Response to anonymous reviewer comments RC1

Please see our responses to each comment in blue below.

General comments

The manuscript investigates the effects of benthic fauna on the biogeochemical transformations of major elements in the Baltic Sea via direct (metabolic) and indirect (bioturbation) pathways, by coupling a recently developed model of benthic fauna to an established hydrodynamic-biogeochemical model. Within regional ecosystem models, benthic components are often represented in a simplified way, where their role is largely limited to returning remineralised nutrients back into the water column. In the Baltic Sea, models accounting for indirect impacts of benthic fauna have been used for a while now, incentivised recently by the need to study the effects of bioturbation by invasive species. However, these models did not consider benthic faunal biomass explicitly, and thus did not allow a separation of effects of bioturbation and direct effects of faunal metabolism on benthic-pelagic dynamics. This manuscript bridges this gap and provides a detailed look at the role benthic fauna plays in shaping biogeochemical fluxes in the Baltic Sea. This study builds on the authors' previous work and contributes to the advancement of benthic system modelling, which has lagged behind its pelagic counterpart in coupled regional models. Although it is an important step in the right direction, I have several criticisms I would like to bring to the attention of the authors. Foremost they are related to the "large-scale" approach taken for this work. It should not serve as a substitute to thorough comparison of model results with data, nor to the detailed presentation and discussion of results (providing quantitative estimates and aligning text narrative with figures). Limitations of the model, unique contributions of this work to the understanding of the Baltic Sea biogeochemistry beyond state-of-the-art knowledge should be appropriately discussed. Below are my detailed comments and questions, that I hope will improve the quality of the manuscript.

We thank the reviewer for the comprehensive and detailed review of our manuscript. Based on the suggestions, we feel the manuscript has been significantly improved. We hope that our modifications of the manuscript detailed below, including additional validation and extended results and discussion, satisfies the reviewer.

Specific comments

Title

1. The title should clearly indicate that it is a model-based study.

We changed the title, removing the slightly ambiguous term "large-scale" and adding "modelling". The new proposed title is:

Modelling the effects of benthic fauna on carbon, nitrogen, and phosphorus dynamics in the Baltic Sea

Abstract

1. It would be informative to add quantitative estimates of changes due to direct and indirect contributions of benthic fauna. "Small proportion of seafloor organic stocks", "decreases denitrification", "increases P retention", "reduction in N fixation" etc all should

be accompanied with quantitative estimates (% change). The way information is currently presented, it is hard to see what a unique contribution of this work is to the understanding of biogeochemical impacts of benthic fauna in the Baltic Sea, as the effects of bioturbation *via* sediment oxygenation on N, P dynamics in the Baltic Sea are generally well known from previous studies.

We added quantitative estimates to the abstract as suggested.

2. Line 16: It should be clear from the text that bioturbation affects denitrification, P retention etc *indirectly*, e.g., increasing oxygen penetration depth and availability.

We modified the sentence to clarify this:

"Further, through enhanced sediment oxygenation, bioturbation decreases benthic denitrification and increases P retention, the latter having far–reaching consequences throughout the ecosystem."

3. Line 20: "chain of indirect effects" is opposed to "direct effects of faunal respiration, excretion and bioturbation". This create some confusion over which effects are "direct" and which are "indirect" ones.

We realize the terms "direct" and "indirect" were used in a confusing way, and now reserve them for referring to metabolism as direct and bioturbation as indirect effects. Consequently, we modified the sentence to avoid confusion: "This chain of effects through the ecosystem overrides the local effects of faunal respiration, excretion and bioturbation." We made corresponding changes throughout the manuscript.

Introduction

1. Lines 43-44. "The combined effects of animal bioturbation and metabolism have seldom been studied together." Several references support this statement, but could the authors elaborate on what are the main reasons for that gap in research, despite recognition of the importance of benthic processes?

2. Could you provide a more solid justification for your approach? Why do we need to implement modelling – does it address the knowledge gap? Why did biogeochemical models of the Baltic Sea so far did not include benthic fauna?

3. When specifying reasons for using Baltic Sea as a model area, to what extent also relatively simple benthic community composition play role?

Response to comments 1.-3:

Several recent reviews and perspectives discuss the gap in research mentioned in the first comment, which essentially is due to the different research traditions, including different foci, scales and assumptions in the several individual fields studying benthic processes (Ehrnsten et al., 2020b; Lessin et al., 2018; Middelburg, 2018; Snelgrove et al., 2014, 2018). We opened up some of the reasoning, and provide additional justification for why we need mechanistic modeling, why benthic fauna has not been included in previous models, and, as suggested in the third comment, why the Baltic Sea is and ideal system to develop this kind of model. Please see an extract of the introduction **with additions in bold** below:

“Even though the importance of benthic fauna for sediment biogeochemistry and benthic–pelagic fluxes has long been recognized (Rhoads, 1974), the combined effects of animal bioturbation and metabolism have seldom been studied together (Ehrnsten et al., 2020b; Middelburg, 2018; Snelgrove et al., 2018). **A long-standing assumption in biogeochemical sediment research is that animals contribute considerably to transport of solids and solutes through bioturbation, but their consumption of organic matter is of minor importance (Middelburg, 2018). However, several studies show that this assumption does not hold in many shallow coastal systems, as recently reviewed by Middelburg (2018) and Ehrnsten et al. (2020b).**

