First of all, we would like to thank the referees and a comment from the community for the thorough reviews of our manuscript. We provide our response on how to consider the referee's suggestions in a revised version of our manuscript.

In the following, we respond to the referee's #2 remarks. Remarks are shown in blue and our response in black italic.

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.

We greatly appreciate the consideration of our manuscript and the comprehensive suggestions for its improvement. In the revised version, we will expand the validation section. We plan to extend Appendix A and present a summary of the findings in the main text.

Additionally, the referee has provided valuable advice for enhancing the discussion and conclusion sections. Our detailed responses to these comments can be found in the "specific comments" section.

Detailed comments:
1. Line 27. Please add references for the examples.

Efficiency of the coastal filter: Nitrogen and phosphorus removal in the Baltic Sea. https://doi.org/https://doi.org/10.1002/lno.10644,
Biogeochemical Budgets of Nutrients and Metabolism in the Curonian Lagoon (South East Baltic Sea): Spatial and Temporal Variations. https://www.mdpi.com/2073-4441/14/2/164
Modeling the long-term dynamics of nutrients and phytoplankton in the Gulf of Riga https://www.sciencedirect.com/science/article/pii/S0924796311000704
Modelling nutrient retention in the coastal zone of an eutrophic sea https://bg.copernicus.org/articles/13/5753/2016/bg-13-5753-2016.pdf
Nutrient Retention in the Swedish Coastal Zone https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2018.00415/full
Biogeochemical functioning of the Baltic Sea: https://esd.copernicus.org/articles/13/633/2022/esd-13-633-2022.html

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

We will add a reference for definition.
MSFD:

https://environment.ec.europa.eu/topics/marine-environment/descriptors-under-marine-strategy-framework-directive_en
In comparison to the good ecological status after WFD:
https://www.eea.europa.eu/en/analysis/indicators/ecological-status-of-surface-waters
Schernewski et al. (2015),
https://www.sciencedirect.com/science/article/pii/S0308597X14002358?via%3Dihub

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?

The seasonal signal is certainly more pronounced. However, for nutrient export to the Baltic Sea and its impact on the environmental status of the Baltic Sea, total loads over longer time scales are important. Here the focus is on the long-term variability. In a recent paper we studied the effects of shorter-term variability:
Ecosystem-Model-Based Valuation of Ecosystem Services in a Baltic Lagoon: Long-Term Human Technical Interventions and Short-Term Variability https://www.mdpi.com/2076-3298/12/2/35

In this study, it became obvious that the present temporal resolution of the input data, especially the available monthly load data of the Oder River limits the accuracy of the model hindcast. This is especially true for the Oder Lagoon, since it is strongly controlled by external Oder River loads. This analysis requires a modified model approach, a detailed field data analysis and a spatial analysis within the lagoon. This goes beyond the possibilities of this paper. Therefore, a separate paper is in preparation that studies the spatio-temporal seasonality in the lagoon and especially the role of extreme events such as droughts and floods as well as hot seasons.

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.

We will extend section A (appendix) with a detailed validation including vertical profiles and additional state variables.

Satellite products with a sufficient spatial and temporal resolution are only available since 2024 (https://data.marine.copernicus.eu/product/OCEANCOLOUR_BAL_BGC_L3_NRT_009_131/description), a time period not covered by the model. Further, satellite products are struggling with coastal waters like the Oder Lagoon, due to the higher turbidity.

The Oder Lagoon shows a strong patchiness of phytoplankton with respect to vertical location and small-scale horizontal distribution of phytoplankton. Satellite data can provide only a limited insight, has very limited absolute reliability and will be taken into account as soon as we study spatial effects in the lagoo, in detail.

5. Figure 2. Does the model represent sea ice dynamics? I found that the modeled water temperature in winter is > 0 oC. 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.

Indeed, we use a coupled sea-ice-ocean model. We will mention this in the model setup section.

6. Figure 2 caption. “Modelsimulation” should be “Model simulation”.

We will correct this.

