Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-33-RC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2021-33

Anonymous Referee #1

Referee comment on "Weekly reconstruction of pH and total alkalinity in an upwelling-dominated coastal ecosystem through neural networks (A_TpH-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-RC1, 2021

Summary and overall impression

The manuscript makes use of ship measurements of physical and biogeochemical properties (temperature, salinity, phosphate, nitrate, silicate, dissolved oxygen, pH, total alkalinity A_T), in combination with measurements taken at 7 time-series stations (same variables as above but without pH and A_T) in the Ría de Vigo, an embayment in NW Spain. They use the ship data to train a feed-forward neural network to then estimate time-series of pH and A_T at the 7 stations. They validate the time-series estimates with ship measurements of pH and A_T near the time-series stations. They then present the time-series estimates with a focus on the seasonal cycle and long-term trends.

This is an interesting study, presenting a new dataset of high-frequency time-series of pH and A_T in the upper 15 m of the Ría de Vigo. The findings of the paper help to quantify and better understand the variability and long-term trend in regional acidification, and could be important for stakeholders with an economic or ecologic interest in the Ría de Vigo. It may also be of interest to scientists who want to estimate time-series based on available proxy data in different regions. However, I do not believe this paper is ready for publication yet. I have some major concerns about the method and its validation, as well as about some of the results, as outlined in the general comments below.

General comments

For training the network, data from the Ría de Vigo and beyond (offshore Galicia) was used. There should be a discussion on the validity of the statistical relationships between

the predictor and training data from within and outside of the Ría de Vigo. Due to the special conditions in the embayment, the statistical relations from outside the embayment might not hold inside the embayment. Presumably, inside the embayment, the effect of the freshwater influx would be a lot stronger than offshore. The fact that salinity is a weak predictor of A_T in the network (Table 2), while salinity has a high correlation with A_T in the embayment (L.276 & L. 315) is a further concern.

The training data extends until 50 m, while the labeling data only extends until 15 m. Similarly, the periods of the two datasets are considerably different (1976-2018 vs. 1992-2019). There should be a discussion on why this is ok. As both time and depth are inputs into the network, this is probably not an issue. It is likely to be a good choice as it increases the amount of available training data. Nonetheless, this should be discussed.

The statistical properties of different set-ups including different predictors and number of neurons are presented (Table 1) but the final neural network is only run once. If a different set of testing, training, and validation data was chosen, i.e. if the network was run again, the statistics could change, and the output would be slightly different, and in the worst case completely different. Thus, the robustness of the training should be checked, e.g., by conducting a boot-strapping approach where each time the network is run, the data that is assigned as "training" and "validation" and "testing" data is different in each run. I suspect that the statistics might change slightly when re-running the network as I find it odd that cW and sW have very different weights when they have a very strong statistic relationship to each other. It might also be that in a different run, salinity would be a more important driver for A_T than it currently is.

I appreciate the transparency of presenting how the final set-up was chosen (e.g., which predictors and how many neurons). However, I think that just letting the network "decide" on the best performance can be dangerous, if it doesn't make physical or biogeochemical sense. E.g., what if the set-up with only DTS had been the best set-up according to the tests? Then the biogeochemistry would have been completely ignored in the network, and there might have been some overfitting. In addition, the list of tested predictor combinations does not include all possible combinations, e.g., DTS are predictors in all of them. My suggestion would be to use all available predictors that make physically and biogeochemically sense (T/S, all nutrients, and oxygen). If some of them don't have a strong statistical relationship to the target variables, the network will weigh them less. Then it can be tested with the presented method which parameters on position or time further improve the network (as well as testing the number of neurons).

The validation of the time-series estimates is done with observations from the IIM database (i.e., the training data) near each of the stations. It is not stated explicitly if this data is independent, i.e. if it is not part of the training data set. As it was not stated otherwise, I assume that the validation data is part of the training data and is thus not independent. This is a major concern as validation should be conducted with independent data. Further, the validation of the time-series estimates is very limited, which is understandable due to the data sparsity. However, with the current testing, we do not know if the trend and the interannual variability is correctly captured.

Table 5 shows that the trend in the pH increases with depth. This is contrary to what I would have expected and should be discussed. Does the trend in pH also increase with depth at other locations, e.g., BATS?

Specific Comments

Title: Consider adding "surface" in the title (unless the Ría de Vigo is only ~15 m deep…)

Title: Consider changing "coastal ecosystem" to "coastal embayment"

Abstract: Mention (for the non-Spanish speakers) that the Ría de Vigo is an embayment.

L.22: From Fig. 5, I cannot see that A_T is significantly lower in the inner stations compared to the outer ones. What is this statement based on?

L.28: -0.0020 (or -0.0019?) and -0.0032

L.28-30: Add uncertainties

L.32: What use is the trained neural network to the community? Can it be applied to other data? If so, it should be stated, if not, that part of the sentence removed.

Introduction: Add the significance of knowing about changes in A_T (why not e.g., DIC)?

L.85: I would have liked to find out how deep the Ría de Vigo is. Is it only 15 m? That matters for the interpretation of the findings.

L.87: Name the citations

L.94: Consider discussing the benefit of high-frequency local data. Why would monthly fields not have been "good enough" to investigate the seasonal cycle and long-term trend?

L.98: Consider specifying what these ecological and economic services are

L.103: Add "e.g." before the references as there are a few more

L.109: Remove "global"

L.132: Should be Sect. 3.1(?)

Eq. 1 and 2: Specify that week here refers to the week of the year, going from 1 to 53 (if that is the case)

L.142: It should be stated why this approach of taking neurons between 28 and 52 was taken. In the lead author's previous study, the approach by Velo et al. (2013) was taken, where the number of neurons that were tested is 32, 64, 128 and 264.

L.197: The details of the Monte Carlo simulations have to be stated here.

Table 1: It looks like network 4 has better statistics than network 3.

L.232: It would be interesting to know if that is a specific region in the embayment, maybe it is a region with lots of freshwater influx. What implication does this finding have on the uncertainty of the time-series estimates of pH and A_T?

L.324: It should be stated where the DIC data comes from. Was this directly measured on the ships? Which dataset is it from?

L.326: Is this difference significant? (0.89 vs. 0.91)

L.334: in L. 47, it says it's between -0.0017 and -0.0026

L.341: How does a lack of data increase the pH?

L.356: There is also a global trend in deoxygenation linked to climate change (e.g., Keeling et al., 2010)

L.373: Add "in the upper 15 m"

L.383: Consider adding implications of these findings, e.g., for the economic and ecological services of the embayment

L.384: Consider stating explicitly what other questions could be answered with the data

Fig. 2: The lines in the inserts are different than in the large plots in a and b, and the x-axis label is missing

Fig. 4 and 5: The trendline and a running mean could be added, as well as the RMSE for each of the subplots.

All Figures should have a higher resolution.

Table B1: Why is P not included as a predictor for Si and N? Why is Si not a predictor for P?

Fig. S5 and S6: Consider changing the y-axis to make it easier to see the data points