Further, empirical studies of faunal effects often focus on temporally and spatially limited parts of the system, omitting important interactions and variability occurring in natural ecosystems (Snelgrove et al., 2014). **It is logistically challenging to study multiple drivers and interactions in the benthic and pelagic realms, such as the interactions between benthic and pelagic production, empirically. Mechanistic or process-based models are powerful tools to conduct such studies (Seidl, 2017).** Here, we extend a physical–biogeochemical model of the Baltic Sea ecosystem (BALTSEM; Gustafsson et al., 2014; Savchuk et al., 2012) with benthic fauna components based on the Benthic Macrofauna Model (BMM; Ehrnsten et al., 2020a). We include both the direct feedbacks from animal growth and metabolism and the indirect effects of their bioturbating activities on biogeochemical cycling to evaluate their relative contributions.

We use the Baltic Sea as a model area for **three** reasons: (i) the shallow depth (mean depth 57 m) and enclosed geography with a long water residence time (about 33 years) contribute to strong benthic-pelagic coupling (Snoeijs-Leijonmalm et al., 2017; Stigebrandt and Gustafsson, 2003), **thus pelagic nutrient dynamics are highly dependent on benthic processes, (ii) the relatively simple, species-poor benthic communities facilitate model development,** and (iii) the major features of biogeochemical cycling of C, N and P in the Baltic Sea are well known due to a wealth of oceanographic measurements and studies performed over the past century, making it an ideal system for process-based modelling (Eilola et al., 2011; Gustafsson et al., 2017; Savchuk and Wulff, 2009, 2001). However, the sediment pools and the role of sediment processes in benthic–pelagic exchange are not as well quantified as pelagic pools and fluxes. **The higher uncertainty in benthic compared to pelagic processes as well as the traditional focus on pelagic eutrophication are probable reason why physical-biogeochemical models of the Baltic Sea have omitted benthic fauna as state variables (e.g. Eilola et al., 2011; Lessin et al., 2018).** Here, we aim to fill this knowledge gap and explore the role of benthic fauna in biogeochemical cycling of C, N and P on a long–term ecosystem–level scale.”

Materials and methods

1. Line 74: the focus of the study is “on the Baltic Proper and the Gulf of Riga”. Could you please explain your choice of focus areas in more detail? While the focus on Baltic Proper as the largest and deepest basin is straightforward, why Gulf of Riga? It is the only sub-basin to the east of the Danish Straits for which there was no validation data

available apart from some literature-based values, so it is not possible to adequately validate the model for this region, and consecutively the confidence in model performance here is lower than in other areas, both in terms of absolute biomasses of benthic fauna and its impact on pelagic biogeochemistry.

We realize we did not properly motivate the choice of focus, which we hope is now clarified with the following addition:

"In this study, we focus on comparing results from the Baltic Proper and the Gulf of Riga (Fig. 1), two basins with a similar benthic community composition but differing in physical and biogeochemical properties such as depth, openness, productivity and bottom oxygen conditions. We expect these differences to be reflected in the strength of benthic-pelagic coupling processes and the role of benthic fauna therein."

Further, we found new data to validate benthic fauna biomasses in the Gulf of Riga (93 data points in the appendix of Gogina et al. (2016)). This new validation analysis conforms with the previous comparison to literature values: the simulated mean total biomass is higher that observed (mostly due to an overestimation of *M. balthica*), but the simulated biomasses of all groups are well within the standard deviations of data. We also added several new estimates of benthic fauna biomass in the gulf found in the literature to Table 1. The new estimates from the 1980s are a fairly close match to simulations. Please see our response to RC2, Line 221 for details.

2. Figure 1: As the text often refers to different regions of the Baltic Sea, a legend for all of them should be provided in addition to numbers. Moreover, model results are presented at intervals of 0-30, 30-70, 70-120 (and 120+) meters, so bathymetry on the figure should use the same gradation.

We thank the reviewer for the suggestions, which will significantly aid the reader. We have changed the figure accordingly.

3. Section 2.2: BALTSEM simulates Baltic Sea as 13 horizontally homogeneous boxes. In the Results and/or Discussion section, it should be elaborated on how this affects model results. Some of the Baltic Sea sub-basins have strong gradients in nutrient distributions, which in turn leads to gradients in productivity and in distributions of benthic fauna. For instance, the eastern part of the Gulf of Finland is heavily influenced by riverine nutrient inputs and have higher primary production rates than its western part. How are these gradients accounted for?

The horizontally averaged approach does mean that finer scale gradients in e.g. nutrient distributions cannot be accounted for. This is why we have excluded data from archipelago areas, as the model does not represent the small-scale complexity in these areas (as explained in the Methods section of Appendix C).

Although horizontal integration indeed is a compromise, comparisons with horizontally resolved models showed that BALTSEM is as good as these in reproducing seasonal and long-term variations, at least outside hot-spot areas like the eastern Gulf of Finland and major river mouth areas (Eilola et al., 2011; Meier et al., 2018).

