Dear Dr. Slomp:

Thanks again for giving us the chance to revise and improve our manuscript! After several months of revision and hundreds of more numerical experiments, we are pleased to let you know we successfully addressed the two reviewer's comments in round one of the submission. Here we are providing our point-to-point responses to the two referees.

Changes in model set-ups:
1. We performed a series of sensitivity runs to evaluate the model's robustness regarding different parameterizations of the sinking velocity of the organic matter (i.e., PON and Opal). And we conclude that a sinking velocity of 5 m/day is a reasonable prescription and updated relevant model results.
2. The riverine DO concentration was changed to be a constant of 258 mmol m^{-3} assuming that the riverine DO was saturated at 25 °C under 1 atm.

Content removed:
1. The model validation for hypoxic thickness was removed since we focused more on the bottom DO dynamic.
2. We comprehensively revised the "Result and Discussion" section. Both reviewers pointed out that the 1st submission lacked originality. In this revision, we focus on 1) the contributions of different biogeochemical and hydrodynamic terms on bottom DO variability in different subregions, 2) the impacts of diatom and the complexity of lower-trophic community on the hypoxia dynamics, and 3) bottom DO's responses to the reduction of riverine nutrient loads with different reduction combination (percentage and nutrient type).

Content added:
1. The model validations of diatom ratios, sediment oxygen consumption (SOC), and ratios of SOC/ overlaying water respiration were added following the reviewers' suggestions.
2. Results and discussion on the factors to bottom DO variability was added (see section "4.1 Factors controlling subregion bottom DO variability" and "4.2 Stratification and Bottom DO Advection/Diffusion"). All terms that directly contribute to bottom DO changes were calculated and evaluated, which include horizontal advection, vertical advection, horizontal diffusion, vertical diffusion, the local rate of change (bottom DO), SOC, and DO changes due to water column biochemistry at the bottom layer. The summation of these terms contributes to the total changes in the bottom DO. A comparison of their contributions was given in different subregions of the LaTex shelf. We found that the most dominant factors to the bottom DO changes are the two advection terms, vertical diffusion term and SOC. The former three terms were associated with the changes in water column stratification. The strong linear correlations between PEA and the advection terms suggest that increased water stability in summer leads to fewer DO exchanges from advection processes. Nevertheless, the relationship between PEA and vertical diffusion of DO across the bottom layer appears to be non-linear.
3. Results from 16 sensitivity experiments for riverine nutrient reduction strategy were added (see section "4.3 Riverine nutrient reductions"). We found that the responses of summer hypoxia to the changing nutrient loads are not linear due

to the impacts of the complexity of the lower-trophic community. Nutrient reductions do not always guarantee a decrease in summer hypoxic areas; instead, due to the interactions among different plankton groups (e.g., competition on nutrients, grazing, and predation behaviors), the hypoxic zone could even increase under some nutrient reduction conditions. The most effective strategy is to simultaneously reduce the nitrogen, phosphorus, and silicon loads. A triple riverine nutrient reduction of 80% can help to fulfill the hypoxia reduction goal of 5000 km^2.

Sincerely,
Yanda Ou and Z. George Xue

Responses to Comments by Referee #1

General comments:
Overall: The manuscript describes the implementation of NEMURO in a ROMS-COAWST Gulf of Mexico model, including several new features that are targeted at studying hypoxia dynamics in the northern Gulf of Mexico. The main novelty is the inclusion of multiple phytoplankton and zooplankton functional types (from the NEMURO model), phosphorus, oxygen and a benthic layer that can accumulate PON. Using a 15 years simulation, the authors first carry out a validation of nutrient and oxygen, find that the model is able to reproduce the mid-summer hypoxic area and then analyze oxygen dynamics to show that 1) oxygen sinks in bottom waters are dominated by sediment oxygen consumption whereas the role of water column respiration is negligible, 2) hypoxia is controlled by SOC or PEA in the western and eastern part of the shelf, respectively, and 3) there is a quadratic relationship between the hypoxic volume and the hypoxic area, which can be used to predict hypoxic volume from the hypoxic area. My general assessment of the scientific content is that the manuscript lacks originality. There are some technical improvements from other models (see my technical assessment below) but the findings are mostly similar to previous studies using both observations and models, which are cited in the manuscript; the question is then what new knowledge does this study brings on the northern Gulf of Mexico hypoxia? This question should be central in the Introduction and in the Discussion.

Response: In this revision, we significantly changed the organization and presentation of the results. Some important new knowledge from our study includes 1)we quantified the physics (diffusion/advection) and biogeochemical control on hypoxia development in different subregions, 2) with the introduction of Si cycle and a complex plankton community, we found the non-linear relationship between riverine nutrient reduction and the changes of the size of the hypoxic zone. In details,

Changes in model set-ups:
1. We performed a series of sensitivity runs to evaluate the model's robustness regarding different parameterizations of the sinking velocity of the organic matter (i.e., PON and Opal). And we conclude that a sinking velocity of 5 m/day is a reasonable prescription and updated relevant model results.
2. The riverine DO concentration was changed to be a constant of 258 mmol m^{-3} assuming that the riverine DO was saturated at 25 °C under 1 atm.

Content removed:
1. The model validation for hypoxic thickness was removed since we focused more on the bottom DO dynamic.
2. We rewrote the entire "Result and Discussion" section. Both reviewers pointed out that the manuscript lacked originality. In this revision we focus on 1) the contributions of different biogeochemical and hydrodynamic terms on bottom DO variability in different subregions, 2) the impacts of diatom and the complexity of lower-trophic community on the hypoxia dynamics 3) built on 2, we further access bottom DO's responses to the reduction of riverine nutrient loads with different reduction combination (percentage and nutrient type).

Content added:
1. The model validations of diatom ratios, sediment oxygen consumption (SOC), and ratios of SOC/ overlaying water respiration were added following the reviewers' suggestions.

