Review #1

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 #1 remarks. Remarks are shown in blue and our response in black italic. Changes made in the manuscript are in red.

In general, the manuscript is well written, well presented and clear, including mathematical expressions and equations. It reaches innovative and robust conclusions that can greatly improve model studies in the area. Especially as little is known about coastal retention in the Baltic Sea. It is also relevant for policy measures, because official nutrient reductions do not account for nutrient retention in coastal waters. However, the conclusions could be more specific and, with additional discussions, could be more relevant for the entire Baltic Sea. While their methodology is valid, clear and relevant, it is difficult to be applied by any other institute studying the Baltic Sea's biogeochemistry, as setting a coastal model for one single region is costly (manpower, budget, time) and not all may have that possibility. However, by expanding their analysis further and adding more discussions on how the Order Lagoon compare to other coastal areas in the Baltic Sea, approximations for N and P retention could be estimated for other areas as well. Published results for the Baltic Sea mentioned in the paper have large discrepancies with specific coastal studies, and therefore new suggestions will be relevant to address including approximations for other coastal areas, especially for those similar to the Oder Lagoon. Their conclusions do mention that the method can be adapted for other regions beyond the Baltic Sea, but as written now is too vague. If this means adapting their model to a new area, it could translate as a costly task (time and computational) and not necessarily straightforward. In this sens, a better clarification on what is possible to extrapolate to other areas with this method is required. In other words, the main missing discussion in this manuscript is: Can the relation found in this study between loads and nutrient retention be used to estimate %retention in other basins? What can be used from these results to improve retention estimates in other coastal areas in the Baltic Sea and other Seas?

The Scientific significance of this study is good and after some revisions, will contribute significantly to our understanding of nutrient retention in coastal waters. The latter is poorly known and therefore, the content of this manuscript is very valuable. The methodology is clear and results relevant to biogeochemistry.

The scientific quality of the study is fair, the methods and approach are valid and innovative. However, it misses additional discussions, described in the specific comments below.

The presentation quality is good, the structure and figures are clear and well explained. However, it misses some additional figures, for example, in the validation section time series of model and observed oxygen should be shown, as well as more on the water column dynamics in the model. The oxygen section can also be expanded, as it is a major parameter affecting both N and P.

Section 4.4 should be expanded or moved to the results section. For details, see specific comment below.

In short, I would strongly recommend this manuscript for publication after minor/medium revision.

We greatly appreciate the positive 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.

We have extended the validation section and included additional data from both observations and model simulations. The comprehensive validation section has been moved to the appendix. In the main text, we provide a summary of the validation and refer readers to the appendix for further details.

We have added a paragraph in the discussion (Section 4.2) to share our perspective on the extent to which our findings could be applicable to other regions.

Additionally, we have expanded the discussion in Section 4.4 to offer a broader perspective on upcoming changes in the Oder Lagoon.

Specific comments

Introduction

line 21: What is meant by extended residence time? Long? If so, replace by “long”

“long” seems to be the right term.

Changed into “long”.

line 26: Please expand and add references. Why are there only a few coastal waters that can reduce nutrient loads in the Baltic Sea? Why not all coastal areas have that capacity? The Baltic Sea has many different types of coastal waters, please mention them as well and why would they be less effective in nutrient retention.

Asmala et al., 2017, which we refer to, give a good classification of different Baltic Sea coastal types and potential filter function. We will summarize the Asmala et al. findings in this context.

We introduced a paragraph on different coastal types and their retention capacity.

Line 28: Why is the Order Lagoon the most critical?

“likely the most critical” We based this statement on Asmala et al. 2017, who identified the Oder Lagoon as the one with the highest denitrification rates. We will make it more clearly in a revised version.

More references have been included supporting this statement. The most pronounced function is removal of DIN due to denitrification depending on residence time.

Line 48: Are the MAI here referred per basin in the Baltic Sea or to the entire Baltic Sea?

We refer to MAIs in general, since they neglect a possible coastal filter. MAIs are based on loads into the "central" basin.

It has been clarified in the text.

Line 55: Models can account for nutrient retention in the Baltic Sea taking values from literature and making assumptions for unstudied coastal areas. They can give good enough nutrient concentration near coastal areas, but can certainly be improved. The problem is not that they do not adequately account for it, but that there is not enough information to adequately account for all different types of coastal areas of the Baltic Sea. In this regard, this study can really improve such model results. Here, I would suggest to focus more on the fact that there is missing information on this for the Baltic Sea (and other Seas) and discuss this further. Importantly, this has been already mentioned in previous modeling studies, for example in:

Eilola et al., 2011, they mention that "the major differences between the nutrient supplies to the different models are due to the different assumptions on the bioavailability of phosphorus loads", referring to the retention of P in coastal waters. They also have a dedicated chapter on bioavailable nutrient loads in their discussions, which should be mentioned in this study.

