

Geosci. Model Dev. Discuss., author comment AC3  
<https://doi.org/10.5194/gmd-2022-174-AC3>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC2

Shuqi Lin et al.

---

Author comment on "Prediction of algal blooms via data-driven machine learning models: an evaluation using data from a well-monitored mesotrophic lake" by Shuqi Lin et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-174-AC3>, 2022

---

Dear reviewer,

Thank you for the constructive comments. Please find our replies following.

Major remarks

- Possible overfitting

**The overfitting issue means the model is too closely aligned to the training dataset so that it can not predict the peculiarity in the testing dataset. We think this is the issue existing in the three workflows we tested here. The algal dynamics varied from year to year (Fig. S3). If the models show relatively lower RMSE and MAE consistently in the training dataset which include the data from 2004-2016, than testing dataset, overfitting is a likely explanation.**

**However, we also agree with your point that peculiarities of the algal bloom in 2019 could also contribute to consistent high errors in testing dataset since the most of observed data points were way higher than usual values. Thus, we adjusted our explanations about the higher RMSE and MAE in testing datasets (Lines 172-175, 189-190, 304-307).**

**We added some information about GBR model in Line 93-97, "The hyperparameters in GBR are optimized via *'RandomizedSearchCV'* function within Scikit-Learn library. The loss function of model is chosen as *'huber'*, which is a combination of the squared error and absolute error of regression. Since the target variable in our research Chl concentration has peak values during algal blooms which could be regarded as outliers, the *'huber'* loss function is more robust and gives greater weight to peak values than the mean squared error function."**

**We have tried different combination of numbers of layers and neurons (1-3 layers, 20-200 neurons), but larger numbers of layers and neurons did not obviously improve the results but increased the computational time a lot, and worse results were achieved when the number of layers and neurons were decreased. We added these details into Supporting Information Text S2.**

- Intrinsic variability in the model's results

**Yes. That's the reason we conducted these two shuffling year tests. As we mentioned in Line 335-338, "We suggest that testing strategies similar to the shuffle methods used in this study are needed to accurately evaluate the expected accuracy of ML models when applied to any given site. The estimated uncertainty in shuffling training year tests (Fig. 4) and shuffling training/testing year tests (Fig. 5) can be used to better represent the uncertainty of ML derived forecasts."**

Minor remarks and typos

- The Supporting Information contains some data and plots that would fit well in the main text. For instance, Table S1 is useful to understand the procedure used in the analysis.

**We have moved the table from SI to the main text (Table 1).**

- "Even the LSTM algorithms could not account for previous condition so far back in time". How long is the expected memory of the model?

**L270: The expected memory to consider the formation and deposition of cyanobacteria akinetes may require couples of months extending to the previous ice-free season (Lines 304-305).**

- L56 "beings": begins
- L136-137: "modified Kappa" is not a common metric. Please give a short description of what it represents.

**L136-137: modified accuracy (Kappa) which considers the possibility of the agreement occurring by chance (Table S2; McHugh, 2012)**

- L228 "even though the GBR model usually performs better in Fig. 5c the testing period chosen for use in Fig. 3, showed the opposite result." Cumbersome sentence, please rephrase it. Moreover, the whole section contains a weird use of commas.

**L228: Consequently, even though the GBR model usually performs better in most of chosen 4-year testing periods (Fig. 5), Fig. 3, which shows the results of 2017-2020 testing period, presented the opposite result.**

- Figure 4. The subplots (b) and (c) are not described in the caption.

**The caption has been modified.**

- Text S3, reference to Wilson et al. (2020): it is not in the bibliography.

**The bibliography has been updated.**

- Figure S3, specify that boxplots refer to the period 2004-2020. Maybe it would be more interesting to compute the boxplots excluding 2019-2020?

**This was a typo. The boxplots refer to the period 2004-2018.**

- Figure "Penal": Panel.

Shuqi and co-authors