Review 2

General comments:
The stepwise FFNN looks like a innovative new approach to enhance the widely popular SOM-FFNN method. The stepwise method is tested using a very comprehensive list of predictors used elsewhere in the literature. The method building component looks thorough, congrats. The prediction of pCO2 based on region-specific predicators selected by the stepwise FFNN algorithm will be a valuable tool when moving to higher resolution, inside regional studies, or getting closer to shore. There are a number of grammatical errors that will need to be cleaned up by the authors or the journal team.

Response: Thank you very much for your appreciation and very valuable suggestions to improve the manuscript!

Specific comments:
1 Introduction
Line 66-82: Appreciate the list of previous works and predictor data used by each, provides justification for use in the stepwise FFNN. Would like to see one more sentence relating use of different predictors leading to varying marine sink estimates.

Response: Thanks for you suggestion. In previous researches, not only different predictors, but also different methods were used. Thus, the differences in estimate of marine carbon sink between previous researches were not only caused by use of different predictors. In addition, there is almost no such research that focusing on the influence of predictor differences on marine sink estimate.

2 Methodology
2.1 Data
Line 106-122: Are all these products retrieved at the same resolution? Are they upscaled or downscaled at all to your needs?

Response: Most of these products were retrieved at 1° × 1° resolution. Some products retrieved at higher resolution were downscaled to 1° × 1° resolution. This description has been added at the end of the 2.1 Data section.

Line 122: "In addition, 8 parameters…." Thanks for listing after this sentence. Which previous research used these as predictors in observation-based pCO2 estimates? List and provide citations like in the introduction. Or is the inclusion of these predictors' novel? If so, highlight that.

Response: These listing 8 parameters have not been used as predictors in observation-based pCO2 estimates in previous researches yet, but nutrients and dissolved oxygen have been used as predictors in observation-based estimates of total alkalinity and DIC. The citation has been added.

2.2 Biogeochemical provinces defined by the Self-Organizing Map
Line 134: These SOM predictors exclude most of the FFNN predictors discussed in the introduction. Was there a reason why "biological" predictors (i.e., nutrients and oxygen) are weighted so heavily in the SOM selection? Were more physical predictors (i.e., mixed layer depth, etc.) used in SOM testing to optimize provinces? Or similar to previous work (Landschützer), using published pCO2 climatology as a predictor to determine provinces?

Response: Thanks for noting this problem. Actually, we also used mixed layer depth, sea surface height and pCO2 climatology from Landschützer, 2020, but mistakenly lost in the text. The description about predictors has been corrected now.

Line 135: Just out of curiosity, did the configuration (3-by-4 size) make much of a difference to SOM province distribution?

Response: In the early work, 4-by-4 or 4-by-5 size were also attempted. Increasing size led to appearance of small provinces inside main provinces, but the distributions of main provinces were similar. To simplify the SOM boundary issues, we choose the 3-by-4 size with less provinces.

Line 141: The 200 m depth boundary is fairly close to shore. Is this a commonly used open oceanic / coastal ocean boundary? If so citations from other studies here.

Response: It is not a commonly used boundary. In previous researches focusing on coastal pCO2 reconstruction, the boundary was defined as 1000m depth or 300 km offshore. We defined the boundary as 200m depth because the SOCAT samples with high predicting error were mainly located in areas shallower than 200m.

Line 144: Unique way to address the SOM boundary issue. Cool.

Response: Thank you for your appreciation.

2.3 Stepwise FFNN algorithm
Line 152-163: Clarify. Was the mean absolute error used for the internal

MATLAB neural network performance loss function (in the training targets and validation targets steps used to end training), also / or as a means for evaluating the FFNN output pCO2 product to withheld data?

Response: The MAE was used for performance loss function, and also in the validation of pCO2 product using a K-fold cross validation method.

Line 172: "…referred to as indicators pool (Start in Fig. 1),…" Keep coming back to this Figure throughout if you can. Makes it easier to read and connect to the Figure.

Response: Thank you for your suggestion. More annotations have been added in the description.

Figure 1: More sub boxes (dotted lines), connected to text could also make it easier to follow. Steps between loop 1 and loop 2, steps between loop2 and end.

Response: Thank you for your suggestion. More sub boxes have been added in the Figure 1.

2.4 pCO2 product
Line 227: Reason why 10 and 70 are chosen? Is there a possibility that even in smaller provinces 10 neurons could lead to overfitting? The polar regions are set right at 10.

Response: We test the number of neurons from 5 to 300. The MAE continuously increasing after 70. The variation of MAE would be difficult to see clearly if all spots were showed, so the results after 70 were omitted. Seems the way the result shows may be misleading, so we redrew the Fig.4a. Not just focusing on overfitting, too few neurons may lead to insufficient learning capacity for complex nonlinear relationship, so we tested the performance of FFNN with different number of neurons.

Line 231: Does this vary neuron number test really limit overfitting? Taking the lowest MAE from the internal train/validation split during FFNN training step just means it is likely replicating training data well and due to random split inside autocorrelated validation data this doesn't change much. Being clear in Line 152-163 about how / when MAE is used to evaluate could clear this up.

Response: The MAE used here was calculated using a K-fold cross validation method grouping by year. The training data and validation data were taken from different years and were relatively independent. The MAE theoretically tend to

increase when insufficient learning capacity due to too few neurons or overfitting problem due to too many neurons appear. The result shows that the MAE did increase when the number of neurons was lower than 10 and higher than 100.