4. Lines 128-130: why does the model consider respiration and excretion of benthic fauna to contribute *ammonia* and phosphate directly to the water column? At least in case of deposit feeders one might expect part of the excreted ammonia to be oxidised

directly within sediments and to be released into the water column in form of nitrate, and some of the excreted phosphate to be bound within sediment, the same way as it is considered for microbially remineralised nutrients in the model?

It is true that part of animal excretion takes place within the sediment and is oxidized before release to the pelagic. However, we do not have good estimates on what that proportion could be. Given that there are no head-down/subsurface deposit-feeders in the Baltic Sea *sensu stricto*, and the main species are either mobile surface dwellers (e.g. *Saduria entomon, Monoporeia affinis*) or infauna living with head or siphons at the surface (e.g. *Macoma balthica, Marenzelleria* spp.), we can assume that the majority of excretion is released to the water column. Following the principle of Occam's razor, we try to keep the model as simple and interpretable/traceable as possible, and therefore refrain from dividing the excretion fluxes. In oxic conditions (which almost always prevail in areas with benthos), the excreted ammonia is immediately oxidized in the water column.

5. Does the model consider sediment resuspension? It seems to be an important factor mediating organic matter availability in the sediments in coastal seas, especially shallow regions such as Gulf of Riga? Omitting it might have important consequences both for model parameterisation and its results?

BALTSEM includes downward relocation of sediments due to resuspension and lateral transport. A description of the process can be found in Savchuk et al. (2012) and in the Supplementary material of Ehrnsten et al. (2020a).

6. Line 134: "degradation" - what type of degradation? At this point, it should be specified in more concrete terms.

We changed the term "degradation" to the more specific term "mineralization".

7. Around line 152: how are silicate transformations handled in the benthos?

We assume that Si does not interact with the fauna (i.e. storage or transformation of Si by fauna is not important for model dynamics), and therefore kept the original formulations for sediment Si from BALTSEM, i.e. it is treated as single pool with sinking diatom Si as a source and mineralization and burial as sinks (see Appendix A, Eq. A4). We added a short mention of this in the main manuscript:

"Sediment C, N and P pools are further divided into three banks of different age to resolve the food limitation of benthic fauna (Fig. 2), while benthic Si is represented as a single pool that does not interact with the fauna."

8. Line 167: is nitrification considered a sink for oxygen in the model?

Yes, nitrification consumes oxygen according to Eq. A17 in Appendix A. We added a mention of this to the sentence.

9. Equation 8, page 7: is sequestered phosphorus considered a state variable in the model, as it is not listed among the state variables in Table A1?

Sequestered P is not separated as a state variable, but is instead returned to the oldest sediment bank SED3P (see Figure 2). We will not endeavour into speculation about how a separate state variable would affect results of the current study, but we are looking forward to see the results of an ongoing study implementing BALTSEM with sediment iron and iron-bound P as explicit state variables (unfortunately without benthic fauna, though).

10. Line 176: what is a definition of “severely hypoxic”?

The oxygen concentration value when P sequestration switches to release depends on salinity and bioturbation, and therefore it is difficult to give an exact value. At salinity 0 and Ebio=0.6, the value is 0 mg O_2 L^{-1}, while at salinity >5 and Ebio=0 it is 1.44 mg O_2 L^{-1}. Please see our response to RC3 for graphs of the oxygen-dependency of benthic N and P processing.

11. Section 2.3: please provide more detail on model forcing and setup in addition to the references, so the interested reader does not need to look for those details within several previous papers. How was benthic fauna initialised?

We added a short description of model forcing and initial conditions. For the interested reader, we refer to the given references for further details.

“The model was run over 1970–2020 forced with observed nutrient loads and actual weather conditions as described in Gustafsson et al. (2012, 2017) with forcing time–series extended to 2020. The physical circulation was forced by 3-hourly meteorological conditions and monthly time-series of river runoff and state variable concentrations and sea level at the North Sea boundary. Monthly inputs of N, P, C and Si from land via rivers and from coastal point sources as well as atmospheric deposition of N, P and C were used as biogeochemical forcing. Initial conditions in 1970 were based on observations for pelagic variables and hindcast simulations for benthic variables as described in Gustafsson et al. (2012) and Ehrnsten et al. (2020a). Shortly, the benthic fauna and their food banks (SED1X and SED2X) were set to 1000 mg C m^{-2}, 100 mg N m^{-2} or 10 mg P m^{-2} throughout the model domain in 1960 and given 10 years of hindcast simulation to spin up, allowing the variables to turn over several times. Initial conditions for SED3X, with a slower turnover rate, were based on a hindcast from 1850 to properly account for the build-up of sediment nutrient pools during past eutrophication (Gustafsson et al., 2012).”

12. Line 184-185. Aggregating the results to means and standard deviations obscures a lot of detail about model capabilities, which cannot be justified by stating that “the purpose of this study was to evaluate large-scale dynamics”. Does the model capture seasonal dynamics of benthic fauna? Are there long-term trends in the model *or* in the data? There is a general lack of knowledge about benthic processes, so more detailed and varied comparison with data would allow to identify critical gaps and steer discussions which will help to identify directions for improvement. This is especially important for a new model implementation, as presented in the manuscript.