Ruvalcaba-Baroni et al., 2024, they refer to the little information available and what can be currently used: "As the response to nutrient removal of different coastal types is poorly quantified... these factors are taken from previous studies in the Baltic Sea (Eilola et al., 2011; Edman and Anderson, 2014; Asmala et al., 2017)." In their discussion, they also mention that "one factor affecting detritus is the fraction of the organic matter coming from rivers that is actually bioavailable and not directly retained in coastal waters." They assume a constant fraction for the entire Baltic Sea and say that "the input of organic matter from rivers, especially nitrogen, could be improved by better accounting for river-specific organic matter retention in coastal waters".

Adding this discussion, will make the point of this study stronger, as it "partly" answers a problem that has been existing since many years.

We would like to emphasize the importance of distinguishing between the non-bioavailability of riverine organic matter (OM) and the bioavailable OM and nutrients that are retained or filtered in coastal zones. This distinction is crucial at the process level, as these components respond differently to changing boundary conditions. Indeed, the understanding of nutrient load transformation and retention between monitored rivers and export from the coastal zone remains limited.

Altogether, the referee raised an important point that we will address in the revised manuscript.

We added a paragraph in the introduction which point to the problem of "purely quantified" nutrient retention in coastal waters as the referee suggested.

Results

line 112: The authors mention that “the model fails to reproduce the low salinities observed between 2004 and 2008, possibly due to an underestimation of runoff in the forcing”, but do not show the forcing used. What evidence do you have for this? It could also be due to the atmospheric forcing (precipitation vs evaporation). Please add more explanation for this, a supporting figure with, for example the runoff used in the high resolution model and how it compares to observations could be helpful here (or in supplementary). Also, why not show salinity and temperatures profiles or some water column performance?

In the extended validation section (see response to general remarks), we will also consider vertical profiles. The suggestion that runoff is the cause is more of an educated guess. The meteorological forcing data comes from a consistent reanalysis, while the runoff data is based on observations from the Oder River. The model reproduces similar low salinities, for example, in 1999 and 2010. Therefore, we concluded that the runoff during the winter seasons of 2004-2008 is underestimated. We will attempt to find more robust arguments to explain the elevated salinity. Additionally, we will present the river data used in our analysis.

We have added an extended validation section and discussed the reasons for model biases more thoroughly. Runoff was previously illustrated in the former Fig. 12a. However, we have introduced an additional runoff figure. The Oder River accounts for 98% of the river discharge. Atmospheric precipitation minus evaporation (P-E) is on the order of 1% of the discharge. Both freshwater sources (precipitation and evaporation) have negligible impact on salinity.

Observations are only available for the surface and near-bottom layers. Therefore, we are unable to present observational profiles. The mixed layer depth derived from model simulations is presented and provides insights into stratification properties.

Line 114: The authors mention that “However, this discrepancy has yet to be confirmed”. This statement falls a bit short. How would you confirm this and why not doing it in this manuscript? Please rephrase or add information.

We believe that we cannot confirm this in the manuscript. Observations are available at a monthly resolution at best and may miss extreme values. We will rephrase this statement accordingly.

The new validation section (appendix) discusses model biases and possible reasons in more detail.

Line 123: The authors say “the model successfully captures the decreasing trend in winter phosphate”. Plotting the trend directly in the figure for both model and observations would make this point more clear. It will also be relevant to plot it for the other biogeochemical parameters. The significance of the trend can then be estimated, which, for example, it is not entirely clear for chlorophyll-a at KHM (is the decreasing trend significant or not?). This information would be valuable for the discussions later on and for line 129, where the model and observed chla are described. From Fig 4, it looks like the model has also a more pronounced decreasing trend than the observations at KHM, but very good match at C.

In the extended validation section, we will include trend lines and statistical estimates. The referee is correct; providing this information will strengthen the discussion.

We added a trend line in the time series figure if the trend is significant.

Line 131-132: The authors mention that "phytoplankton response to changing ambient light conditions and (???) may vary between 23 and 60". I think the authors missed writing "Chl:C ratios". Please add the missing part. If so, would the Chl:C ratio be that different between C and KHM? Would this alone explain the discrepancies in model performance between C and KHM?

We will re-phrase the sentence.

We corrected the sentence referring to Chl:C ratios. Performance discrepancies between stations C and KHM, we cannot explain beyond some speculations.