2.5 Validation

Line 237: Unique use of the K-fold cross validation method grouping by year.

Response: Since samples within 500 km in the same period were correlated, grouping by year makes the training data and validation data relatively more independent.

3 Results and discussion

3.1 Biogeochemical provinces and corresponding predictors of pCO2

Table 3: Add more to the caption on the order of the predictors listed.

Response: Thank you for your suggestion. The caption has been modified.

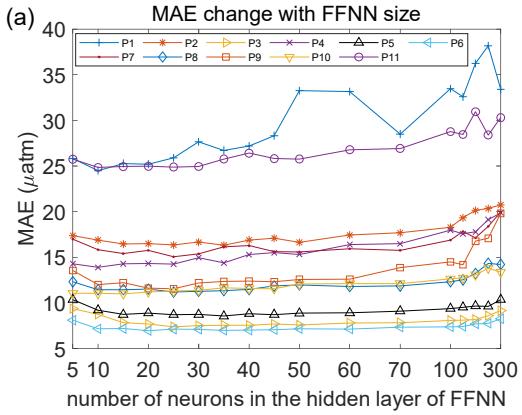
3.2 pCO2 product

Line 317: Does this mean you first went through the stepwise FFNN process using the same neuron number? Then when the best predictors where determined you used the varying neuron number test (from 10-70) to find the best neuron number? Then you used the K cross validation to test robustness? Clarify. Link to back to Figure 1 if you need to.

Response: Yes, the stepwise FFNN process use the same neuron number. Since the result of varying neuron number test shows there is almost no insufficient learning capacity or overfitting problem when the number of neurons was in 10-70 and the MAE differs a little. Any number of neurons in this range was considered suitable. Although a loop of "stepwise FFNN – neuron number test – stepwise FFNN - …." to use different number of neurons in the stepwise FFNN process may further decrease the predicting error, the effect was not so significant and a stable end is not easy to find. In the future work the role of the varying neuron number test may be reconsidered, but now it is used for avoiding insufficient learning capacity or overfitting problem in spite of the low possibility of appearance, and decreasing the predicting error slightly.

Figure 4: Still not sure on this test limiting overfitting. Looks like they all (except at the poles maybe because it is not well constrained…? As in your Table 4) just level out. Using the same FFNN predictors and the same targets how reproducible is this Figure? Or is it dependant on the initialization on that run?

Response: The Fig. 4a has been modified to show the additional result of 70-300. The MAE increased after 100. In MATLAB we used "setdemorandstream(pi)" to set initial state stable, thus the result using the same FFNN predictors and the same targets is completely reproducible.



Line 334: Good to state this up front. Other than these regions it does look good. However, if the goal from the introduction is get at the air-sea flux, how important are these regions for the global marine CO2 flux? Suggest in conclusions what could be done in the future to improve these regions?

Response: The east equatorial Pacific is the most important CO2 source, while the subpolar Pacific was a sink in summer and a source in winter. The CO2 flux in the Southern Ocean near the Antarctic continent was near zero due to ice cover. For the future work to improve these regions, maybe more parameters related to biological activities, El Nino and La Nina, or remote sensing parameters will be added to constrain the pCO2 in these regions.

Figure 6a: Would be nice to also have the atmospheric xCO2 product on this Figure for comparison.

Response: Thank you for your suggestion, the atmospheric CO2 has been added in the Fig. 6a.

3.4 Validation based on independent observations
Line 395: Nothing is obvious to every reader. Remove and clarify.

Response: Thank you for your suggestion. The unbefitting description has been removed.

Line 442: "… was credible." Is consistent with and improves upon? Readers should want to believe in what you did! Got to sell it a bit!

Response: Thank you for your suggestion. The text has been modified as "suggesting that pCO2 predicting based on regional different predictors

selected by the stepwise FFNN algorithm was better than that based on the globally same predictors."

4. Conclusions

Line 447-465: This needs a bit of a rework. Feels like recycled sentences from throughout. What should readers take away from your work? How can this approach be applied in other studies? Who benefits from this improvement? Where is more work needed (e.g., polar regions), how could improvements be made?

Response: Thank you for the suggestion. This part has been rewritten as "A stepwise FFNN algorithm was constructed to decreasing the predicating error in the surface ocean pCO_2 mapping by finding better combinations of pCO_2 predicators in each biogeochemical province defined by SOM method, based on which a monthly 1°×1° gridded global open-oceanic surface ocean pCO_2 product from January 1992 to August 2019 was constructed. Our work provided a statistical way of predictor selection for all researches based on relationship fitting by machine learning methods, and shows that using regional-specific predictors selected by the stepwise FFNN algorithm retrieved lower predicting error than using globally same predictors. This stepwise FFNN algorithm can be also used in pCO_2 mapping researches for higher resolution and coastal regions, and also in other data mapping researches using SOM or other region dividing method. The prepare work was only collecting as many parameters, which are possibly related to the target data and need to be sufficiently available in time and space. However, high predicting error in special regions still remains to be improved, such as polar regions and equatorial Pacific. Since the result of the stepwise FFNN largely depends on the way biogeochemical provinces divided, improving of SOM step is still necessary. Besides, the FFNN can be replaced by any suitable type of neural networks. A possible way to improve the performance of stepwise FFNN algorithm is to modify the structure of FFNN or to use better networks.".