2. Results and discussion on the factors to bottom DO variability was added (see section "4.1 Factors controlling subregion bottom DO variability" and "4.2 Stratification and Bottom DO Advection/Diffusion"). All terms that directly contribute to bottom DO changes were calculated and evaluated, which include horizontal advection, vertical advection, horizontal diffusion, vertical diffusion, the local rate of change (bottom DO), SOC, and DO changes due to water column biochemistry at the bottom layer. The summation of these terms contributes to the total changes in the bottom DO. A comparison of their contributions was given in different subregions of the LaTex shelf. We found that the most prevailing factors to the bottom DO changes are the two advection terms, vertical diffusion term, and SOC. The former three terms were associated with the changes in water column stratification. The strong linear correlations between PEA and the advection terms suggest that increased water stability in summer leads to fewer DO exchanges from advection processes. Nevertheless, the relationship between PEA and vertical diffusion of DO across the bottom layer appears to be non-linear.

3. Sensitivity experiments for riverine nutrient reduction strategy were added (see section "4.3 Riverine nutrient reductions"). We found the responses of summer hypoxia to the changing nutrient loads are not linear due to the impacts of the complexity of the lower-trophic community. Nutrient reductions do not always guarantee a reduction in summer hypoxic area, instead, due to the interactions among different plankton groups (e.g., competition on nutrients, grazing, and predation behaviors), the hypoxic area could even increase under some nutrient reduction conditions. The most effective strategy is to reduce the nitrogen, phosphorus, and silicon loads simultaneously. A triple riverine nutrient reduction of 80% can help to fulfill the hypoxia reduction goal of 5000 km^2.

Technical assessment: The model developed and used in this study seems appropriate, although I would like to discuss a few points that might need to be revised. These points are discussed in the specific comments below.
1) the main issue is the choice of a fast-sinking rate for the particulate organic matter. This choice results in the dominance of the sediment oxygen sinks, which is also a main conclusion of the study.

Response: Concerning the possibly improper sinking velocity applied in the model, we added two sensitivity experiments with different sinking velocities: 1 m/day and 5 m/day. A comparison of SOC against the observation by McCarthy et al., (2013) (Fig. RC1-1) suggests that the experiment with a sinking velocity of 5 m/day provides the best estimates (Fig. RC1-2). The root-mean-squared errors (RMSEs) are 567 μmol m^{-2} h^{-1}, 713 μmol m^{-2} h^{-1}, and 452 μmol m^{-2} h^{-1} for sensitivity tests with a sinking velocity of 15 m/day, 1 m/day, and 5 m/day, respectively. The model results (tests with sinking velocity = 5 m/day) generally overestimate the SOC at site F5 and C6 except for January 2009 and May 2010 at site C6, but underestimate SOC at site B7 and MRM. However, the simulated and observed SOC are generally in the same order of magnitude. Times series also reveals that the magnitude of simulated SOC simulated with a sinking velocity of 5 m/day is generally within the measured range (Fig. RC1-3) over the entire year. The magnitude of simulated SOC by tests with a sinking velocity of 15 m/day is out of the upper measured bound especially in summers. Modeled SOC by the test with a sinking velocity of 1 m/day is always below the lower measured bound.

We further compared the simulated ratio of SOC/overlaying water respiration against measurements (Fig. RC1-4). The test run with sinking velocity = 5 m/day provides the best-simulated ratio with a low averaged RMSE of 4.23 over site F5, C6, and B7 compared with 4.58 (sinking velocity = 15 m/day) and 6.51 (sinking velocity = 1 m/day) derived by the other two tests. At site MRM, both the two tests with faster (5m and 15m/day) sinking velocity highly overestimate the ratio in August 2009. We ascribe such bias to the relative course bathymetry near the river mouths. Point sources are applied in the model for diverting momentum and concentration tracers from the river to the computational grid cells. The scheme can lead to an overshot of river water at the near-mouth grid cells, which, may further result in shorter residence time of organic matter and plankton.

We also compare the hypoxic area simulated by the three different experiments. The simulation with 5m and 15m/day settling velocity show a similar hypoxia zone, while the using 1m/day is generally greater than the former two (Fig. RC1-5). Both the former two estimations (5m and 15m/day) can reproduce the magnitude and interannual variability of the measured hypoxic area. Compared to the shelf-wide observations, the simulated bottom hypoxic extent (Fig. RC1-6– RC1-8) by the experiment with a sinking velocity of 5m/day produces less bias among the three experiments. Based on these results, we changed the sinking velocity of PON from 15 m day^{-1} to 5 m day^{-1} in the baseline simulations.

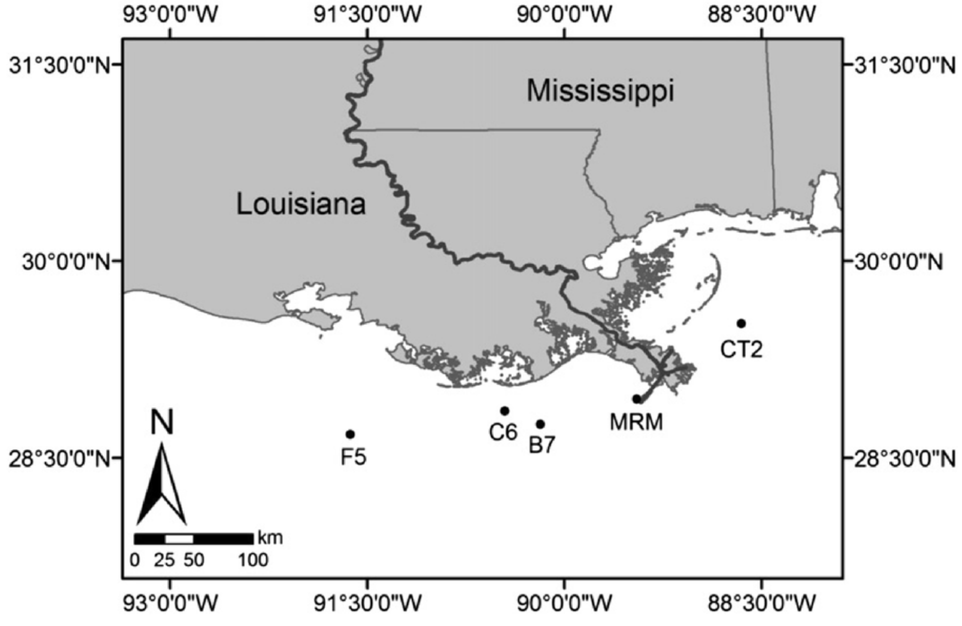

Fig. RC1-1 Map showing the location of sampling site in the northern Gulf of Mexico (McCarthy et al., 2013).

