

Interactive comment on “FFNN-LSCE: A two-step neural network model for the reconstruction of surface ocean pCO₂ over the Global Ocean.” by Anna Denvil-Sommer et al.

Luke Gregor (Referee)

luke.gregor@uct.ac.za

Received and published: 5 February 2019

The study introduces a new method that calculates pCO₂ as climatology + anomaly. I think that this is a novel approach and is definitely relevant to the community. The authors do a good job of explaining the method and comparing it with past studies. I really like the approach and the method seems to perform well relative to the other methods in the SOCOM ensemble. I enjoyed reading the manuscript; however, there are many typos scattered throughout the document. I have noted most of these with track changes in the PDF document using Adobe Reader (attached as "supplement"). Below are some general comments, then repeated mistakes that can be fixed with

C1

find/replace.

General notes

The authors limit their study from 2001 to 2016. I realise that data is sparse before 2001, but it would be interesting to see how the model fares. It is useful for the community to know if this method (and in general the climatology-regression approach) can predict in data-sparse periods. It seems to be good at predicting in data-sparse regions.

It is really good that the authors use quasi-regular spacing to define the train-test-validation splits, but I would like to know more about how they defined these splitting points. On this point, I also think it would be useful to have a measure of how much a model overfits. The authors could compare the test RMSE with the RMSE of the model trained with 100% of the data. If the latter is much smaller than the 50:25:25 splits, then there is overfitting. This is important for when the method is compared with other gap-filling methods.

Regarding the figures, I struggle to see the difference between the brown and the black lines as I'm slightly colourblind. I would like to see another colour for the Jena-MLS method.

Repeated mistakes:

- absolute bias = mean absolute error ($MAE = \text{abs}(\hat{y} - y).mean()$) – though I would like to see biases too ($\hat{y} - y$)
- data base = database
- data set = dataset

C2

- southern/northern hemisphere = Southern/Norther Hemisphere

Please also note the supplement to this comment:

<https://www.geosci-model-dev-discuss.net/gmd-2018-247/gmd-2018-247-RC2-supplement.pdf>

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2018-247>, 2018.