Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-37-RC1, 2020

Open Access 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 #1

Received and published: 21 May 2020

The authors describe a neural network approach to obtain TCO2 from geolocation, depth, temperature, salinity, oxygen together with nutrients (nitrates, phosphates, silicates). This approach is trained on GLODAPv2.2019 and LDEO data, therefore with a good surface and in-depth amount of data. Using this neural network, NNGv2LDEO, they produce a monthly TCO2 climatology for the global ocean at different depths. Additionally, they also produce a surface pCO2 climatology.

The manuscript is globally well-structured and understandable and the associated clmatologies could be useful for modelers. However, a few elements still require clarification.

- In section 2.1, "an additional criterion based on the influence of each input variable on the TCO2 extracted" is mentioned. More details on that criterion and its results would be appreciated.

- In section 3.1, when justifying the choice to continue with NNGv2LDEO over NNGv2QCLDEO, you highlight that Gv2 has more data points than Gv2QC in the Mediterranean Sea for example. While these additional data are important for the quality and spatial coverage of your climatology, the absence of these data in Gv2QC might be because of their poor quality. How do you thus justify their use? Additional data in an underrepresented area is a good thing but if these data are of poor quality, does it really constitute an improvement? It would be interesting to compute the errors of NNGv2LDEO specifically in the areas not in Gv2QC. This is kind of visible in Figure 3 but could gain from being developed.

- In section 3.2, it is mentioned that "the high correlation holds for all depths". It would be interesting to see the variations in these correlations according to depth layers.

- In the dataset provided, there are some extremely high pCO2 values (>5000, North of Russia). These are not addressed in the manuscript ad absolutely should be. When plotting the data without tuning the colorbar, it's the first thing a user sees. If stemming from an error, please consider correcting the dataset. If not, this should be described in the manuscript as to provide the data user with a warning of these extreme/erroneous data.

- What is the monthly distribution of the data used and thus how does this affect the monthly climatology (i.e. underrepresentation of winter months leading to a lower confidence in the data product)? - A pCO2 climatology product is created and compared to other existing products. Several figures (including the supplementary material) highlight these comparisons but it would be useful to have a general climatology representation (something like Fig 7). This would avoid/nuance the problem mentioned earlier of the extreme values. - Be aware that the Figures 6, 7 and S4 are not readable by a

colour-deficient individual. Please consider changing the colormap chosen.

- L65: “being these changes reflected” -> these changes being reflected

- L105: the list of predictor variables is missing the date of sampling

- L132: “minimize de errors”, remove de

- L133: wrong spelling of author name

- L255: “The variable selected (. . .) was that labeled as spco2_raw”. Please indicate the complete name of the variable, or what it refers to

- L397: “are the same of as those”, remove of

- Table 6: The RMSE vs r^2 table is kind of hard to read. Maybe find another way to arrange it. Or describe it more in the title.

- Figure 1: “first level are always the same for each of network”, remove of

- Figure 5: NGv2LDEO -> NNGv2LDEO; HOL ALOHA SURFACE -> HOT ALOHA SURFACE

- Figure S5: spelling of the word climatology

- Figure S8-S9: spelling of the word microatmospheres

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