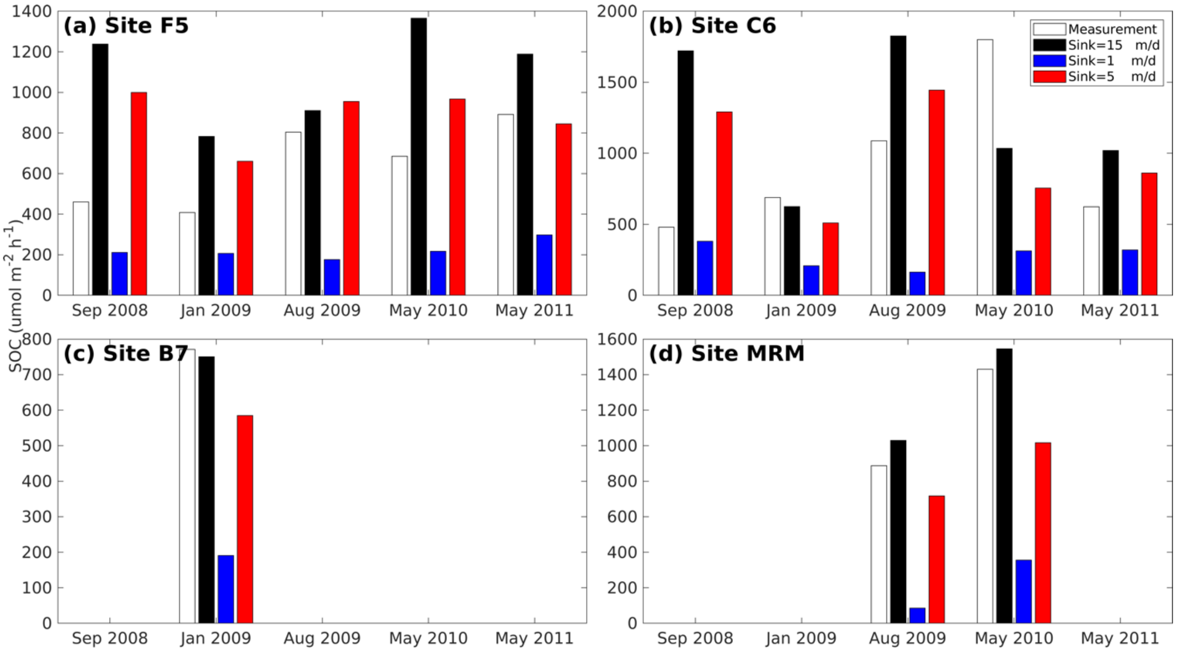

Fig. RC1-2 Comparison of observed SOC (in μmol m^{-2} h^{-1}) by McCarthy et al., (2013) and simulated SOC by different sensitivity tests.

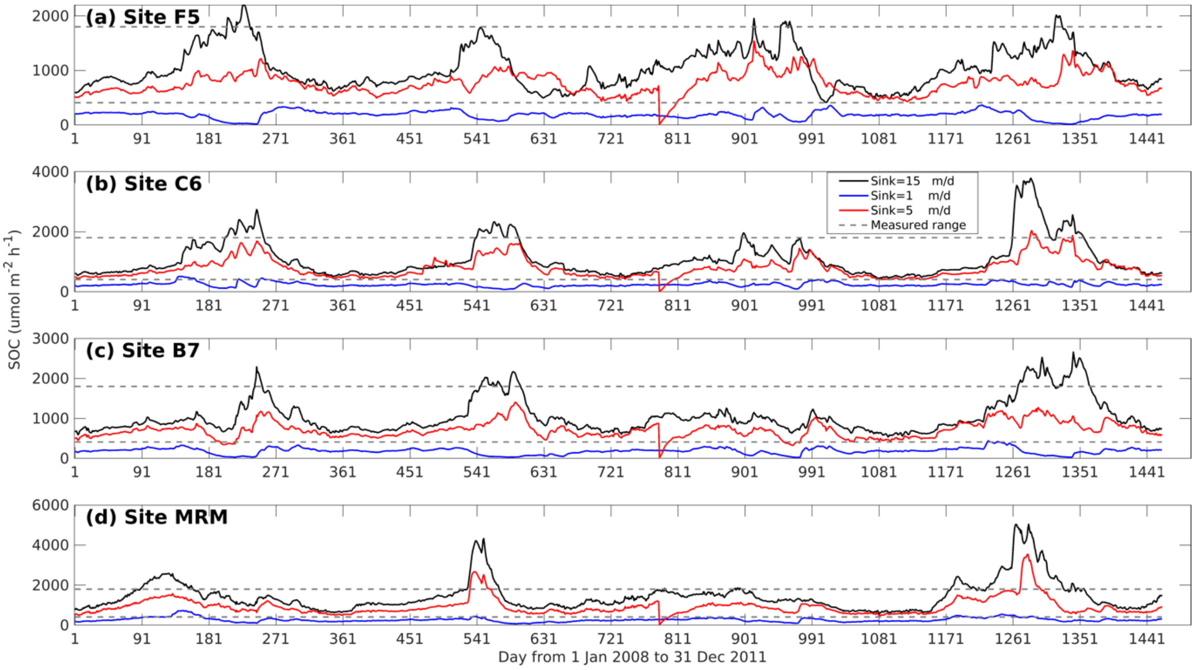

Fig. RC1-3 Daily average of simulated SOC by different sensitivity tests.

