

Report #1

The authors proposed a two-step ML method to overcome the limitation of insufficient observational data for algal bloom prediction. There are still three major problems: (1) Unclear certainty of the two-step ML method; (2) The differences between different models applying the two-step ML method were not evaluated; (3) The impact of the predicted data as training data is unknown.

(1) The evaluating metrics (*MAE*, *RMSE*, *R2*) of three workflows were provided in Table 2, indicating the performance of the methods. Regarding the uncertainty, we used the results of the shuffling years test (using different years for training the model) as the main indicator of uncertainty. Two shuffling-year tests (section 2.4) were conducted based on the two-step ML method (workflow 3) to evaluate the uncertainties induced by variations in the data used to train the ML models (section 3.4) and the uncertainties related to sample frequencies of lake nutrient variables (section 3.5). Figures 4 and 5 visualise the uncertainty related to the ML predictions, and Table S3, 4 show the coefficient of variation.

(2) Based on our results, GBR and LSTM generated similar results in the two-step ML method (Fig. 2). LSTM gave slightly better results in predicting *Chl* concentrations in terms of evaluating metrics (Table 2) and was more accurate in predicting algal bloom events in terms of TPR (Fig. 3a). We added a conclusion about the difference between two ML models in sections 3.3 and 4.1 (L207-208, L273-275)

(3) The impact of the predicted data as training data can be seen in the comparison between workflow 1 and workflows 2, 3 (Table 2, Fig. 2), where the usage of real nutrient data is compared with the predicted data. We concluded the impact in section 4.1 (L280-286),

By using predicted nutrient data as training data, workflows 2 and 3 still capture the seasonal pattern of algal *Chl* concentrations. Especially in workflow 3, LSTM predicted the concentration with $RMSE = 5.81$ mg/L, which is very closed to the $RMSE$ in workflow 1 (5.64 mg/L). With only a minor decrease in performance compared to workflow 1, workflows 2 and 3 had a wider potential range of applications (e.g., interpolation, reconstruct historical data, algal bloom forecast) via making daily forecasts with less-than-daily measured nutrient observations.

Also, if you want to know the daily prediction of *Chl* based on physical environmental factors (without any lake nutrient observations), we presented the results from both ML models below. The performance of GBR model under this scenario was much worse than that using the workflows (1-3). In contrast, LSTM model showed fairly good results, comparable to workflow 2.

