Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-247-AC3, 2019

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "FFNN-LSCE: A two-step neural network model for the reconstruction of surface ocean pCO_2 over the Global Ocean." *by* Anna Denvil-Sommer et al.

Anna Denvil-Sommer et al.

anna.sommer.lab@gmail.com

Received and published: 13 March 2019

We would like to thank Luke Gregor for his positive and helpful review. We have answered his comments below.

Luke Gregor: 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

Printer-friendly version

Discussion paper

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 find/replace.

Authors: Thank you for the positive assessment of our study. We apologize for the typos.

L.G.: 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.

A.: Thank you for your suggestion. We agree that the extension of the time series is a priority for future developments.

L.G.: 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.

A.: To be used by the FFNN algorithm data have to be formatted as a list “latitude by latitude”. Data are read “line by line” with each line corresponding to a latitude. Figure S1 presents the method of selection that was applied. This approach ensures that information from almost all regions and months are used for training. The RMSE for the case when 100% of data were used for training amounts 14.8 μatm, only slightly smaller than the one computed for cross-validation (17.97 μatm). We added this information to the manuscript.

L.G.: 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.

A.: Colour was changed to red.

L.G.: "Methods based on ANN are able to represent the large class of pCO2 -driver relationships, but they are sensitive to the number of data used in the training algorithm and can generate artificial variability in regions with sparse data coverage.". - I find the first part of the sentence a bit unclear. Not sure what is meant by "large class of pCO2-driver relationships". The second part of the sentence requires a reference.

A.: "Large class of pCO2-driver relationships" refers to a large variety of combinations of drivers/predictors. The first part of the sentence was modified to "Methods based on ANN are able to represent the relationship between pCO2 and a variety of predictor combinations (e.g. pCO2=f(SSS,SST,SSH) or pCO2=f(SSS,SST,xCO2,CHL,MLD)). A reference was added.

L.G.: "The model is easily applied to the global ocean without any boundaries between the ocean basins or regions". - This amazes me! Good results with one domain. A great advantage to the method!

A.: Thank you!

L.G.: "Based on Rodgers et al. (2009) who reported a strong correlation between natural variations in dissolved inorganic carbon (DIC) and sea surface height (SSH), SSH was added as a new driver to this list". - I like the addition of SSH to the list, but then why not extend the predictions to 1993 (SSH limiting) or 1998 (Chl-a). I agree with Reviewer 1 that this can be discussed more.

A.: In this study, the priority was given to data coverage for the development of the model. The period retained (2001-2016) represents 77% of the SOCAT dataset. The extension of the period of reconstruction will be considered in future studies.

L.G.: I would have hoped that the climatology approach would be better at estimating pCO2 for periods and regions where data is sparse.

A.: It is difficult to assess the skill of the model over regions with sparse data coverage based on observations only. An alternative approach would consist in subsampling a numerical model with the temporal and spatial coverage of real observations. These pseudo-data will be used for reconstructing surface ocean pCO2. The comparison between reconstructed and modelled pCO2 distributions will provide an upper estimate of the accuracy.

L.G.: "Monthly global observed physics reprocessed products distributed through the Copernicus Marine Environment Monitoring Service (CMEMS) (0.25ox0.25o) (http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=MULTIOBS_GLO_PHY_REP_015_002) were used for SSS, SST and SSH". - ARMOR3D L4? Reference: Guinehut et al 2012?

A.: Yes, it is the ARMOR3D L4 data. We added the reference Guinehut et al. (2012).

L.G.: "For MLD, daily data from the "Estimating the Circulation and Climate of the Ocean" (ECCO2) project Phase II, at 0.25ox0.25o resolution (Menemenlis et al., 2008) were used". - Please add the type of data product.

A.: The product name is ECCO2 and the release name is Cube 92. The information was added to the manuscript.

L.G.: Lignes 140-142 - Converted to radians first.

A.: Yes, of course. The information was added.

L.G.: "There is a well-known empirical rule advising to have a factor of 10 between number of patterns (data) and number of connections, or weights to adjust". - Please add a reference. COMMENT: Landschutzer 2013 used 30. Are there any signs of overfitting to the training data? i.e. is the training RMSE score much lower than the test RMSE score?

A.: Landschutzer et al. (2013) based their choice on Amari et al. (1997). A factor of 30 allows avoiding cross-validation. However, we believe that a problem such as the reconstruction of the time and spatial distribution of surface ocean pCO2 requires cross-validation. In line with Amari et al. (1997), we use a factor of 10 that necessitates a cross-validation to avoid overfitting.

L.G.: “Validation and evaluation data sets were chosen quasi-regularly in space and time to take into account all regions and seasonal variability”. - This is good, but I would like to know more about how the quasi-regular sampling was done.

A.: As mentioned above, to be used by the FFNN algorithm data have to be formatted as a list “latitude by latitude”. Data are read “line by line” with one line per latitude. This approach ensures that information from almost all regions and months are used for training.

L.G.: “In order to improve the accuracy of the reconstruction, the model was applied separately for each month”. - So in effect, 12 models were trained? This is mentioned later in the text. Move that point here.