Oxygen dynamics (section 3.2)

This section is very relevant to the manuscript. While the model shows oxygen depletion, there is missing evidence on how depleted the oxygen is in the area. Why not add oxygen time series in bottom waters from both observations and model? Is there any oxygen measurements in this area?

Observations for oxygen are available. However, near-bottom observations are affected by the measurement process itself. Factors include how close the oxygen sensor can be lowered to the floor and whether the water column is disturbed by the measuring platform (vessel). Experimental evidence supporting these limitations is provided by Fredriksson et al. (2024). We will present both simulated and observed bottom oxygen concentrations, acknowledging that the observations may not accurately represent the true near-bottom oxygen levels.

Available oxygen data are presented in the new validation section. However, observations near the bottom suffer from the measurement technique which we discuss as well.

This requires further discussions either to show more evidence for oxygen depletion or mention the lack of observations and what are the alternatives to "validate" the model in such case. What are the references for oxygen depletion in this area?

This is also relevant for the statement that the authors write in line 126 "..., the amount of phosphate released is insufficient to significantly increase the surface concentration." This is linked to oxygen, but also to mixing. Neither are shown. Water column dynamics can be shown in different ways (e.g., salinity and temperature profiles, mixing water depths, etc). It is difficult to judge how well the model represents the oxygen concentration in bottom waters or phosphorous retention in the sediments. It may as well be that the release of phosphorus is not insufficient in the model, but that the lower values in the surface are due to a biased runoff forcing that stratifies the water column too much in the model. Please add water column dynamic information (it can be in the appendix) and evidence for low oxygen concentrations in bottom waters from observations (e.g., time series for C and KHM of bottom oxygen).

We present oxygen from bottom water observations and discuss the inherent difficulties using these data.

In figure 5, an additional map showing the actual model oxygen concentrations in bottom waters for the months with oxygen depletion and available observations will be valuable.

We show bottom oxygen in the appendix for both available stations.

Line 164: Phosphorus retention in the sediments is also affected by bottom oxygen concentrations. It will be interesting to plot retention rates vs oxygen concentrations in bottom waters, especially for P. Could this (partly) explain the lack of relationship between rates of retention and P loads?

A high retention rate could impact the phosphate release during anoxia. However, too strong retention on the other hand would yield too low phosphate concentrations during oxic conditions.

Unfortunately, we do not have the retention rate directly in our diagnostic model output. Instead, we show the phosphate liberation together with near bottom oxygen.

Lines 199-205: Relates to comment above on Oxygen Dynamics. This paragraph does not have any reference. How do we know the Order Lagoon experiences anoxia? Oxygen is a relevant topic for this type of study. Please expand the discussions on this and evidences for oxygen depletion available for the Order lagoon.

Aim of the model application was to show that the model sufficiently well represents the seasonal dynamics and the long-term development in the Oder lagoon. However, during our study, we observed short-term ecological effects where the model was not well in agreement with the monitoring data. 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. The monthly monitoring data taken one meter above the ground does not indicate hypoxia. On the other hand, observed mussel and fish-kills, beside internal eutrophication, are clear indications for hypoxia and even anoxia. We decided to tackle this complex short-term problem in a separate paper, that is presently under preparation, because of all the implications for management and ecosystem assessment. Scientific literature indicating hypoxia and internal eutrophication in the Oder Lagoon does not exist.

Here we would expand the discussion and refer to the observed mussel and fish-kills as well as add literature from comparable lagoon systems that underpin the likelihood of hypoxia and internal eutrophication.

We will show water column dynamics and discuss possible consequences.

We added a figure showing model stratification and phosphorus release from the sediment.

Line 175: How can the higher resolution model overestimate nitrogen retention?

In the coarse-resolution model, the near-bottom oxygen concentration is overestimated. This enhances the coupled nitrification-denitrification process in the sediment, consequently increasing nitrogen retention.

Impact of model resolution is postponed to another study.

Discussion

Line 185: The discussion relates to runoff forcing and would benefit from a runoff figure (as mentioned in comments above for line 112).

As responded to line 112 comment, we will show river data.

Runoff has been shown already in former Fig. 12a. However, we introduced another runoff figure.

Line 195: See comment above on Oxygen dynamics (section 3.2). This discussion should be expanded.

Additional results from the extended oxygen analysis will be discussed.

Oxygen concentration and its impact on phosphate release is discussed in the discussion section and the validation section.

198: There is much more literature available for P recycling in the Baltic Sea that can help tuning the model. Please add more references.

We thank the referee for this advice.