The main aim of this study is to examine the effects of benthic fauna on biogeochemical cycling of C, N and P on a system scale. For this purpose, we believe that aggregation of results into long-term (20 year) means and standard deviations is adequate. Before we can get to the main aim, we need to validate the model, and we believe that this is the part the reviewer is primarily referring to in this comment.

As much as we would like to add validation of further benthic variables and processes, unfortunately there no data available for further validation. While there probably is a large amount of data on e.g. sediment concentrations and fluxes measured for research and monitoring purposes, to our knowledge there is no comprehensive collection of the data available in the form of an open-access database or review article. As we comment also in response to RC2 and RC3, we restrict the validation to these types of transparent, open-access sources and believe it would be far beyond the scope of this study to do a comprehensive collection of data from unpublished sources or literature. Thus, we restrict the quantitative validation to pelagic variables (Appendix B: salinity, temperature, oxygen, ammonium, nitrate and phosphate), benthic fauna (Appendix C) and sediment C:N:P ratios (results section 3.1). In addition, we put our results into context of previous research, including quantitative and qualitative evaluations of sediment processes and the role of fauna therein, in the Discussion.

The dynamics of the two component models has been extensively validated before, including seasonal and long-term dynamics in BALTSEM (Eilola et al., 2011; Gustafsson et al., 2012, 2014; Meier et al., 2018; Savchuk et al., 2012), as well as long-term dynamics of the Benthic Fauna Model (Ehrnsten et al., 2019, 2020a). We are not aware of any data for validation of seasonal dynamics of benthic fauna, but we also believe that the relevant time-scale of benthic processes in a system perspective is years rather than seasons. Thus, we would not like to add further comparisons of temporal dynamics to this manuscript, since (1) the manuscript is already 52 pages long, (2) many aspects have been validated previously, and (3) to our knowledge, there is no data available for further validation of benthic processes.

13. Lines 187-188. This is an important topic for benthic modelling and should be discussed in some more detail. What should observational scientists measure to help constrain the model?

What would be needed first and foremost is a comprehensive collection of observational data from the benthic realm in a format that can be used for validation of basin-scale stocks and fluxes, similar to existing databases for pelagic variables. We have added a section discussing this to the conclusions and outlook, reproduced below on pages 11-13.

Results

1. The model run is 1970-2020. Why was validation of pelagic state variables performed for 1970-2015?

2015 is the latest year we have a full compilation of validation data for, therefore the validation only extends to 2015.

2. Figure 8 shows that there was a noticeable impact of inclusion of benthic fauna/bioturbation on primary productivity and nitrogen fixation, so even if overall relative bias remained almost unchanged, there was an impact on pelagic dynamics. Please provide more detail on how the models with BMM and without BMM compare. Given the impacts of fauna and its activity on nutrient fluxes is the main scope of the paper, a more detailed assessment of changes in pelagic environment is justified.

As the model with BMM has been recalibrated compared to model without BMM, it would not be correct to compare the two model versions directly to infer the effects of fauna.

We realize that this was not properly explained in the manuscript, and have now included a section on the recalibration process in the Materials and Methods section. Shortly, BALTSEM without BMM uses a simple first-order formulation for mineralization of sediment stocks. This implicitly includes mineralization by microbes and benthic fauna. When the benthic fauna was added as explicit variables that perform mineralization through metabolic processes, the first-order mineralization rates (a_{ZSEDC}, a_{ZSEDN} *and* a_{ZSEDC} in Table A2) were decreased to avoid "double-counting". Instead of comparing model versions, we use the sensitivity analyses, scenarios and comparison of the two contrasting basins to infer the effects of fauna under various settings.

The recalibration also explains why the overall relative bias is unchanged: we aimed for producing a model with added dynamics that works as least as well as the previous version of BALTSEM, which has been developed over several decades to reproduce the main biogeochemical cycling processes in the Baltic Sea. Thus, we did not aim primarily at improving model performance in this sense, but rather to create a tool for exploring the effects of benthic fauna on biogeochemical dynamics. In fact, it took us about two years to get to the stage where the coupled model performs about equally well as the uncoupled model.

Regarding the second part of the comment, we agree that more detail on the effects of fauna on the pelagic processes is warranted, and have added text and figures on changes in primary production, nitrogen fixation, sedimentation and oxygen conditions to the results and discussion as suggested here and in several other comments.

3. Line 214: how is "reasonable accuracy" defined in case of benthic variables? As the authors demonstrate, benthic fauna biomass shows high variability, hence large standard deviations, especially when aggregated over long time and large areas, while standard deviations in the modelled fauna are relatively small. Does this high std in data really justify using cost function (Appendix C) as a validation metric?

"Reasonable accuracy" is defined according to Eilola et al. (2011): model results can be interpreted as good if the model mean is within one standard deviation of the observed mean ($0 \leq CF < 1$), reasonable if $1 \leq CF < 2$ and poor if $CF \geq 2$. This was explained in Appendix C, but is now also defined in the main text.

High variability in benthic fauna data is inherent to variability that is not expected to be found the model results (e.g. due to patchiness of habitats and resources), so therefore we think it is reasonable to investigate whether the model mean is within the variation of the data. We are open with the fact that large standard deviations in data give some "slack" to the Cost Function estimates (discussed both in the main manuscript and Appendix C), but we still think that Figure C1 gives a good overview of the model performance with respect to benthic fauna. All data behind the cost function is shown in Figure C2 to allow the reader to judge the performance of the model in greater detail.