A.: The corresponding sentence was added.

L.G.: “We have developed a FFNN model with 5 layers (3 hidden layers)”. - Sentence is unclear in this context. (Put earlier).

A.: We placed this sentence earlier in the manuscript.

L.G.: “There were thus 12 models sharing a common architecture but trained on different data”. - This could move to the first step to clarify the point a bit earlier. I have noted where this point could be moved to.

A.: We have added a sentence to the description of the first step. We kept the sentence as it corresponds to the 2nd step of the model and we feel that it is needed for clarification.

L.G.: “At this step, in order to increase the amount of data during training and to introduce information on the seasonal cycle, the model was trained using as a target pCO2 data from the month in question as well as those from the previous and following month during the entire period 2001-2016”. - My feeling is that having monthly models would really reduce the number of data in the Southern Ocean during winter. Really very little data. I have made a comment about this for Figure 1.

A.: We agree that the Southern Ocean is particularly under-sampled with a bias towards summer months. However, the use of a single model for the reconstruction of surface ocean pCO2 over the full period will not significantly improve the accuracy in the Southern Ocean, but it will reduce it over the global ocean.

L.G.: “Figures 1 (b) and 1 (c) show an example of data distribution for the sole months of January over the period 2001-2016 (Fig. 1 (b)) and for the three months time-window December-January-February 2001-2016 used in the training algorithm of the January FFNN model (Fig. 1 (c))”. - I would like to see this for July rather - This would really highlight the missing data for the Southern Ocean. Also, the figure is very small. I would prefer if the subplots were vertically stacked (single column).

A.: The aim of these figures is to show a possible distribution of data and an impact of data from two neighboring months. We did not want to focus on a particular month and discuss associated problems. Figures were vertically stacked.

L.G.: “Each sampling was tested...” - "sampling fold" would be a lot clearer.

A.: We changed the text accordingly.

L.G.: “...different initial values that are chosen automatically”. - how are these chosen automatically? There must be an initialisation or a grid of hyper-parameters passed to the model. Please clarify.

A.: The values are chosen randomly.

L.G.: “Note that a higher number of parameters did not show a significant improve-

ment of accuracy (not shown)". - Remove (not shown). Makes me want to see this. COMMENT: This is good. shows that over-training is being minimised.

A.: Done.

L.G.: "In order to provide the final output, the selected FFNN architecture is trained on all available data: 100% of data for training, 100% for evaluation and 100% for validation". - I would like to see RMSE from validation data pre- and post-training, but only necessary in the supplementary materials. The authors can then state that their method is or is not prone to overfitting. This would show if there is overtraining.

A.: Thank you for the suggestion. We agree that it would be interesting to present these values to provide additional information on the model development. We have noted it for the next model release. The RMSE for the case of 100% of training data is 14.8 uatm. The information was added to the manuscript.

L.G.: Ligne 225: FFNN-LSCE

A.: Corrected.

L.G.: ". The global climatology was reconstructed with a satisfying accuracy during step 1 with a RMSE of 0.17 μatm and r2 of 0.93". - This is very low! might be worthwhile commenting on. This shows that pCO2 deviation from the climatology contributes nearly all the error in estimates!

A.: The high scores during step 1 underline the capacity of our FFNN approach to reconstruct the surface ocean pCO2 when there is a good data coverage. However, it is hard to conclude whether the 1st or the 2nd step has more or less contribution to total model accuracy. As you and Christian Rödenbeck mentioned, the estimation of the climatology during step 1 includes data outside of the reconstruction period, which might influence final results. This needs more investigation using numerical models, for example. However, we expect that the influence of the data outside of the period will not significantly modify our results.

L.G.: Ligne 253: absolute bias - Rather refer to this as the mean absolute error and I would also like to see the bias.

A.: We have found that total bias is difficult to explain as it varies a lot over the global ocean and its small value does not mean that the biases are small everywhere, often it results from compensation. We added a table presenting the bias to the Supplementary material. “Absolute bias” was replaced by “mean absolute error” (MAE).

L.G.: “The RMSE between SOCAT data and the climatology of pCO2 from Takahashi et al. (2009) equals 41.87 μatm, larger than errors computed for the regional comparison between FFNN and SOCAT (Table 1)”. - Is this given as a normalisation/comparison - I think this is quite useful.

A.: Indeed, this is given as a comparison. The comparison stresses the skill of the model to reconstruct interannual anomalies. An error between reconstructed pCO2 and SOCAT larger than between the climatology by Takahashi et al. (2009) and SOCAT would indicate that the model failed the capture the interannual variability.

L.G.: Ligne 269: “...Subtropical Pacific...” - assuming NH, but better to clarify.

A.: Done.

L.G.: “Despite large time mean differences computed over the eastern Equatorial Pacific, scores are satisfying at the regional scale indicating error compensation by improved scores over the western basin”. - I’m glad this is spoken about. Please add a number to explain "satisfying".

A.: Done.

L.G.: Ligne 273: “. . .South hemisphere...” - Southern Hemisphere, each should be capitalised.

A.: Done.

All typos and mistakes were corrected.

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

Printer-friendly version

Discussion paper