

Interactive comment on “ML-SWAN-v1: a hybrid machine learning framework for the prediction of daily surface water nutrient concentrations” by Benya Wang et al.

Benya Wang et al.

benya.wang@research.uwa.edu.au

Received and published: 6 June 2020

Two additions are recommended related to this observation: 1. Presentation of the relationships between in the inputs and outputs would be informative. As noted in the discussion (line 445) correlation tests can be used to assess the which parameters would be useful for pre-generation. Potentially a scatter plot matrix, with correlation values on the lower diagonal, would be useful to show the relationships concisely.

Response: Thanks for your comments. Please find the attached figure. We did this analysis before the modelling phase. There are strong relationships between nutrients. This is also one of the assumptions for this hybrid model that if we can pre-generate

C1

these nutrients and they can be further used in the final model to predict final target nutrient. In that case, the model could have higher accuracy. We will update the manuscript to include this figure in the result section or in the supplement.

2. It would be useful for the authors to speculate in the discussion, or even undertake the modelling, if training the model on the observed values for the hybrid model, rather than the generated values, would further improve performance. Possibly this would allow the data driven model to fit the underlying relationships more accurately, even if when implemented less accurate predicted values must be used as the inputs. Or possibly the hybrid model is correcting for some of the errors in the generated inputs if there is a systematic bias, and this would degrade performance.

Response: Thanks for the suggestion. However, there is only very limited number of data (less than 30 samples) that have completed nutrient species to do the modelling. The hybrid model should have higher accuracy using observed values instead of the pre-generated data. We will add this in the discussion.

Line 200: Please provide further information on how the 80:20 data split was implemented. Was the last 20% of the time series used for testing, or was a more complex method used? If this was the approach, does this explain why the DON, and DOC for Murray River, was the most important input, because this data is only present in the testing period? As noted in the manuscript, there appears to be a long term trend in the dataset, and only using this data that is more representative of the testing dataset may improve performance? Possibly data from the 1990s does not represent the responses experienced in the more recent testing period? The authors should further comment on these temporal and data length issues.

Response: The main aim of this research is to rebuild the historical nutrient data and explore the short- and long-term changes. The first step is to verify the model performance. In that case, we randomly divided data into 80:20, built the model, tested on testing data (20%), and then repeated all steps for 30 times to further test model

C2

uncertainty and stability (Figure 3). After this, all data points including the testing data were then used to rebuild the historical nutrient data (Figure 4). The different feature importance of DOC and DON in Murray River and Ellen Brook may due to the different nutrient sources and water pathways. We think it is a really good idea to compare model performance on more recent data and old data but it may out of this paper's scope.

Line 316: claims that the generated data enables the hybrid ML to capture long-term trends. Please elaborate on how this is possible, as it is not clear to me. A given set of inputs to the stand-alone ML to predict the generated data will give the same result in 1990 as 2018, so how can long term trends be captured? As outlined above, possibly because using inputs only collected in the more recent period is more representative of the testing data?

Response: The stand-alone model and the hybrid model used the same dataset to build and test model performance. If the pattern only exists in the recent samples, then both stand-alone and hybrid model should have similar fluctuation. The pre-generated nutrient is the only difference between the stand-alone model and the hybrid model. If there are long-term trends in the nutrient concentrations (e.g., TN), similar trends should also exist in the components of TN (either DON or DIN). The pre-generated nutrients emphasise this impact. That is why we suggested the pre-generated data in the hybrid model helped the model to capture long-term trends.

Section 4.2 moves from considering all nutrients to TN only. Please add a justification why this was the case. Is this because TN is the more relevant for the application, or for the sake of brevity only one parameter is analysed in more detail.

Response: Thanks for the comment. TN was selected because TN is the most important and most frequently measured nutrient in many places. In addition, we want this paper to be more concise. That is why only TN was analysed in detail. This hybrid method can be used for other nutrients

C3

Minor points The first sentence of the abstract would benefit from being rewritten. Suggest making this two sentences and reworking, e.g. nutrient data is monitoring, not necessary for monitoring, for example. I found the term "temporal data" confusing, I was assuming this meant lagged data, $Q_{b,t-1}$, for example. It wasn't until I found Table 1 that this means Julian day of the year. It is suggested to be more specific, and call this "Julian day" or "seasonal component" or similar. Line 70: Another relevant study for hybrid modelling may be Hunter et al. (2018) <https://doi.org/10.5194/hess-22-2987-2018> Line 164: remove are from GBM model are generally have less. . . Line 214: Can be divided into there stages, add s to stages. Line 268: I was confused by this sentence, in that lower RMSE and increased MEF are both improved performance, but opposite patterns. Suggest "overall, the scaled RMSE improved from LM. . . And the same pattern was found form MEF. . . Line 277: suggest add to the end of this paragraph, "as such, the WRTDS results are not directly comparable to the other methods" Line 389: the temporal data was more useful for the perennial catchment, Murray River. Could this be because seasonal information is captured in other inputs, e.g. Q for the more ephemeral catchment? Line 399: Please elaborate further on the method used in this section, it is not clear what has changed. Should "gradually" be "sequentially"?

Response: Thanks for these detailed comments. We will update the manuscript accordingly.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2020-4>, 2020.

C4

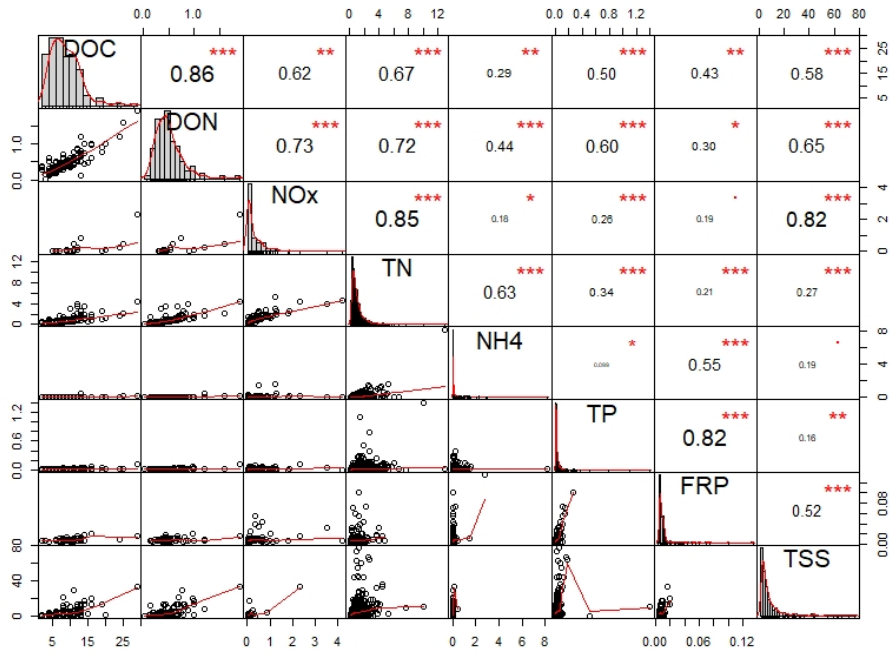


Fig. 1. correlation_between_nutrients