4. Line 221-224: see my comment above: why focus on the Gulf of Riga?

Please see our response to comment 1. (Materials & methods) above.

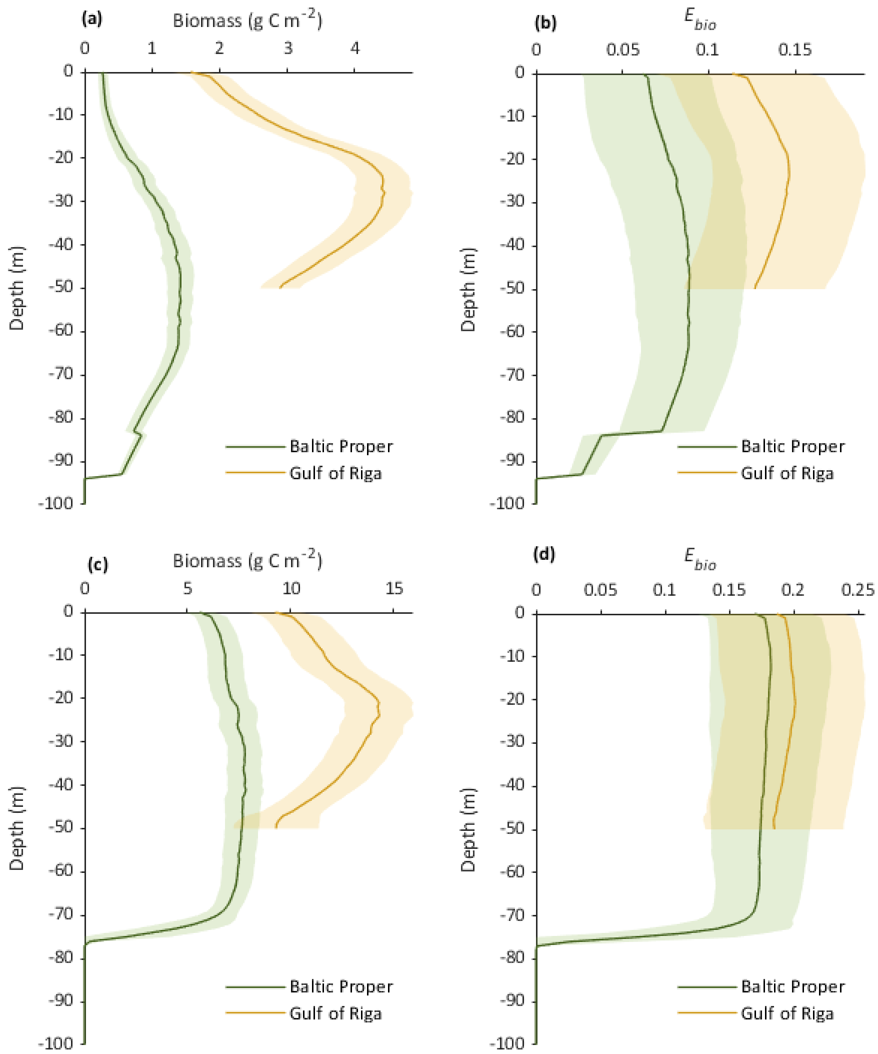
5. Section 3.3, from line 247: this is a key section in the manuscript, as it demonstrates the direct and indirect effects of fauna. Yet it is very short and just skims through the results and through the figures. It should be more detailed, which would also make it

easier to relate figures 7-9 to the text. It would also benefit from detailed quantitative information, in particular regarding relative changes.

We have added more detail to the existing text (especially % change), and also added new paragraphs delving more into the pelagic dynamics. As suggested further down, we added result graphs on the extent of hypoxic and anoxic areas under different levels of bioturbation to Figure 8.

6. Line 264: for this, a figure similar to figure 5 could be provided.

We could add a figure on the depth distribution of fauna and bioturbation coefficients, as suggested. However, we already have 11 figures with multiple panels in the main manuscript plus several more in the appendices. We are not sure that this figure would add enough new information compared to Fig. 10 that gives the same information aggregated over depth. We leave it up the reviewer's and/or editor's discretion to judge if a figure similar to one below should be included. In that case, we can rerun the model with higher depth resolution output to get smoother curves.



Possible new figure. Depth distribution of benthic fauna biomass and the bioturbation coefficient E_{bio} in the upper 100 m of the Baltic Proper and the Gulf of Riga. Averages (lines) and standard deviations (shaded areas) of biweekly values 2080-2100 in the BSAP (a-b) and HIGH nutrient load scenarios (c-d).

7. Lines 265-271: figure 11 (as well as several other figures of the manuscript) contain multiple subfigures, so it would be very helpful to provide relevant pointers in the text, e.g., reference to Fig 11a on line 268, Fig 11b on line 269 etc, so the text and figures could complement each other.

We have added references to subfigures in results and discussion as suggested.

Discussion

1. Line 284: both terms “long-term” and “large-scale” have been used multiple times in the manuscript without proper definition. It creates overall impression of vagueness. Does “long-term” stand for “multidecadal” or “long-term mean”? The words “large-scale” could be omitted altogether without impact on meaning.