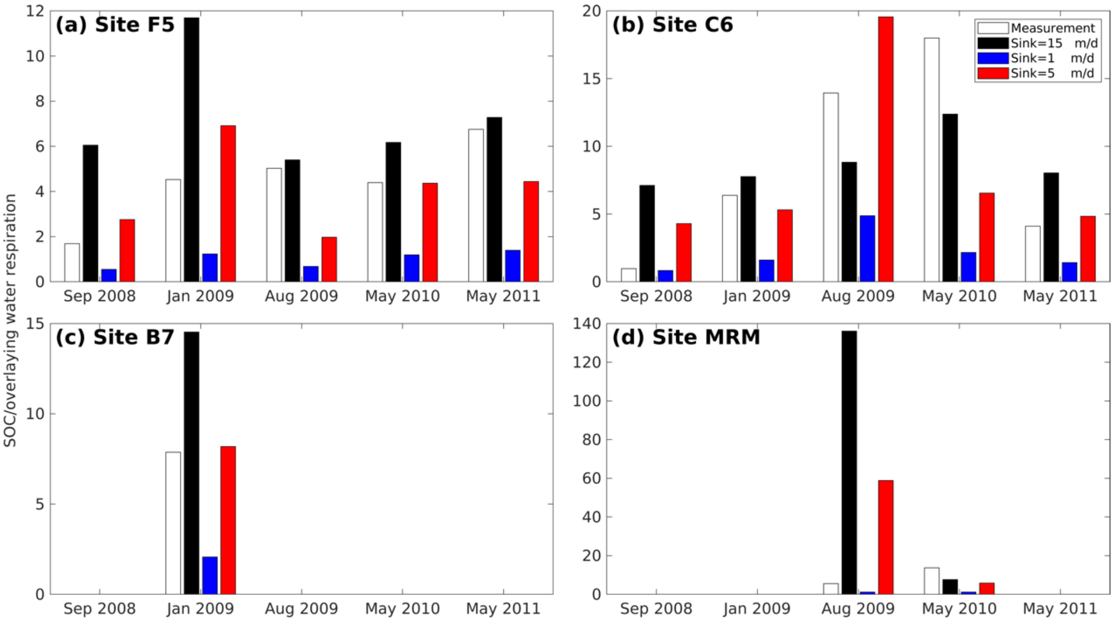

Fig. RC1-4 Comparison of observed ratio of SOC/overlaying water by McCarthy et al., (2013) and simulated ratio by different sensitivity tests.

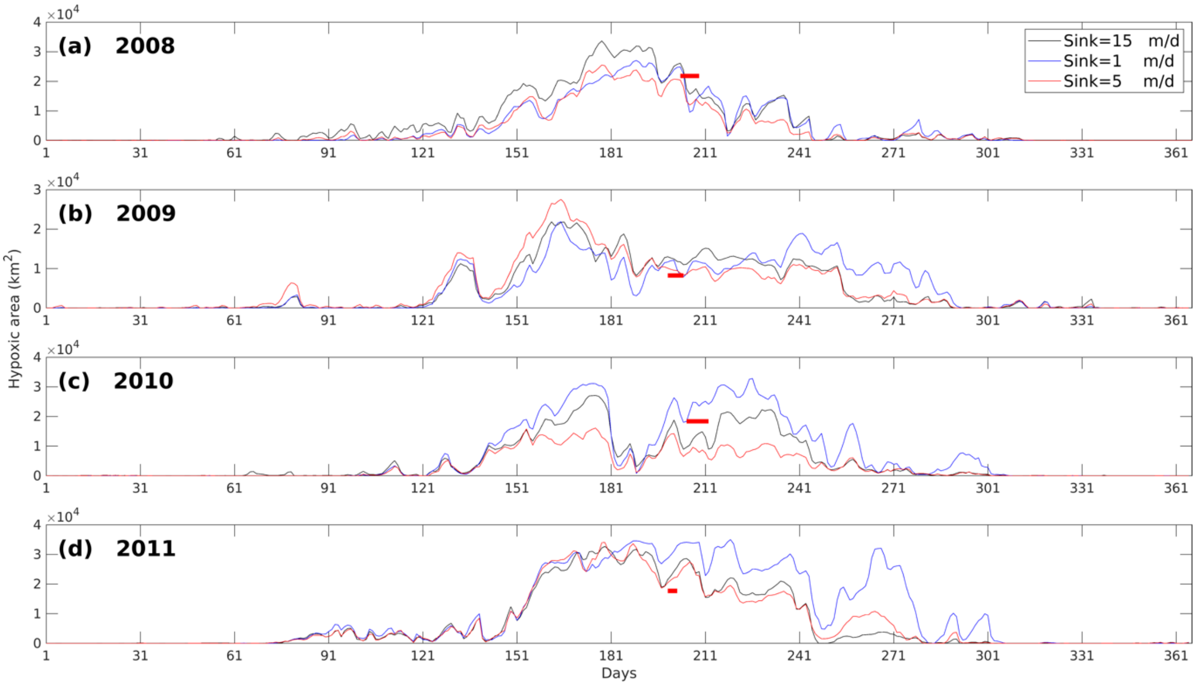

Fig. RC1-5 Comparison of observed and simulated hypoxic area. Note that the horizontal red thick bars denote the shelf-wide cruise measurements.

