Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-37-RC2, 2020

Earth System Science Data Discussions

Interactive comment on "A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach" *by* Daniel Broullón et al.

Anonymous Referee #2

Received and published: 22 May 2020

This work applied the method of Broullon et al. (2019) to TCO2 and extended the NN model by including year as an input and including TCO2 computed from LDEO pCO2 in the target. The manuscript is clearly written except for a few elements that require clarification; and the climatology TCO2 data are useful for other modelers.

General Comments:

While including LDEO is expected to improve modeling TCO2 dependence on input variables in the surface waters, it raises two questions. The first is the increase of the spatially biased sampling, which could lead to model optimization more weighted toward fitting the surface measurements. The second is the unknown system bias

of the computed TCO2 relative to GLODAPv2 TCO2. This bias could be estimated if there were enough overlapped points within the space and time resolutions of the training data. If you do the regression of Fig.2a using only the surface data, you may detect the bias. As the regression slopes of Fig.2a and 2b are 1, subtraction of the two predictions indicates the system bias of computed LDEO TCO2. You mentioned on line 280 that "Interestingly, CANYON-B is able to reproduce the TCO2 data derived from the complete LDEO dataset with a lower error than the one it obtains for the complete Gv2QC dataset in the surface ocean. . .". Another explanation to this is that because of the unbiased nature of a NN model (the overall prediction error is close to zero), the system bias of LDEO TCO2 could happen to fall between the prediction biases of Gv2 TCO2 in the surface and interior.

Absolute errors are often used in tables and figures. They hide the information whether the errors show under-estimate or over-estimate; Therefor showing negative errors are recommended.

Specific Comments:

Line 132: The reference of Rumelhart et al. (1986) is missing.

Line 149-152: Could you give more details on how to ensure biogeochemical variables have a larger influence than position variables?

Line 254: Why average 1981 to 2015 to obtain 1995 climatology? You have 20 year from 1995 to 2015, but only 14 year from 1981 to 2015.

Line 261: This is an important criterion to select the NN for making prediction, but no detail available. Could you supply more information in the supplement material on the influences of position variable of the networks?

Table 2. Are the errors absolute? If so, please state explicitly. Also, the global errors should be added. Showing negative errors are more meaningful.

Table 6: Does the label "NNGv2" means NNGv2LDEO?

Table 7: Is the bias absolute? If so, please state explicitly. The global errors should be added. Showing negative errors are more meaningful.

Figure 1b: “y=1x +- 7.8” should be y=1x - 7.8.

Figure 3: Showing negative errors are more meaningful.

Figure 4b: The error bar for depth < 50m should be added using the surface errors.

Figure S1.b: There should be a “+” operator between bj*a0 and SUM(wij*ai) in the activation function.

Figure S2: How the std is calculated for T, S, and pCO2. Modelled TCO2 is larger than observed TCO2 for all pCO2 STD > 4. How to explain this?

Figure S3. If the difference is absolute, please state clearly. Showing negative errors are more meaningful.

Figure S5 and S6: Showing negative errors are more meaningful.

Figure S7a: The model produces a much larger seasonal amplitude in the surface water. Unless measurements are not available in all months, the seasonal amplitude of the climatology should be no larger than that of the measurements. Does this indicates either over-fitting or extrapolation in seasons of no measurements.

Figure S8 and S9. Plotting land with colors confuses grasping the contours of differences.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-37, 2020.