We have used the terminology of BALTSEM, which stands for “Baltic Sea Long-Term Large-Scale Eutrophication Model”. In this context, the former refers to the decadal time-scale of simulations (i.e. long-term mean), and the latter to (a) the organisational scale of the ecosystem and (b) the spatial scale of basins. We agree that the terms were used in a slightly vague sense in many places and have removed or replaced the terms with more specific definitions where applicable. We also removed the term “Large-scale” from the title (see first comment above).

2. Line 285: as it currently stands, it should be “Baltic Sea”, not “coastal sea”, as the manuscript shows the model is not yet applicable in high-salinity regions.

This is a good and valid point, we changed the expression as suggested.

3. Line 288-289: to support the statement that benthic fauna can alleviate the ‘vicious circle’ of eutrophication, could you show the differences in extent of hypoxic area (e.g. with Figure 8) for simulations with different levels of bioturbation?

We thank the reviewer for the good suggestion. We added graphs showing the extent of hypoxic and anoxic areas under different levels of bioturbation to Figure 8, as well as a section to results giving more detail regarding changes in oxygen conditions, N fixation and primary production.

4. Similarly, what about primary production and N-fixation for the two future scenarios (Fig 10)? I think these would also support the discussion on faunal impacts on ‘vicious circle’?

We added graphs on primary production and N-fixation in the load scenarios to Figure 10 as suggested. These graphs nicely illustrate the interesting result that the relationship between primary production and sedimentation changes with nutrient loads, as discussed in section 4.3.

5. From line 307: as the overall model performance was not improved by adding benthic fauna, and related processes are not well understood, could the authors discuss on what understanding is currently lacking and which model improvements are desired.

Please see our response to comment 2. in results above.

6. Lines 313-316: same as the previous comment, which data is currently lacking? Which mechanisms are not sufficiently understood? The sentence on “the main strength of this study” is too general (“many interlinkages”) – could it be replaced with something more concrete and relevant to the actual modelling work?

We realise that this paragraph was badly formulated, as also pointed out by Reviewer 3. It is now removed from the manuscript. Instead, we have expanded section “5. Conclusion and outlook” with a discussion on challenges, opportunities and ways forward. Please find the new text on page 12 below.

7. Lines 326-334: please keep referencing to figures and sub-figures as discussion requires.

We have added references to figures and sub-figures as suggested.

8. Line 361: is it worth presenting fast running time as an advantage, since the manuscript presents results from only several simulations with sensitivity analysis limited to a single parameter?

While the manuscript presents results from just a few simulations, the work behind the manuscript comprises hundreds if not thousands of simulations for different purposes including integration of code from the two models, debugging, parameterization, calibration etc. We doubt that we would have been able to implemented the model integration and analyses as successfully if we would have used a complex 3D-model with a running time counted in days rather than minutes. We have added a paragraph to Materials and methods explaining the calibration process.

Conclusions and outlook

1. Line 397: “much-studied bioturbation” - I suggest replacing this with “bioturbation, relatively more studied in the modelling context”.

Changed as suggested.

2. Could the authors provide some outlook on the potential future directions of benthic modelling, in the context of Baltic Sea in particular?

3. Lines 399-402: These conclusions are far too general. Could it be something emerging from the study rather than vague statement that benthic-pelagic coupling is “modified by multiple drivers, which may change over time”?

4. Based on your work, has the time arrived for the regional models (of the Baltic Sea) to extend their formulations to explicitly include benthic fauna?

Response to comments 2.-4.

We rewrote the section to (a) be more specific about the findings of this study, and (b) provide an outlook for future directions of benthic modelling in the context of physical-biogeochemical models of the Baltic Sea and beyond. The new version is reproduced below:

“Using a newly developed modelling tool, significant effects of benthic macrofauna on C, N and P cycling were simulated in the semi-enclosed brackish-water Baltic Sea, with impacts on the ecosystem from the extent of hypoxic bottoms to the rates of pelagic nitrogen fixation and primary production. Our results suggest that in addition to bioturbation, relatively more studied in the modelling context, the metabolism of benthic fauna should be given more attention in future studies as it may play a significant role in benthic mineralization of organic C, N and P in coastal seas and estuaries.

The magnitude of effects of benthic fauna on biogeochemistry generally decreased with depth and increased with productivity, as shown by the comparison of two basins and different nutrient load scenarios. Thus, these simulations confirm the notion that benthic–pelagic coupling is strongest in shallow coastal areas (Griffiths et al., 2017; Nixon, 1981), but they also show that this relationship is modified by multiple interacting physical and biological drivers, which may change over time. For example, we found that the proportion of primary production reaching the seafloor decreased with increased nutrient loading and increased temperature, as both led an intensification of pelagic nutrient cycling. Further, simultaneous positive and negative feedback loops led to complex relationships between, e.g. productivity and P cycling (as seen in Fig. 11f). On the one hand, increased productivity can increase the amount of bioturbating fauna, stimulating P sequestration. On the other hand, increased productivity can increase benthic oxygen consumption for mineralization of sinking organic matter, leading to deteriorating oxygen conditions and increasing P leakage from sediments. Unravelling the many interacting drivers and responses on a system scale is important to understand how coastal and global biogeochemical cycles are responding to changes in, e.g. nutrient loads and climate.