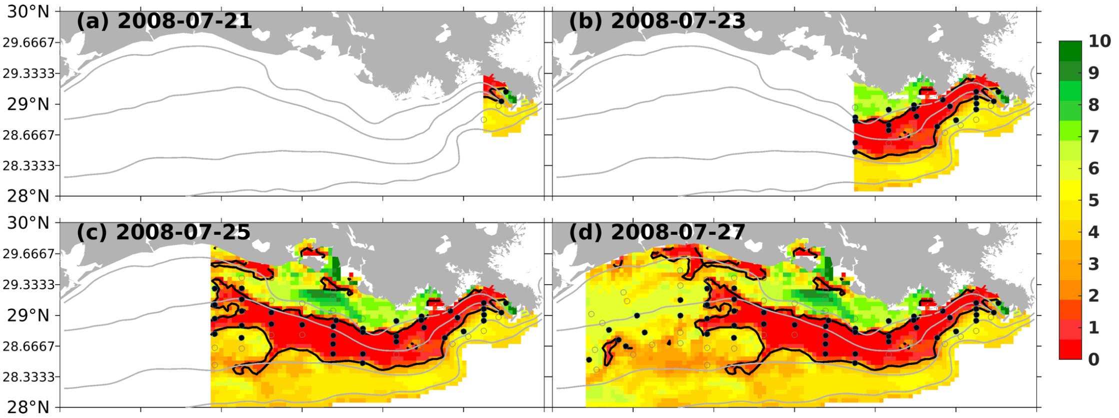

Fig. RC1-6 Evolution of simulated bottom water dissolved oxygen concentration (unit mg l^{-1}) by the sensitivity experiment with a sinking velocity of 15 m day^{-1}. The black filled circles and open circles indicate the hypoxic site and non-hypoxic site, respectively, according to the Shelf-wide cruise observations. The grey curves denote bathymetry of 5, 10, 20, and 50 m.

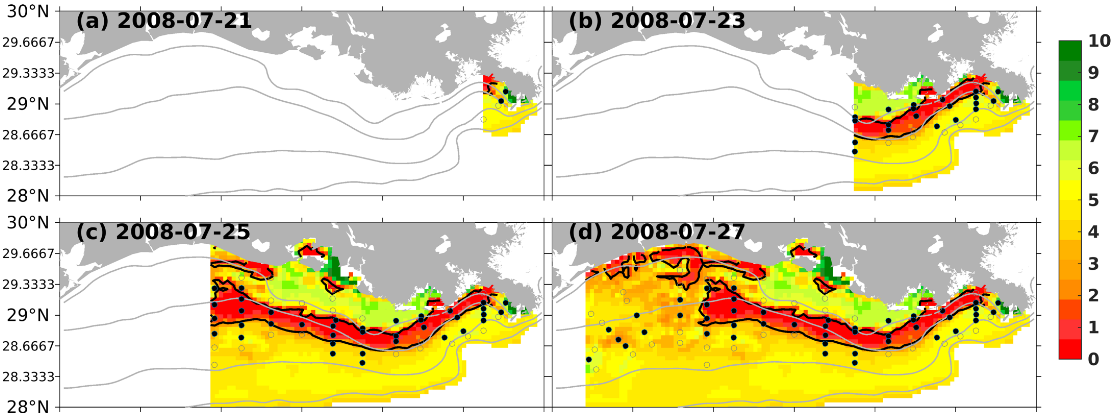

Fig. RC1-7 Same as Fig. RC1-6 but for the sensitivity experiment with a sinking velocity of 1 m day^{-1}.

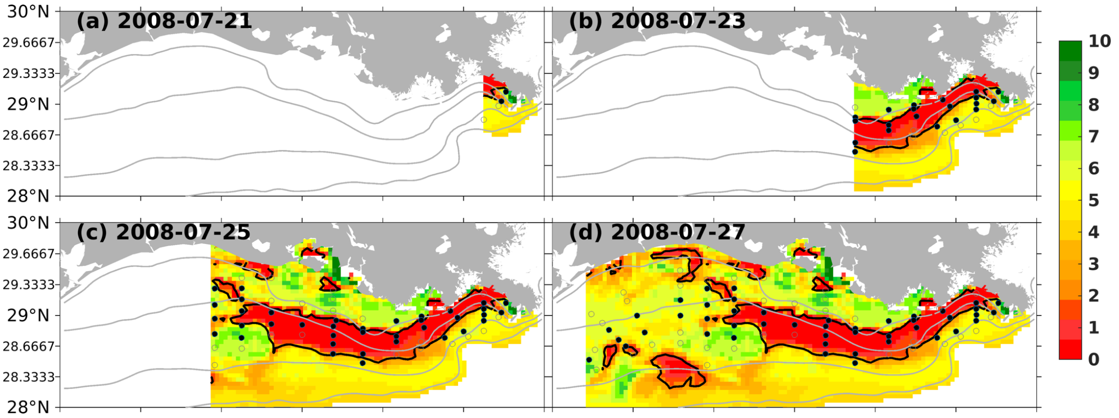

Fig. RC1-8 Same as Fig. RC1-6 but for the sensitivity experiment with a sinking velocity of 5 m day^{-1}.

2) Looking at the results, it is not clear if the model is appropriately initialized/spun up. Hypoxia occurs in deep waters and a long-term deoxygenation trend occurs both inshore near the Atchafalaya and offshore. This seems to indicate that PON accumulate in the benthic layer nearshore throughout the simulation and that there is a drift in subsurface oxygen offshore.

Response: We initialized the nitrate, phosphate, silicate, and dissolved oxygen based on the observations provided by the World Ocean Database and World Ocean Atlas. Other nutrients and plankton concentration terms were initialized spatially homogeneously as a small value. Physical terms were initialized using the HYCOM global analysis products. We did not find any "drift" in the sediment PON throughout the studied period. Sediment PON fluctuated between peaks and troughs and exhibited a salient annual cycle.

