

The manuscript has been improved now. However, there remains a few comments not well addressed.

Regard model description, while I understand that not all details need to be included, the paper should be self-explained. For critical model components, a brief explanation is needed followed by reference to previous papers, rather than requiring readers to find answers in another paper without any hints. The authors should do a better job of explaining their model for the reader. For example:

1. **Sediment Processes:** The text mentions the binding and release of PO₄ with iron oxides. However, it is unclear whether this is just a description of a natural process or if your model actually simulates it. There also lacks discussion of this mechanism in the following part of this manuscript. Also, it seems this process is simulated by a sediment model, but this sediment model is not mentioned until the Discussion section. This makes it very hard for most readers unless they already know your work well.
2. **Atmospheric Deposition:** Similarly, it is still not stated whether the model's forcing data includes atmospheric deposition fluxes of nitrogen and phosphorus. The authors should clearly state which external drivers are included in their model.

Regarding model validation,

1. The appendix still states that the model results are more realistic than the observations. Although this was acknowledged and corrected in the response letter, the corresponding text in the manuscript has not been revised.
2. The authors note that a key reason for the bottom dissolved oxygen (DO) underestimation is that observations are from the lowest ~1 meter, while the model output is from the lowest 20 cm. However, they keep using the 20 cm model data for comparison. This doesn't make sense as it produces severely underestimated oxygen levels (reaching hypoxic levels), which would significantly alter the simulated fate of N and P in the bottom water. It would be more scientifically sound to use model data comparable to the observed depth (1 m). Currently, the 1-meter comparison is only provided for one station as appendix; please add this comparison for the other station as well.
3. Concerning the validation of stratification, if the authors refuse to include a statistical analysis, please improve the figure. In their current form, it is difficult to assess whether the model correctly simulates the stratification process.

Regarding sections 3.3, it would be better if this section ended after you show the long-term changes in nutrient loads. The parts that come after would fit much better in the Discussion section. In addition, you can also connect these results with your earlier

finding in Section 3.2 about the relationship between load and retention, and strengthen the Discussion by addressing the following points:

1. The implications of the long-term load trends for nutrient retention within the system.
2. A comparison of your findings with larger-scale models that apply a fixed retention value: do your results suggest these models tend to overestimate or underestimate retention? This discussion should be clearly and explicitly
3. A direct discussion on the limitations of the fixed retention approach used in previous larger-scale models, in light of your evidence that retention is load-dependent. This is a key motivation for your study (as stated in introduction) but currently lacks sufficient discussion.