Even though these large-scale simulations contain a large degree of uncertainty, they are an important complement to empirical studies, which for practical reasons can only consider temporally and spatially limited parts of the system (Boyd et al., 2018; Snelgrove et al., 2014). To improve the confidence in simulation results, we see two major ways forward. First, as all models contain different formulations, assumptions and uncertainties, implementing benthic fauna components in other physical-biogeochemical models and comparing the results would greatly increase the strength of evidence for those results where different models agree. This kind of ensemble modelling is increasingly used in climate change research, and has also been applied in the context of Baltic Sea biogeochemistry (Meier et al., 2012, 2018; Murphy et al., 2004). We hope that the publication of the benthic model formulations stimulates the development of benthic fauna modules in other models of the Baltic Sea ecosystem and beyond. Even though the current model implementation is only applicable to the brackish parts of the Baltic Sea due to a lack of functional groups present in the marine parts, the inclusion of additional functional groups using the existing groups as a template would be straightforward technically. The main challenges are the parameterisation of group-specific rates as well as managing the increased complexity.

Second, a comprehensive compilation of observational data on sediment stocks and fluxes would be needed for improved model validation. Such data is collected for monitoring and research purposes by a great number of institutions around the Baltic Sea, but a comprehensive, open-access, quality-controlled collection of this data is lacking. The Baltic Environment Database (BED) has been invaluable for both model development and validation of pelagic physics and chemistry. While this data can be used as indirect validation of benthic model processes in the strongly coupled system, we call for the development of a “Benthic BED” to facilitate future model development. A comprehensive collection of observational data would also facilitate the identification of knowledge gaps and future research priorities."

Appendix A, line 817: why is mortality rate chosen to be linear for *Limecola balthica*, and quadratic for the other two groups?

Quadratic mortality rate is generally used as a closure term to represent predation by groups not present in the model. This is the reasoning behind using a quadratic term for the benthic predator group and linear terms for its two prey groups, the deposit-feeders and *M. balthica* in the first version of the fauna model (Ehrnsten et al., 2019). As the model was extended to the Baltic Sea scale, the mortality rate of deposit-feeders was changed to quadratic to compensate for missing predation pressure in some areas with low or absent biomass of predators (primarily in the oligotrophic Bothnian Bay, Ehrnsten et al., 2020a).

Technical corrections

1. Line 89: "extension" - should be "extent" Corrected as suggested.

2. Line 110: "Ponotporeia" - should be "Pontoporeia" Corrected as suggested.

3. In equation 1, should "Uci" be "UBFiC" in the denominator? Corrected as suggested.

4. Page 6: equation 6 is missing. The mistake in equation numbering was corrected. Equations are now numbered consecutively from 1 to 9.

5. Line 175 and equation 10: the fitting constants 5-8 are not featuring in the equation. The extra fitting constants were removed.

6. Figures 7 and 11: in both cases, "phosphorus (c, d)" should be "phosphorus (e, f)". Corrected as suggested.

7. Line 319: "suggests" - should be "suggest" Corrected as suggested.

8. Table 1: what is "total" biomass? Should it be "mean" instead? Corrected as suggested.

9. Line 712: "and" - should probably be "an" Corrected as suggested.

10. Line 742: Nitrite and nitrate should be "NO2" and "NO3", respectively. Corrected as suggested.

11. Line 784: "and" at the end of the line should be "are"? Corrected as suggested.

12. Line 794: "feces" - "faeces". Corrected as suggested.

13. Line 907: "120-120" should be "70-120" Corrected as suggested.

14. Figure C2: please use a different y-scale for the Bothnian Bay. Scale changed and a note about differing scale added to figure legend.

References in review response

Ehrnsten, E., Norkko, A., Timmermann, K. and Gustafsson, B. G.: Benthic-pelagic coupling in coastal seas – Modelling macrofaunal biomass and carbon processing in response to organic matter supply, J. Mar. Syst., 196, 36–47, doi:10.1016/j.jmarsys.2019.04.003, 2019.

Ehrnsten, E., Norkko, A., Müller-Karulis, B., Gustafsson, E. and Gustafsson, B. G.: The meagre future of benthic fauna in a coastal sea – benthic responses to recovery from eutrophication in a changing climate, Glob. Chang. Biol., 26, 2235–2250, doi:10.1111/gcb.15014, 2020a.

Ehrnsten, E., Sun, X., Humborg, C., Norkko, A., Savchuk, O. P., Slomp, C. P., Timmermann, K. and Gustafsson, B. G.: Understanding Environmental Changes in Temperate Coastal Seas: Linking Models of Benthic Fauna to Carbon and Nutrient Fluxes, Front. Mar. Sci., 7, 450, doi:10.3389/fmars.2020.00450, 2020b.

Eilola, K., Gustafsson, B. G., Kuznetsov, I., Meier, H. E. M., Neumann, T. and Savchuk, O. P.: Evaluation of biogeochemical cycles in an ensemble of three state-of-the-art numerical models of the Baltic Sea, J. Mar. Syst., 88(2), 267–284, doi:10.1016/j.jmarsys.2011.05.004, 2011.