3) the model does not include a light attenuation term from river sediment (near the river mouth). This could influence the timing and distribution of primary production over the shelf, and therefore affect the conclusions of the study.

Response: The reason for the exclusion of a sediment module is to guarantee the model efficiency, which is also the practice of most of the hypoxia modeling efforts in this region. Our research group indeed published a paper regarding the impacts from sediment-induced light attenuation on shelf primary productivity (see Zang et al., 2020 Biogeosciences). Yet we found that introducing the sediment model into hypoxia simulation is computationally heavy. The model already needs 170 hours (~1 a week) with 500 CPU cores to finish a 15-year hindcast experiment. Nevertheless, based on extensive model validation (nutrient and DO profiles, hypoxia area distribution), we are confident that our current setup is capable of reproducing the general feature of hypoxia events.

In term of model validation, model results are compared with many nutrients and oxygen data. However, the format of the model-observations comparison is questionable and does not result, in my opinion, in a satisfactory validation of the model.

Response: We made significant changes in model validation in the revision. We replotted all the figures for model validation of nutrient and oxygen profiles (Figure 3 and Figure 6 in the manuscript). Probability histograms of nutrient differences between modeled and observed results were given for the upper 50 meters and upper 5 meters, respectively. Probability histograms of DO differences were given for the upper 50 meters.

For the validation of the diatom ratio, we compared the mean±1SD between simulated and observed values (Table 1 in the updated manuscript).

For the validation of SOC and the overlaying water respiration, we compared the mean SOC and the mean SOC/overlaying water respiration between simulations and observations using bar plots (Fig. 4 and 5 in the updated manuscript).

For the validation of the time series of hypoxic area, we provided the averaged modeled and measured value in the time series plot (Fig. 8 in the updated manuscript) and also the 5-year running R^2 (Table 2 in the updated manuscript).

Manuscript assessment: both the Introduction and the Results/Discussion sections need some revisions. The Introduction review the literature of the northern Gulf of Mexico but does not assess what are the gaps in the knowledge. Rather, the authors propose technical improvements, which are welcomed but not sufficient. It is not clear, by the end of the manuscript, if using a more complex ecosystem model is an improvement over previous models. Although previous work is discussed relatively extensively in the Introduction, there is little discussion in the Results/Discussion section. Since similar studies have been carried out before, their results/findings should be compared. It would help to see what is the novelty of this study.

Response:

In the section “1 Introduction”, we added the following content in the last two paragraphs restating the main findings from the previous studies, the knowledge gaps, and the scientific questions we aimed to answer (Line 78-109). In this revision, we aimed to 1) understand the contributions of different factors in hypoxia development in different parts of the LaTex shelf and 2) to assess the outcomes of different riverine nutrient reduction scenarios regarding the reduction of the hypoxic zone within a complex lower-trophic ecosystem.

In the section “4.1 Factors controlling subregion bottom DO variability”, our results indicated that the variability of bottom DO on the LaTex shelf was mostly controlled by four processes: horizontal advection, vertical advection, vertical diffusion, and SOC (Fig. 9b in the updated manuscript). Their impacts on bottom DO variability varied in different regions across a year, which, to our best knowledge, has not been reported by other studies. Although the importance of DO advection and SOC on bottom DO balance was also documented by Ruiz Xomchuk et al. (2021), vertical diffusion was proposed as a minor contributor in their study. Such a disagreement could result from the water layers investigated. Vertical diffusion of DO across the layer 10 m above the bottom was discussed in Ruiz Xomchuk et al. (2021), while here we estimated vertical diffusion of DO across the bottom layer.

In the section “4.2 Stratification and Bottom DO Advection/Diffusion”, significant negative correlations were found between the PEA and the two absolute advection terms of bottom DO, while bottom DO flux due to vertical diffusion was found positively and moderately correlated to the PEA (Fig. 10 in the updated manuscript). Previous studies pointed out that water stratification affects the DO replenishment at the bottom layer, but did not provide evidence on how the stratification was correlated to the DO transport processes at bottom layer, especially for the non-linear relationship between PEA and DO vertical diffusion across the bottom layer (e.g., Hetland and DiMarco, 2008; Bianchi et al., 2010; Fennel et al., 2011, 2013, 2016; Justić and Wang, 2014; Wang and Justić, 2009; Feng et al., 2014; Yu et al., 2015; Laurent et al., 2018).

In the section “4.3 Riverine nutrient reductions”, we provided 16 sensitivity experiments (Table 3 in the updated manuscript) which suggested that reductions in riverine nutrients loads would not guarantee a reduction of hypoxia area due to the impacts of interactions among different species (e.g., competition on nutrients, grazing, and predation behaviors) in a complex lower-trophic community (Fig. 11–15 in the updated manuscript). Scenarios of nitrogen reduction even illustrated an increase in hypoxic areas. The results were different from previous studies which were mostly built upon a highly simplified lower-trophic community model (e.g., Fennel et al., 2006, 2011, 2013; Fennel and Laurent, 2018; Justić and Wang, 2014) or statistic models based on the relationship between hypoxia and total nitrogen loads (e.g., Scavia et al., 2013; Obenour et al., 2015; Turner et al., 2012; Laurent and Fennel, 2019). According to our simulations, it is expected that the 3-year mean hypoxic area can reach the hypoxia goal of 5000 km^2 if all nutrients (nitrogen, phosphorus, and silicon) are reduced by nearly 80% which is much more demanded than the recommended percentage indicated by previous model studies.

Specific comments:

L25: The rationale/discussion to support your study is not very convincing and also quite vague, you need to provide better arguments that explain why you conducted this research
Response: We restated the main findings from the previous studies, the knowledge gaps, and the scientific questions we aimed to answer in the Introduction section (see the above responses). We address the focus of this study as 1) the contributions of different factors in hypoxia evolution in different parts of the LaTex shelf; 2) the impacts of different riverine nutrient reduction scenarios on the variability of bottom DO and shelf hypoxia within a complex lower-trophic ecosystem.

