

Biogeosciences Discuss., referee comment RC2  
<https://doi.org/10.5194/bg-2021-33-RC2>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on bg-2021-33

Anonymous Referee #2

---

Referee comment on "Weekly reconstruction of pH and total alkalinity in an upwelling-dominated coastal ecosystem through neural networks ( $A_T$ pH-NN): The case of Ría de Vigo (NW Spain) between 1992 and 2019" by Daniel Broullón et al., Biogeosciences Discuss., <https://doi.org/10.5194/bg-2021-33-RC2>, 2021

---

The reconstruction of regional pH and  $A_T$  time series could be important for policy makers and the public to evaluate and understand the long-term acidification in local coastal regions. In this study, the authors use a neural network approach to reconstruct pH and  $A_T$  time series in the Ria de Vigo from 1992 to 2019. The seasonal cycle and long-term trends of pH and  $A_T$  are then determined and discussed using the reconstructed time series data.

I have some major concerns that need to be addressed before the publication of this manuscript.

### *Concerns:*

The training, testing, and validation datasets all come from the IIM database and are separated randomly. However, randomly splitting the IIM data into three datasets can be very risky due to the potential auto-correlation among these three datasets, especially for periodic data. If the testing dataset is correlated with the training dataset, then it is not suitable to be used to assess the generalization of the network. In addition, this could be dangerous to the determination of long-term trends. A complete independent dataset

could be used for testing. If an independent dataset is not available, the model could be built on anomalies with periodic cycles removed. Alternately, the training dataset and the testing dataset could be separated in a way that there is no time overlap between them.

Please explain why you choose 28, 34, 40, 46, and 52 as the tested number of neurons in the hidden layer. The number of neurons in the hidden layer seems too large considering the size of the input variables. This will likely lead to overfitting. Please justify that overfitting is not an issue for this model.

I am particularly concerned about the relative importance of input variables on  $A_T$ . I assume that the overall connection weights have been calculated. The weak connection between salinity and  $A_T$  is a red flag to the neural network. However, both strong salinity- $A_T$  correlation and the connection weights/inputs relative importance, which are contrary, have been used to explain the change in  $A_T$ .

The description of the best network selection should be clarified. According to Table 1, LLDTSPNSiOYW trained with 34 neurons instead of DTSPNSiOYW with 28 neurons provide the lowest RMSE and highest  $r^2$  on the test dataset for pH. There is no need to trade-off between RMSE and  $r^2$  for pH model selection. For  $A_T$ , DTSPNSiOW with 46 neurons is better than LLDTSPNSiYW with 52 neurons considering RMSE and  $r^2$  for the testing dataset. The best networks in Table 1 are different from the best networks stated in lines 225-231. By the way, it is not clear what "the most accurate representation of the validation data" means. As stated in lines 155-157, the validation dataset is used to stop the training process when the performance on this dataset does not improve during six consecutive iterations of the training process. So why it is used as one of the criteria to determine the best networks? It should be more clear how the best networks are selected.

It is worth calculating and discussing the long-term changes in the seasonal cycle amplitudes of pH and  $A_T$ . The pH seasonal cycle amplitude change can be predicted according to the change in buffer capacity. Therefore, it could be used as a potential way to check the robustness of the model.

*Below are some minor comments:*

Section 2.2, what are the uncertainties for pH and  $A_T$  measurements?

Provide a table with the number of observations in the training data for each depth range (0-5 m, 5-10 m and 10-15m) and zones (inner and outer/middle). This will be helpful to model performance discussion.

Figure 2, specify in the caption the differences between (a) and (b) as well as between (c) and (d).