Comments on "Transformation Processes in the Oder Lagoon as seen from a Model Perspective" by Neumann et al.

General comments:

In the manuscript by Neumann et al., the authors discussed the function of the Oder Lagoon in nutrient retention using a 150-m 3D ecosystem model. The lagoon was found to significantly affect the Baltic Sea eutrophication given the high nutrient retention capacity by the estuary (12% of riverine phosphorus and 40% of riverine nitrogen). However, the lagoon itself was found in a highly eutrophic condition which requires further attention and studies. The manuscript is well presented, clear, and easy to follow. However, as a model paper, the model evaluation section is too weak, lacking assessment of some key feature like, oxygen concentration, chl-a concentration (against satellite), and T/S vertical profiles. In addition, there is a lack of in-depth discussion in nutrient retention. I will list my comments in detail as followed.

Therefore, I recommend a major revision before publication.

Detailed comments:

1. Line 27. Please add references for the examples.

2. Line 49. Could you be more specific about what the definition of "a good environmental status" are?

3. Line 61. Why did the authors study the inter-annual variability instead of other scales like seasonality which should be more apparent than inter-annual signal for many earth systems?

4. Section 3.1. As a model paper, I think it is not enough to evaluate surface T, S, nutrient, and chl-a concentration over such big area, especially as the authors pointed out that water stratification was found as a factor to bottom dissolved oxygen deficiency. Evaluations of T/S vertical profiles, nutrient profiles, oxygen profiles, and chl-a spatial patterns (against satellite estimates) are crucial to validating the modeled results.

5. Figure 2. Does the model represent sea ice dynamics? I found that the modeled water temperature in winter is > 0 ^{o}C. As the authors mentioned at Line 113, the SST observations were not available in winter due to sea ice coverage. So, in reality, there should be sea ice covered the lagoon in winter. Please clarify the model's capability in sea ice modeling.

6. Figure 2 caption. "Modelsimulation" should be "Model simulation".

7. Lines 110-114. More quantitative comparison is needed like providing R^2, RMSE, and relative RMSE.

8. Line 111. As I observed, salinity was overestimated at Station C, but was underestimated at Station KHM between 2004 and 2008. Also, the authors need to provide evidence to support "possibly due to an underestimation of runoff in the forcing". The runoff forcings are from observations. If the statement is true (i.e., runoff is underestimated), it may indicate that the number of river point sources are not enough, or the atmospheric precipitation is underestimated. Please show the evidence to support the causes of underestimation of salinity.

9. Lines 118-120. Why there is such a great difference in transport through the Dziwna between your estimates and others?

10. Figure 3. Is the plot based on monthly average? Please provide the way how the transport contribution is calculated.

11. Lines 125-127. There is a lack of evidence to address (1) that low bottom DO triggers the release of iron-bound phosphate from sediment and (2) that the amount of phosphate released is insufficient to increase the surface concentration.

12. Line 139. I am not convinced that the deeper channel acts as a sediment trap. Instead, I believe water stratification is the primary factor contributing to bottom DO depletion in the main channel. First, surface chl-a concentrations do not show a distinct pattern inside versus outside the main channel (Figure A1e). This suggests that the amount of sinking organic matter should follow a similar spatial distribution, especially in such a shallow lagoon (< 10 m), where organic matter does not drift far from where it sinks. Second, water column stratification plays a crucial role in the development of bottom hypoxia, as demonstrated by numerous hypoxia studies. As the authors later mention, the model does not account for mixing processes due to heavy traffic in the main channel, which leads to discrepancies between the modeled and observed bottom DO concentrations (hypoxia is modeled but not observed in the measurements). This suggests that bottom hypoxia in the main channel may be more influenced by strong (or overestimated) stratification.

13. Figure 4. Statistics like R^2 and RMSE should be provided.

14. Lines 151-157. I think this is the core of this work. The authors may need to provide a diagram for quantification of each source and sink terms for both N and P. I found the Figures 9-10 attempt to address it, but there is a lack of quantification for the P sources and sink terms. Also, it is better to move Figure 9-10 to here.

15. Line 157. As I understand, denitrification occurs at anoxic conditions. That is, denitrification rate at sediment should decrease as oxygen concentration increases.

16. Line 160. Why don't you use the daily mean to increase the sample size? As I observed, the sample size in Figure 7 is small which may weaken the conclusion drawn.