L33-34: this is true only in a dual reduction strategy
Response: Yes. Have addressed it in L29-33 of the updated manuscript.

L46-48: All of these authors agree that SOC depends on organic matter in the sediment but because sediment OM is unknown they use a relationship between bottom O2, bottom temperature and SOC. They assume oxic respiration, which is why they find a direct relationship between SOC and bottom O2. Justic and Wang (2014) use a sediment tracer that depends on the abundance of deposited OM and is the source for SOC.
Response: We have rewritten this part as followed and updated it in L45–46:

L52-53: I don't understand this sentence. SOC would be overestimated at the peak of bloom and underestimated during the post-bloom period. This is probably what you meant to say but this is not what I read
Response: In the manuscript, we stated that in L49–L51,"However, the instantaneous parameterization tends to underestimate SOC at the peak of blooms yet overestimate SOC once the blooms started."

L57-58: This is why the models cited previously used a relationship with T/O2 or instant remineralization. I think what you try to say here is that these earlier parameterizations are not satisfactory and you will try to do better. You should discuss how your SOC implementation will be better than Justic and Wang (2014) because this is the most similar.
Response: Here, we want to address that we did not couple a sediment model for consideration of computation efficiency. The SOC scheme we adopted is based on Fennel's et al., (2006, 2011) scheme but with additional sediment, PON term to correct the misestimations by the instantaneous remineralization parameterization. Indeed, our SOC scheme is somewhat similar to Justic and Wang's (2014). But, in their model, the lower-trophic ecosystem supporting the modeled organic pool is simple. We added the comparison against their study in the manuscript (L98–103).

L66-68: This is because diatom is the dominant functional group, e.g. Murrell et al. (2014), Lehrter et al (2017). Also, the fact that these models are not a true representation of the reality is not the main point. Here you should point out what these models are doing wrong because of their simple representation of the phytoplankton community and why adding more groups of phytoplankton (and zooplankton) would improve the representation of oxygen sinks and hypoxia on the shelf. More is not always better.

Murrell et al: Murrell MC, Beddick DL, Devereux R, Greene RM, Hagy JD, Jarvis BM, Kurtz JC, Lehrter JC, Yates DF (2014) Gulf of Mexico hypoxia research program data report: 2002–2007. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-13/257
Lehrter et al: 10.1007/978-3-319-54571-4_8

Response: More is not always better. Yes, we strongly agree to it! In the Introduction section (L78–93), we added a summary of the previous studies and pointed out the knowledge gaps and the problems that a simple ecosystem model could cause.

In the Result and Discussion section, we provided sensitivity experiments to address the importance of the development of a more complex ecosystem in hypoxia simulation. Please refer to the above responses for the brief description and the manuscript for detailed discussion. In the validation section, we added validation for the diatom ratio (section "3.3 Diatom ratios"). We compared our simulated ratio of diatom (Table 1 in the updated manuscript) to two in-field studies by Schaeffer et al. (2012) and Chakraborty and Lohrenz (2015). The ratios provided by the former study are based on the biovolume of different phytoplankton groups, while those provided by the latter study are calculated by chlorophyll *a* attributed to different phytoplankton groups. We added a paragraph for the validation in L330–342.

The high contribution of the diatom group to the shelf phytoplankton community emphasizes that the inclusion of the diatom group in the numerical model is critical to well present the phytoplankton dynamics and the associated hypoxia events in the LaTex Shelf. Our model can reproduce well a diatom-dominated community on the shelf compared to cruise studies.

L79-80: there are lots of discussions about the factors controlling bottom O2 in the papers you cited above.
Response: We removed this sentence but stated the knowledge gap and the aim of this study in the Introduction section (L79–85 for knowledge gap and L103-106 for updated study objectives)

L85: you did not discuss silicate above
Response: We added the impacts of silicate on the bottom DO changes in the manuscript. Please see section "4.3 Riverine nutrient reductions" for details.

L90: what is there to see in the accompanying paper?
Response: The accompanying paper pointed out that statistical models were built based on the daily output by the coupled model in this study and successfully performed promising predictions of hypoxic area in the LaTex Shelf. Predicted hypoxic area showed a high agreement with the ROMS hindcast time series (RMSE=3256 km^2, R^2=0.7721). When compared to the shelf-wide cruise observations from 2012 to 2020, our prediction model provides a more accurate summer hypoxic area forecast than any existing forecast models with a high R^2 (0.9200); a low RMSE (2005 km^2); a low scatter index (15 %); and low mean absolute percentage biases for overall (18 %), fair-weather summer (15 %), and windy-summer (18 %) predictions. The accompanying paper has been published recently.

Ou, Y., Li, B., and Xue, Z. G.: Hydrodynamic and Biochemical Impacts on the Development of Hypoxia in the Louisiana–Texas Shelf Part II: Statistical Modeling and Hypoxia Prediction, Biogeosciences Discuss., 2022, 1–23, https://doi.org/10.5194/bg-2022-4, 2022.

L98: do you have sediment transport in your model (since you are using COAWST) and if so, why not having sediment biogeochemistry as in Moriarty et al (2018)
Response: Although we built our model on the COAWST, we did not include
The sediment transport model with a consideration of model efficiency. For a long-term hindcast purpose, we simplified the sedimental processes. We wanted to demonstrate the advances of a simplified treatment of sediment model in a long-term hypoxia hindcast.

L120: It is obvious why you want to add oxygen but you should discuss the addition of phosphorus, either here or in the introduction
Response: We have added a discussion on hypoxia responses to the nutrient reductions on phosphorus in section "4.3 Riverine nutrient reductions" for details.

L124: Can you develop? You mean phytoplankton and zooplankton are in N currency, but there is opal, DOP and DON
Response: Nitrate, ammonium, phosphate, and silicate are the limiting nutrient in phytoplankton growth. Opal, DOP, and DON are recycled back to silicate, phosphate, and ammonium during decomposition processes.