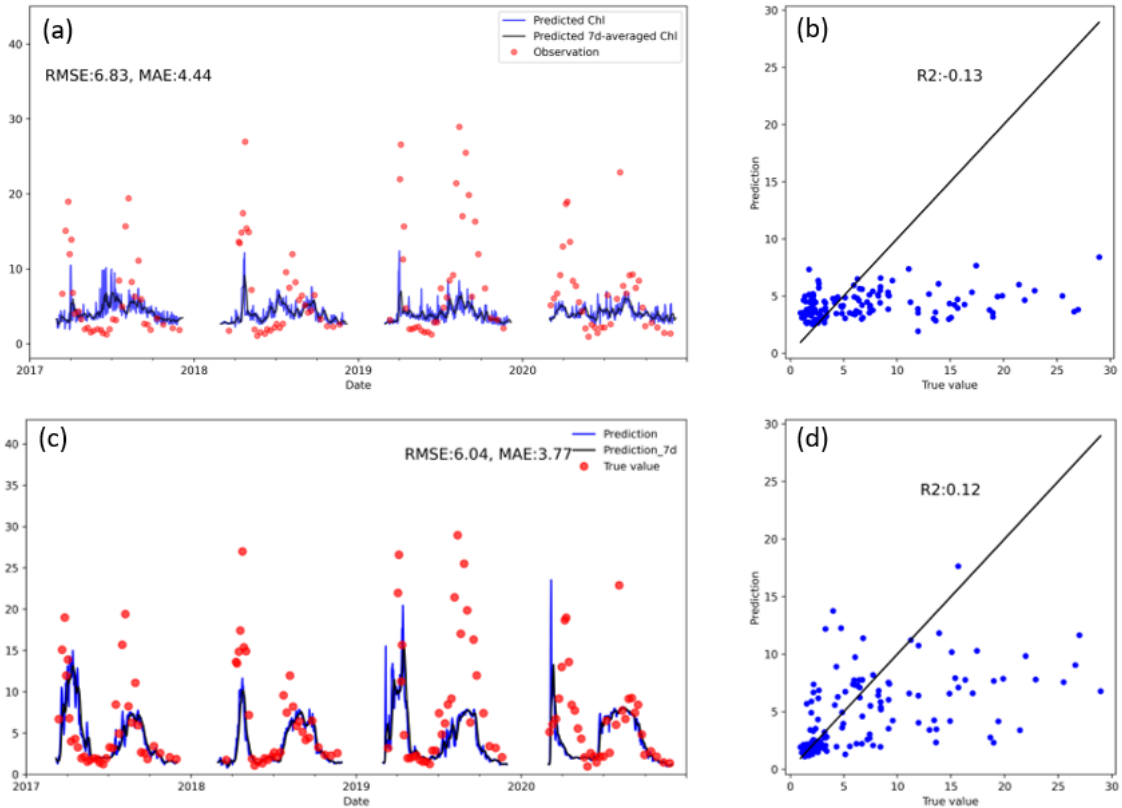


Figure. Timeseries of observed and predicted *Chl* from GBR (a) and LSTM (c) models, and the corresponding scatter plots are shown in panels (b) and (d).

Minor remarks:

1. The authors mentioned the overfitting a few times, whether it's caused by the correlations of the data.

The algal dynamics varied from year to year (Fig. S3). If the models show relatively lower *RMSE* and *MAE* consistently in the training datasets which include the data from 2004-2016, than testing dataset, overfitting is a likely explanation (Table 2).

Thus, we gave more detailed explanation about the higher *RMSE* and *MAE* in testing datasets in L167-169, 297-299. We only mentioned overfitting twice in the manuscript,

but we have changed L. 181 to “likely suffered from overfitting”, as we agree with the reviewer that alternative explanations are possible.

2. In the first step of the two-step ML approach, “each pre-generated nutrient prediction is included in the training data for the next nutrient prediction”, the authors should analyze the prediction accuracy of this approach.

We have assessed the accuracy of nutrient prediction by providing the visual time-series comparison and evaluation metrics (Fig. S6). The predictions from ML models for the 6 nutrient variables assessed were better than that from the process-based model (L185-186).

3. The purpose of designing the three workflows is not explained by the authors.

The logic and purpose of designing three workflows has been clarified in the Introduction (Lines 49-56). And the details of designing the three workflows was depicted in section 2.4.

4. Problems with the narrative logic of the literature abstract in this manuscript, the authors mainly used a data-driven model, but this piece is less described.

In the Introduction section (L31-46), we now retraced the applications of ML models in algal bloom studies and mentioned the shortcoming of the ML models, including lack of interpretability and generalization, limitation induced by sparse nutrient monitoring data. And then, we added the main objectives of our study and logic line behind the three workflows in the following paragraph to make the narrative logic clearer (L47-56).

5. The logic of the literature review of ML models is not clear enough, which makes it difficult to find the development of the models.

See the response to the last comment. We have summarized the development of ML models application in algal bloom prediction in the Introduction section (L31-46).

6. Page 2, Line 36-37. Recknagel et al.'s model is based on a neural network, while Nelson et al.'s model is based on a random forest algorithm.

We have corrected the citation here.

7. Subsection 2.3.2. The authors need to conclude the reasons for comparing GBR and LSTM.

We have added a sentence in the last paragraph of this section (L104-106):

‘Compared to GBR model, LSTM has more complex model architectures, carrying the ‘memory’ from the previous time steps. In this study, GBR and LSTM were applied, respectively, to assess the performance of ML models with and without ‘memory’.’

Report #2

The authors replied to the referees’ remarks in an appropriate way.

The only issue that I can highlight is that I could not find a precise correspondence between line numbers in the response and those in the revised manuscript. Also, the reference to Table S2 in the response is not correct (likely it is to the new Table S1), and the reference to McHugh, 2012, is not reported in the text. I invite the authors to be more precise with these details.

We have corrected the original response and the related information in the manuscript.

And yes, the reference to Table S2 is to the new Table S1.