

Review of the paper entitled "Reconstruction of global surface ocean pCO₂ using region-specific predictors based on a stepwise FFNN regression algorithm" by Zhong et al.

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 predictors 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.

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.

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?

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.

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?

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

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.

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

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?

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.

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.

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.

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.

2.5 Validation

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

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.

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 were 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.

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?

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?

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

3.4 Validation based on independent observations

Line 395: Nothing is obvious to every reader. Remove and clarify.

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!

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?