17. Figure 7. The plot read confused to me. The authors mix the output of the control and reduced load experiments when generating this plot. However, the plot includes two types of signals: (1) annual signal of the retention rate which changes as nutrient loads; (2) retention rate changes due to the changing system when nutrient is reduced. The former is the one what the authors want to analysis. However, regarding the latter one, when the total nutrient loads are reduced, the entire ecosystem will adapt to such changes and turn out to be a new system. For example, some phytoplankton species can become the dominate species given their higher adaptation to low-nutrient environment. Such changes may affect the sinking organic matter not

just in spatial distribution but also in temporal phases. So, my suggestion is to plot Figure 7 for individual system (i.e., don't mix the output from the control and reduced loads experiments).

18. Figure 6. Statistic tests are needed to test if the differences in nutrient retention are significant between the control and reduced load experiments. I strongly suggest the authors use daily output instead of monthly mean to increase the sample size.

19. Line 165. Could you please provide the definition of "retention capacity," as it sounds like professional jargon to me?

20. Line 167. "…while the phosphorus retention capacity remains largely independent of load variations". This conclusion is drawn from Figure 7d. However, according to Figure 6b, P relative retention capacity seems to change significantly (need statistic test) when nutrient loads are halved. That is, Figure 6b contradicts Figure 7d. Please also see the comment 17.

21. Section 3.4. What is the purpose of this section? Is that designed to find the minimum resolution that can simulate the retention capacity well enough? If so, you already have the 150-m model and there is no need to try coarser ones.

22. Line 172. Running a 5550-m model may not be meaningful, as the resolution of the parent model is 2000 m. Instead, the authors can add a finer resolution test (maybe at 50 m).

23. Section 4.1 looks like another result section but lacks in-depth discussion. It reduplicates what has been shown in section 3.

24. Lines 185-186. There is no evidence shown to support this statement. Please see the comment 8.

25. Line 193-197. More quantitative analysis is needed to address the contribution of various sink terms to the net fluxes. Such analysis may ask for modification of model parametrization. As shown by DIN validation, the model also failed to capture the peak value in high-DIN period, which may result from the overestimated nutrient uptake rate by phytoplankton.

26. Line 200-202. Please see the comment 12. The authors may need to compare the contribution of sediment oxygen consumption and water stratification to the bottom DO changes.

27. Line 202. Please pinpoint the Grobes Haff in the map.

28. Lines 204-205. This study is not a hypoxia study. If there is a great discrepancy found between modeled and observed DO, then I suggest the author focus more on the nutrient retention.

29. Lines 206-207. This is a very strong statement. I've seen a low-trophic model with 11 phytoplankton functional groups.

29. Lines 208-209. I am not sure if it is true for the Baltic Sea. As I learnt, parameter tuning is usually needed for most ecosystem model when study region is changed due to the changes in multiple ecosystem aspect, like dominate species, lower-trophic complexity, and pollution conditions. So, I would suggest the authors be cautious when saying “without parameter tuning”.

30. Lines 212-216. Isn’t it obvious?

31. Lines 240-241. This may not be true. The authors should test the significance of P changes due to reduced nutrient loads. Also see the comments above.

32. Lines 244-245. To my understanding, it is not correct.

33. Line 245. Usually, at water surface, oxygen decreases as primary production decreases.

34. Figure 10. Are the changes between the control and reduced loads experiments? Please clarify it in the caption.

35. Figures 9 and 10. Need similar plots for P. Please also make the line styles and line colors consistent for the same term in both plots. Please also use the same name for the same terms.

36. Figure 11. Please clarify how the water residence time is calculated in the main text. It is important to show it because there are at least two definitions of water residence time as I know.

37. Lines 268-270. This conclusion confuses me. Do the authors mean that the riverine nutrient loads are mainly control by riverine water discharges rather by riverine nutrient concentration?

38. Figure 12. In (a) N and P loads decrease in recent years. Such negative trends may be contributed by (1) negative trend in riverine water discharges and (2) nutrient reduction actions in recent years. It is very interesting to compare these contributions to see if the human’s efforts in nutrient reduction matter regarding water quality improvement.

39. Section 4.4. I am not sure why the authors are interested in the interannual signal. As shown by Figure 12a, the range of discharges and nutrient loads increase in recent years (e.g., ranges summarized every 5 moving years). It is likely due to the climate changes which cause more extreme events like droughts and floods. So, it would be more interesting to discuss the climate-change induced uncertainty in nutrient loads and the related implication.

40. The conclusion needs to be updated according to the revised contents.

41. Figure A1. Legends and text in the figure are hard to read. The color for the normalized RMSD is hard to see as well. And I believe RMSE (root-mean-square error) should be a more appropriate term to use when comparing model output and observations.