7. Lines 110-114. More quantitative comparison is needed like providing R2, RMSE, and relative RMSE.

The revised validation section will include this analysis.

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.

We will attempt to evaluate the reasons for the discrepancies. However, conducting extensive sensitivity studies is not feasible. We consider these discrepancies to be relatively minor and they do not affect the vertical column properties, which we will present in the revised manuscript. If we cannot confirm that runoff uncertainties are the main reason, we will remove this statement.
Since salinity is controlled by the water exchange with the Baltic Sea via the deep (10.5 m) Swina Channel, the results in the lagoon depend on external forcing (Baltic Sea model) as well as the assumed connectivity between both systems. Single inflow-events play a major role for salinity in the lagoon. Location C is in the channel and it can be assumed that regular chip traffic plays a role in mixing saline water along the channel. However, we consider the agreement between data and simulation of salinity as sufficient for this study, since the differences have only negligible effects on nutrient cycling and pelagic ecology.

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

In reality, the Dziwna channel is a long and shallow channel until it reaches the Baltic Sea. We did not include the appendix of this channel in our model. We will note this in the revised version.

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

*It is calculated as usually, velocity*rho*area. This calculation is done online, that is, for each model time step. The monthly average is performed afterwards.*

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.

We will demonstrate it in more detail in the validation section of the revised manuscript.
During our study, we observed that the model was with respect to some parameters not well in agreement with the monitoring data, especially when it comes to short-term ecological effects. One aspect is hypoxia and the other, related aspect, are the sudden summer peaks of inorganic phosphorus.

Mass balances clearly indicate that the P-peaks have to result from a release from the sediment, very likely under anoxic conditions (Fe bound P), the so-called internal eutrophication. This process is strongly influenced by the pollution history of the lagoon, namely the amount of Fe-P stored in the sediment at the beginning of the simulations 1995 (5-10 years after the pollution peak). This requires a different approach. We decided to tackle these complex aspects of anoxia and internal eutrophication in a separate paper and here focus on long-term developments.

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.

The navigation channel acts as a sediment trap. Sedimentary material is continuously (by animals) or as an event (storm) re-suspended, transported horizontally by currents, and eventually ends up in deeper areas such as the navigation channel. (See figure below.)

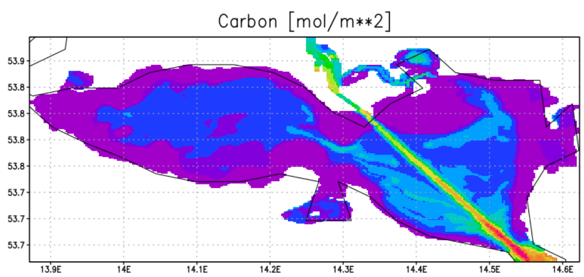

Figure 1: Carbon content in model sediment Sep. 2005

There is little literature assessing the effect of dredging in the channel:
Ecosystem Services Supporting Environmental Impact Assessments (EIAs): Assessments of Navigation Waterways Deepening Based on Data, Experts, and a 3D Ecosystem Model [*https://www.mdpi.com/2073-445X/13/10/1653*](https://www.mdpi.com/2073-445X/13/10/1653)

Unfortunately in German only, a diploma thesis compiles data and makes a budget of dredging which clearly demonstrates the "sediment trap" property of the channel.
Minning, M. Der Schifffahrtskanal im Oderhaff. Eine Sediment-, Nähr- und Schadstofffalle?.Diplomarbeit. Christian-Albrechts-Universität, Kiel, 2003. Available online: https://eucc-d-inline.databases.eucc-d.de/files/documents/00000695_Diplomarbeit_Minning2.pdf.

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.

For anoxia both requirements are needed: (1) Oxygen consumption, in shallow water mainly by sediments, and (2) prevention of oxygen supply for example due to stratification. This is the case in our model. In the real system, heavy ship traffic to and from Sczezin harbor exists. These ships have a draught close to the floor of the navigation channel which regularly mixes the water column. This process is missing in the model.

