Reconstruction of global surface ocean pCO2 using region-specific predictors based on a stepwise FFNN regression algorithm.

The authors developed a new ML approach to reconstruct global surface ocean pCO2 that considers an impact of different predictors in different ocean regions. Based on Self-Organizing Map method authors defined 11 biogeochemical provinces. A stepwise FFNN regression algorithm was applied to each of these provinces to establish a set of predictors that are highly responsible for pCO2 variability in considered province. Based on selected predictors and analysis of FFNN size (number of neurons) a monthly 1°x1° surface ocean pCO2 product from 1992 to 2019 was constructed. The results show a good agreement with validation data and independent observations.

I found this work well-organized and easy to read. It was interesting to see new predictors (phosphate, nitrate, silicate, dissolved oxygen) and their role in pCO2 variability. The authors presented important results for the Indian Ocean where due to the lack of observations different methods show their disagreements.

Below, I listed several points that need to be clarified before publication.

Comments:
- Boundaries between provinces. In the text we can find "To obtain a smoother distribution, we defined that the grid within 5 1x1 grids of province borders belong to all provinces adjacent to the nearest province border. Samples in these grids were involved in the FFNN training process of multiple provinces, but only counted once in the validation." Please could you clarify what you mean by "only counted once in the validation"? Is only an output from one province used in the validation? If yes, how do you chose a province from which you take an output?
- Independent observations. Please could you provide geographical positions and period of stations used as independent observations?
- Set of selected predictors. In table 3 authors presented two sets for most of the regions that depends on the availability CHL-a data. Please could you present more explicitly that the final product is built on two FFNNs, one trained for the period 1992-2001 based on one predictors set and another – for 2002-2019 based on the second predictors set?
- More explicit figures' captions. Please provide more explicit figures' captions, period of presented results, or results averaged over xxxx-xxxx, what are horizontal lines in Fig.6b?
- Not correct conclusion. On page 15 lines 375-379 authors concluded that the difference between FNN1 and FFNN3 is relatively small, because predictors used in FFNN1 and FFNN3 were related to main drivers of pCO2, such as CHL-a, xCO2 and MLD. However, same drivers are used in FFNN2. Thus, it cannot explain why FFNN2 shows higher differences with observations.
- On page 18 line 430 authors said that the pattern of reconstructed pCO2 climatology was close to SOCAT in the Indian Ocean. I would say that it is not so close to mention it in this sentence.

Page 10 lines 298-300: For better structure of paragraph the sentence "In the province P1 located in the Arctic, the silicate concentration and temperature were considered as the most crucial predictor of pCO2." could be moved at the end of paragraph where authors mentioned the phosphate, nitrate, silicate, etc.

Page 13 lines 336-337: The sentence about results in the Indian Ocean can be removed if you put the Indian Ocean in the previous sentence, or please add "Also" at the beginning of the sentence dedicated to results in the Indian Ocean.

Page 16 lines 390-392: Please could you reformulate this sentence ("The interannual variability and seasonal pattern..") as it is difficult to read?

Page 18 lines 432-436: two sentence can be combined: "Compared with previous climatology product (Landschuster et al., 2020), the global distribution pattern of surface ocean pCO2 was basically well consistent: inconsistent spatial distribution also existed in the Arctic and parts of the Southern Ocean near the Antarctic continent."

Typo:
Page 2 line 41: "**s**urface ocean pCO2" should be replaced by "**S**urface ocean pCO2".
Page 9 line 261: "**v**alidation group" should be replaced by "**V**alidation group".
Page 13 line 332: "Based the K-fold" should be replaced by "based **on** the K-fold".
Page 20 line 459: "based improved FFNN size" should be replaced by "based **on** improved FFNN size".