and comparison with these existing results was not found in the manuscript.

2) The CHL data used was only from 1992 to 2019, and was not available in the Arctic and the Southern Ocean in winter, the details about how the reconstruction was carried out when CHL was not available should be declared clearly in the method section.

3) The subskin temperature correction (Watson et al. 2020) should be considered in the estimate of sea-air  $\text{CO}_2$  flux.

4) The author should reconsider the topic of this manuscript. If the author want focus on the CO<sub>2</sub> flux of global open oceans, additional work was necessary rather than only discussing spatial distribution or interannual variability, because the reconstruction method in this manuscript and the results was not novel. If the author want focus on the CO<sub>2</sub> flux of global coastal oceans, which was still a research gap, much more works are needed to make the result convince.

5) Line 76: "An ensemble of 100 FFNNs was used to reconstruct monthly pCO<sub>2</sub> fields.....", How are these 100 models built? Why did you do that? How are the results of 100 models selected?

6) Line 84-85: "The random extraction and the FFNN training were repeated 100 times so that 100 versions of the monthly FFNNs have been obtained", Why is it the "100 times"? How is the "100 times" determined? Does it converge after 100 iterations?

7) Table 2, What is the meaning of two numbers in the rightmost column of Table 2, for example:  $0.07 \pm 0.04$ ,  $0.30 \pm 0.13$

8) Line 444-446: "The global open ocean uptake obtained in this study of  $1.344 \pm 0.111$  PgCyr<sup>-1</sup> lies between the observation based net sink estimate by Wanninkhof et al. (2013) ( $1.18$  PgCyr<sup>-1</sup>) and the global sum of regional best estimates given in Table 2 ( $1.8$  PgCyr<sup>-1</sup>)". In table 2, I can't find the value of  $1.8$  PgCyr<sup>-1</sup>

9) Line 462-463: The discrepancy is possibly due to an overestimation of Arctic pCO<sub>2</sub> by the CMEMS-LSCE-FFNN (see in Sect. 3.1.2) and to the lack of estimates over a large portion of the seasonally sea-ice covered regions. This sentence means that the data in the Arctic are not accurate at present. So the data in the Arctic is not suitable for use at present.