For frequency of ship trafic see:
Ecosystem-Model-Based Valuation of Ecosystem Services in a Baltic Lagoon: Long-Term Human Technical Interventions and Short-Term Variability *https://www.mdpi.com/2076-3298/12/2/35*
About 3300 cargo ship arrive in Szczecin harbor per annum, which are about 18 cargo ships per day.

13. Figure 4. Statistics like R2 and RMSE should be provided.

See our response to 7.

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.

We will think about a revision of figures 9-10. Currently, we think about an additional table.

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

Denitrification occurs around the redoxcline. In the case of oxic bottom water, the redoxcline is located in the sediment. The denitrification at the sediment redoxcline is much more efficient as in the water column. The small spatial distance supports the coupled nitrification-denitrification.

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.

Unfortunately, river load data are on a monthly basis. These data are needed for the relations with loads and relative retention.

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).

We will use different colors for the experiment and analyze whether the results will change. However, we will explain the figure and consequences in more detail as the referee suggested.

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.

See our comment to 16. We will improve the statistics.

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

Retention capacity (in this context) refers to the ability to retain nutrients within a system. This term is also used in other contexts, such as the retention of water in soils or energy in a battery. We use it synonymously with filter capacity or function.

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.

See our comment to 18. We will improve statistical analysis.

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.

We will remove section 3.4.

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).

See our comment to 21, and we will not setup a 50m model.

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

We will revise sections 3 and 4.1 to reduce repetitions.

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

See our comment to 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.

We show our model results with reasonable care. Further "sensitivity studies" are not productive and beyond the scope of this study.

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.

See our comment to 12.

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

Will be done.

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.

DO is reproduces by the model fairly well which we will show. An Exception is the navigation channel for reasons we explained in our comment to 12.

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

In our opinion, model complexity depends on the scientific question.

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".

We document what we did. We will replace "open sea" with "Baltic Sea" which we simulate with the same model. The referee has a point; an altered model structure may require a re-calibration. However, we did not change the model.

30. Lines 212-216. Isn't it obvious?

Certainly yes, but usually one have to make a compromise between quality and costs. Our intention was to give some guidance for larger scale models. However we will remove section 3.4.

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.

See our comment to 14.

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

See our comment to 15.

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

The sediment-water interface is meant, line 244-245.

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

Yes, the caption could be more precise.

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.

We thank the referee for this hint.

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.

The total volume of the lagoon was divided by the river discharge. This simple calculation follows the international lake approach going back to Vollenweider (1976) and many subsequent publications. It allows the estimation of critival loads.
Vollenweider, R. A. (1976). Advances in defining critical loading levels for phosphorus in lake eutrophication. Memorie dell'Istituto Italiano di Idrobiologia, 33, 53–83.

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.

Yes, thank you, we have to rephrase it, to better make clear that the concentration of both nutrients, N and P, is largely independent from river discharge. This has been observed already earlier and is the reason why the long-term assessment of critical loads used a discharge correction/normalization (e.g. Friedland et al. 2019). However today, we see the tendency of a climate change induced generally reduced annual water discharges. This means in recent years, your point (1) is more important than (2), which dominated the load reductions in the 1990's. We can elaborate a bit more on it, because a tendency to more extreme floods can partly counteract that.

Friedland, et al. 2019: https://doi.org/10.3389/fmars.2018.00521,2019.

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.

See earlier comment: yes, therefore, a separate paper is in preparation that studies the spatio-temporal seasonality in the lagoon and especially the role of extreme events such as droughts and floods as well as hot seasons.

We show and discuss that high and low runoff adequately modifies nutrient loads. A relation to climate warming cannot be established from our relatively short simulation period.

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

Yes.

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.

We agree that the colors in Figure A1 are hard to distinguish and will adjust the figure in the revised manuscript. Using the normalized RMSD instead of the RMSE gives us the possibility to make the numbers comparable between the station and parameters. Further, the normalized RMSD gives us a good measure for the model skill, as values below 1 indicate that the standard deviation of the observations is higher than the RMSE, meaning that the model results stay within the natural variability.