We indicate that the 2D sediment module's memory is too short to remember the eutrophication period over a longer time and to respond with high phosphate liberation to anoxia. Instead, it stores the history of a few years. Thus, it is not a matter of parameterization but of the model structure.

Line 205: So there are observations? Why not show them?

Some are available but with the restrictions discussed in Fredriksson et al. 2024.We will show them.

We show them now.

Line 220: Pleas add "As done in Eiola et al., 2011 and Ruvalcaba-Baroni et al., 2024, where they use a factor to account for nutrient retention" or similar (see comment above for line 55).

We can do so, but we should be aware that bioavailability and retention/filtration are distinct processes.

Introduction has been extended with the known uncertainties in availability of riverine loads.

Line 221-222: Please expand. This has more potential and general suggestions on how to improve model estimates in the Baltic Sea can be given based on the relevant results in this manuscript. The big picture would be to say something on how these very nice results (based on chapter 4.2, 4.3 and 4.4) can be used by the larger scientific community and perhaps even policy. This could be discussed in a chapter on its own (4.5?).

We will consider including an additional section.

We introduced a new paragraph in section 4.2 summarizing relevant fidings for retention and transferability to other regions.

Line 222: As mentioned before, not all coastal areas can be modeled the way it was done in this manuscript (it will be nice, but it will take time). Please expand on how your results could differ in other areas, based on bathymetry, residence time, oxygen content, etc and how they relate to published values (some of which are already mentioned in the introduction, but there are a few more, see discussions in Eilola et al., 2014). Or is it that every single coastal area in the Baltic Sea needs to be modeled separately to be able to estimate their coastal nutrient retention? Perhaps, but some relationships must be similar.

We believe that the most crucial coastal regions can be simulated with reasonable effort. Asmala et al. (2017) provide a useful classification for determining which regions are most important. We will address this issue in the discussion.

We discussed our opinion on what could be transferred to other regions and which coastal types require additional studies.

Line 240: There is no statistics shown in the paper, how do you know it is “not statistically significant”?

We did this analysis and will show it in a revised manuscript.

We show the statistics now.

Line 258: This is an important statement that could be repeated in the conclusion.

Will be done.

This part of the conclusions now.

Line 291: Please replace “.” by “,”

Will be done.

Done.

Interannual variability of discharge and loads and its consequences (4.4)

This entire section is structured more as results than as discussions. Please expand (add discussion and references) or move to results.

The referee is correct. We will extent the discussion on this topic.

Section 4.4 is extended by an additional paragraph.

Line 308: Please replace “emissions” by “inputs”

Will be done.

Done.

Line 310: How do you know that interannual variability is likely controlled by extreme events? Is there a reference or some evidence? It may seem obvious, but the relative or quantitative impact remains quite unknown for the Baltic Sea, as far as I know. Also what do you mean by “extreme events”? Storms, heatwaves and/or what else?

We will clarify the type of event we are considering. A detailed analysis is beyond the scope of this study and is the subject of our ongoing research.

We addressed the kind of extremes and possible impact in the additional discussion on this topic.

Line 315: How easy/difficult would it be to expand it to other regions in and outside the Baltic Sea? Please clarify.

As we have already argued, we believe that additional studies can be conducted with reasonable effort. A prerequisite is the availability of reliable runoff data, which are accessible for most regions of the Baltic Sea.

An additional paragraph on our opinion of transferability in section 4.2 is added.

320: I am not sure that the word “realistically impossible” can be used in this study, as it does not assess economical impacts and costs. Maybe rephrase to something less strong.

We will consider this advice.

We rephrased it to “unrealistic”.

Figures

Figure 1. It would be good to mention already in the caption the navigation channel. It is quite obvious that there is a channel, but is not redundant to mention it. It would be also interesting to know when was this channel made (before the 60s?) to have an idea on how far back in time we should go to get “pristine conditions” in this area and to better interpret the observational data sets in this area as it seems to have a large impact for nutrient retention.

We will modify the figure, give additional information of the channel and provide references. Modifications to the Swina River date back to 1721 and in 1880, a shortened and deepened artificial channel was completed. Subsequent, deepening projects included an increase to 9.6 meters in 1939 and to 10.5 meters in 1984 across the entire lagoon. Between 2018 and 2023, the entire waterway across the Oder Lagoon was deepened to 12.5 meters (Schernewski et al. 2025, https://doi.org/10.3390/environments12020035).

We modified the figure and text accordingly.

Figure 5. Please add the period mean in the caption.

We will add this information. The mean annual of anoxic days is based on the years between 1995 and 2019.

We added this information.