Gogina, M., Nygård, H., Blomqvist, M., Daunys, D., Josefson, A. B., Kotta, J., Maximov, A., Warzocha, J., Yermakov, V., Gräwe, U. and Zettler, M. L.: The Baltic Sea scale inventory of benthic faunal communities, ICES J. Mar. Sci., 73(4), 1196–1213, doi:10.1093/icesjms/fsv265, 2016.

Gustafsson, B. G., Schenk, F., Blenckner, T., Eilola, K., Meier, H. E. M., Müller-Karulis, B., Neumann, T., Ruoho-Airola, T., Savchuk, O. P. and Zorita, E.: Reconstructing the development of Baltic Sea eutrophication 1850-2006, Ambio, 41(6), 534–548, doi:10.1007/s13280-012-0318-x, 2012.

Gustafsson, E., Deutsch, B., Gustafsson, B. G., Humborg, C. and Mörth, C. M.: Carbon cycling in the Baltic Sea - The fate of allochthonous organic carbon and its impact on air-sea CO2 exchange, J. Mar. Syst., 129, 289–302, doi:10.1016/j.jmarsys.2013.07.005, 2014.

Gustafsson, E., Savchuk, O. P., Gustafsson, B. G. and Müller-Karulis, B.: Key processes in the coupled carbon, nitrogen, and phosphorus cycling of the Baltic Sea, Biogeochemistry, 134(3), 301–317, doi:10.1007/s10533-017-0361-6, 2017.

Lessin, G., Artioli, Y., Almroth-Rosell, E., Blackford, J. C., Dale, A. W., Glud, R. N., Middelburg, J. J., Pastres, R., Queirós, A. M., Rabouille, C., Regnier, P., Soetaert, K., Solidoro, C., Stephens, N. and Yakushev, E.: Modelling marine sediment biogeochemistry: Current knowledge gaps, challenges, and some methodological advice for advancement, Front. Mar. Sci., 5, 1–8, doi:10.3389/fmars.2018.00019, 2018.

Meier, H. E. M., Edman, M. K., Eilola, K. J., Placke, M., Neumann, T., Andersson, H. C., Brunnabend, S.-E., Dieterich, C., Frauen, C., Friedland, R., Gröger, M., Gustafsson, B. G., Gustafsson, E., Isaev, A., Kniebusch, M., Kuznetsov, I., Müller-Karulis, B., Omstedt, A., Ryabchenko, V., Saraiva, S. and Savchuk, O. P.: Assessment of Eutrophication Abatement Scenarios for the Baltic Sea by Multi-Model Ensemble Simulations, Front. Mar. Sci., 5, 440, doi:10.3389/fmars.2018.00440, 2018.

Middelburg, J. J.: Reviews and syntheses: to the bottom of carbon processing at the seafloor, Biogeosciences, 15(2), 413–427, doi:10.5194/bg-15-413-2018, 2018.

Rhoads, D. C.: Organism-sediment relations on the muddy sea floor, Oceanogr. Mar. Biol. An Annu. Rev., 12, 263–300, 1974.

Savchuk, O. and Wulff, F.: A Model of the Biogeochemical Cycles of Nitrogen and Phosphorus in the Baltic, in A Systems Analysis of the Baltic Sea. Ecological Studies, vol 148, edited by F. W. Wulff, L. A. Rahm, and P. Larsson, pp. 373–415, Springer, Berlin Heidelberg., 2001.

Savchuk, O. P. and Wulff, F.: Long-term modeling of large-scale nutrient cycles in the entire Baltic Sea, Hydrobiologia, 629(1), 209–224, doi:10.1007/s10750-009-9775-z, 2009.

Savchuk, O. P., Gustafsson, B. G. and Müller-Karulis, B.: BALTSEM - a marine model for desicion support within the Baltic Sea Region. BNI Technical report No 7., 2012.

Seidl, R.: To model or not to model, that is no longer the question for ecologists, Ecosystems, 20(2), 222–228, doi:10.1007/s10021-016-0068-x, 2017.

Snelgrove, P. V. R., Thrush, S. F., Wall, D. H. and Norkko, A.: Real world biodiversity-ecosystem functioning: A seafloor perspective, Trends Ecol. Evol., 29(7), 398–405, doi:10.1016/j.tree.2014.05.002, 2014.

Snelgrove, P. V. R., Soetaert, K., Solan, M., Thrush, S., Wei, C.-L., Danovaro, R., Fulweiler, R. W., Kitazato, H., Ingole, B., Norkko, A., Parkes, R. J. and Volkenborn, N.: Global carbon cycling on a heterogeneous seafloor, Trends Ecol. Evol., 33(2), 96–105, doi:10.1016/j.tree.2017.11.004, 2018.

Snoeijs-Leijonmalm, P., Schubert, H. and Radziejewska, T., Eds.: Biological Oceanography of the Baltic Sea, Springer Netherlands, Dordrecht., 2017.

Stigebrandt, A. and Gustafsson, B. G.: Response of the Baltic Sea to climate change - Theory and observations, J. Sea Res., 49(4), 243–256, doi:10.1016/S1385-1101(03)00021-2, 2003.