

Geosci. Model Dev. Discuss., author comment AC2  
<https://doi.org/10.5194/gmd-2022-174-AC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC1

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-AC2>, 2022

---

Dear reviewer,

Thank you for the comments. Please find our replies below.

- The authors mentioned some machine learning models in the Introduction. These are important in the development of algal bloom prediction. The authors should analyze the disadvantages of these ML models and the improvements of their own model.

**I have added some sentences in Line 37-46 that summarizes the disadvantage of the present ML models, and their limitations in water quality and plankton dynamic prediction. The following paragraph illustrates the improvements we made in our models.**

- The literature review of ML models is too simple, which makes it difficult to find the development of models.

**As I replied to the last comment, I have added the information about the development and application of ML models in water quality and algal bloom prediction (Line 37-46).**

- Page 3, Line 73. The significance of designing three workflows needs to be further clarified.

The significance of designing three workflows has been clarified in the Introduction (Lines 47-53), and the details about the workflow were illustrated in the 2.4 section.

- Page 3, Line 80. Why do the authors use GBR and LSTM?

**The characteristics and benefits of these two ML has been added (2.3.2; Line 89-105)**

- The advantage of two-step method is accurate prediction when observations are insufficient. However, workflow 1 performs better than workflow 2 or 3 (Table 1). From this comparison, the two-step method is not an important step that affects the accuracy.

**Yes. But workflow 1 can only predict Chl concentration when lake nutrients observations are available, which could be infrequent in most of lakes, and it is also hard to apply this workflow in the algal bloom forecast due to lack of water quality forecast. The advantage of workflows 2 and 3 is therefore a wider potential range of application (e.g., interpolation, reconstruct historical data, algal bloom forecast) at only a minor decrease in performance for this particular lake. The advantages of workflows 2 and 3 were illustrated in the discussion (4.1 Performance of ML models)**

- From Fig. 3 (e.g., Kappa scores), the PB model also works well. What is the advantage of ML models?

**The advantages of ML models were revealed by the higher TPR. Although PB works well in terms of Kappa scores, it means PB model can correctly predict chlorophyll concentrations during most of the no-algal bloom period. However, our goal is to capture the algal bloom onset as skillful as we can. ML models have a clear advantage in terms of predicting algal blooms event, which only happened in relatively small proportion of time (Fig. 3a, b).**