Ref.:

Fredriksson et al., 2024: https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.12607

Schernewski, G.; Neumann, T.; Piehl, S.; Swer, N.M. Ecosystem-Model-Based Valuation of Ecosystem Services in a Baltic Lagoon: Long-Term Human Technical Interventions and Short-Term Variability. Environments 2025, 12, 35. https://doi.org/10.3390/environments12020035

Review #2

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. Changes made in the manuscript are in red.

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.

We have extended the validation section and included additional data from both observations and model simulations. The comprehensive validation section has been moved to the appendix. In the main text, we provide a summary of the validation and refer readers to the appendix for further details.

We have added a paragraph in the discussion (Section 4.2) to share our perspective on the extent to which our findings could be applicable to other regions.

Additionally, we have expanded the discussion in Section 4.4 to offer a broader perspective on upcoming changes in the Oder Lagoon.

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

We added several references.

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

We added references for the definition of the "good environmental status".

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.

An analysis of short-term and seasonal variability will be addressed in a future publication. The focus of this study is on the lagoon's nutrient filtration function, which operates on a longer time scale.

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 lagoon, in detail.

The validation section has been significantly expanded and moved to the appendix. Satellite imagery products, particularly for chlorophyll, are not available in reasonable quality for the inner coastal waters of the Baltic.

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.
The sea ice model is noted in the model description.

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

We will correct this.

Done.

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

The revised validation section will include this analysis.
The validation section provides these quantities.

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.

By accident, we loaded incorrect observations for Station C in the evaluation program and have revised the figure accordingly. We would like to thank the referee for their critical review. It now appears that the model underestimates salinity on some occasions at both stations.

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.

We have addressed this issue in the text and recommend incorporating a realistic long channel in a revised model setup.

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.

We show the relation between bottom oxygen concentration and liberation of phosphate from the sediment and hypothesize that our two-dimensional sediment module cannot store enough phosphorus for a later release.

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.)

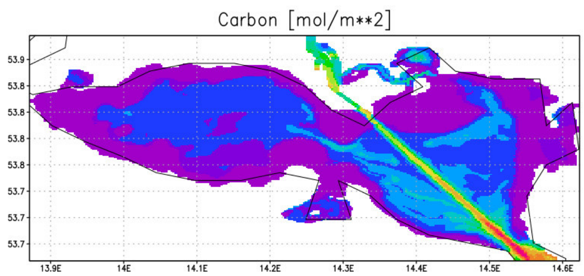

Figure 1: Carbon content in model sediment Sep. 2005

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

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.

We added a reference referring to continuous dredging in the channel.

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.
Comprehensive validation statistics are 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.

We will think about a revision of figures 9-10. Currently, we think about an additional table.
We revised the figure and added a table of sources and sinks for N and P.

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.
Figure 6 demonstrates the contribution of pelagic and sedimentary 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 the 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.

Former Fig. 7 (Fig. 5) is updated and show color coded the two different model 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.

See our comment to 16. We will improve the statistics.
A statistical evaluation is added.

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.
Statistical analysis shows that the relative P retention is independent on loads which is given in the revised manuscript.

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.
Section 3.4 is removed from the manuscript.

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.
We revised section 4.1

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

See our comment to 8, and what we did in response to 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.
We show our model results with reasonable care. Further "sensitivity studies" are not productive and beyond the scope of this study.
In addition to our comment above, we do not see that the model fails predicting DIN.

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.
We present stratification data and phosphate liberation (a proxy for oxygen deficiency) at station KHM. Unfortunately, temperature data are unavailable for station C, which hinders stratification analysis at this location.

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

Will be done.
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.
Bottom oxygen concentrations are presented in the validation section. We discuss uncertainties stemming from both measurement limitations and the potential impacts of ship traffic.

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.
Section 3.4 is removed from the manuscript.

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.
We proofed the statistics.

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.
Caption has been modified.

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.
The figures have been modified to maintain consistent style and color schemes. Given that phosphorus has only one sink, a separate figure similar to that for nitrogen was deemed redundant. Instead, a table has been introduced (Table 1).

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.
We clarified how residence time was estimated.

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?
Indeed, this represents our conclusion regarding the interannual and multiyear variability of nutrient loads. However, it is important to note that this conclusion may not apply to long-term perspectives, particularly when significant changes in the catchment begin to take effect.

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.
As clarified in our response to this comment above, we have addressed the raised concern regarding load changes.

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.
We have incorporated a paragraph discussing the potential impacts of climate change. However, a comprehensive analysis of the effects of extreme events and climate warming will be the subject of a future study.

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

Yes.
The conclusion is updated.

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.

We use a modified figure in the revised manuscript.