L126: can you provide a reference, a link to the observations? Would it be possible to get a time series of the observations in a supporting figure (for PO4, POP, DOP, silicate since they are new tracers? Also a map of all the gages would be useful since your model domain is quite large.
Response: The river gages and nutrient concentration measurements can be found in the USGS National Water Information System (https://maps.waterdata.usgs.gov/mapper/index.html). We added the locations of all river point sources in Fig. 2a (in the updated manuscript). The time series of river discharges of freshwater, nitrate, phosphate, and silicate from the Mississippi and Atchafalaya rivers was also added in Appendix C (Fig. C1 in the updated manuscript).

L129: I don't really understand what are your DOP and POP pools here (see next comment)
L138-139: These terms seem to be just
*dDOP/dt = dDON/dt * RPO4N*
*dPOP/dt = dPON/dt * RPO4N*
can you confirm? in this case you only have PO4 in your model
Response: The inclusion of DOP and POP pools here is to complete the phosphorus cycle in the model. Therefore, the changes of DOP and POP pools due to biochemical processes were set to follow those of DON and PON pools, respectively. And it is also a reason why we only consider a sinking velocity for PON but not for POP. The ratio of riverine DON/DOP and PON/POP may not follow the Redfield ratio. However, the measurements of DOP and POP are usually rare, we thus assumed the Redfield ratio is valid when measurements of the riverine DOP and POP are missed and applied the Redfield ratio for the reconstruction of riverine DOP and POP according to the measurements of DON and PON.

L161-172: please review this paragraph, the clarity could be improved
Response: We have reviewed this paragraph and have rewritten some sentences. Please see L177–213 in the updated manuscript.

L163-164: this is the opposite

Response: The estimated SOC is smaller at the bloom peaks but is larger after the peaks. We quoted Fennel's et al., (2013) comments below:

"An important limitation of this parameterization is that it neglects temporal delays in SOC which occur in nature and would result in smaller SOC at the height of blooms and larger SOC after bloom events in late summer and fall and further downstream from nutrient sources."

L164: Note for earlier that the formulations of Hetland and DiMarco (2008) and Lehrter et al (2011) include temperature, which mimics the delay because warmer water occurs after the peak of production

Response: Yes. But neither study considered the organic matter in their SOC schemes. Although the organic matter can be neglected in the SOC scheme if the empirical function is good enough, this is not the way we consider the improvement of the instantaneous remineralization parameterization scheme of SOC. We aimed to provide a scheme that more represents reality. Thus, we mimic the delay by adding a sedimental PON pool in the SOC scheme.

L180-187: this is a bit difficult to follow, could you make it easier?

Response: We think the most difficult part shall be the derivation of the fraction of denitrification-produced CO_2 to the total CO_2 production, x. Some assumptions were made (1) that only a portion of NH_4 provided by the aerobic respiration according to Eq. (R1) is used as the source element in the following nitrification according to Eq. (R2) and (2) that all NO_3 produced by nitrification is used as the source element in denitrification according to Eq. (R3). Such a portion of NO_3 can be explicitly provided by the linear relationship of denitrification and SOC rate shown in Eq. (5). Indeed, we can set the x as such a portion, but, by setting the x as the fraction of denitrification-produced CO_2 to the total CO_2 production, we can simplify the calculation and can also have the linear relationship of denitrification and SOC rate applied. Once we have the x determined (or the linear relationship of denitrification and SOC rate applied), we can derive the SOC and sedimental PON consumption as functions of M, the production rate of NH_4 by aerobic decomposition. We rewrote this part in the manuscript (L197–205).

L181: How come M is expressed in m-3 since it represents the integrated OM decomposition in the sediment. If you express it in m-2 you can remove the THKbot terms which simplifies the equations

Response: Yes. The production rate of NH_4 in sediment can be expressed in the real rate. However, we assumed that the oxygen consumed at the sedimentary layer contributes directly to the DO at the bottom water layer where DO is expressed in mmol m^{-3}. The THKbot is thus always needed in the equations. In the discussion, we aimed to compare the SOC rate, another rate of changes due to advection and diffusion, and the bottom DO total rate of changes, thus, we expressed these variables in the same volumetric unit.

L192: Do you use the same expression for the water column respiration?

Response: The expression for the water column respiration was slightly different since the concentration of PON used there was represented as mmolN m^{-3}. In other word, in water column, aerobic decomposition rate is:

$$M = PON \cdot VP2N_0 \cdot \exp(K_{P2N} \cdot TMP)$$

L199: do you have anaerobic respiration occurring in this case and if not, why?
Response: We did not consider the anaerobic respiration in the water column. Denitrification occurs as the oxygen level is low enough, however, the replenishment of DO in the water column is usually faster than in the sediment. Thus, although the water column could reach below hypoxia condition during some period, the fast replenishment would weaken the denitrification in the water column.

L211: although this seems fine for the plume region, it seems very short for the entire GoM and may influence you results as the interior GoM is still adjusting during your analysis period. The fact that hypoxia occur>100m later on suggests that this is the case. Also you need to show that your sediment layer reach a seasonal steady state (later on it seems to accumulate throughout the simulation near the Atchafalaya). How was the benthic layer initialized? can you provide a time series of PONsed?
Response: Our PONsed did reach a seasonal steady state starting from 2007 to 2020. We initialized the PONsed as 0 mmol m^{-2} due to the lack of observation. We are confident that our model spin-up period is long enough. In previous ROMS applications to the shelf, the spin-up period was usually as short as a few months even though the model was initialized with the averaged climatological profiles (e.g., Hetland and DiMarco, 2008; Fennel et al., 2011, 2013, 2016). For comparisons, in our model, initial conditions for the physical terms were derived from the Hybrid Coordinate Ocean Model (HYCOM) global analysis products. Initial conditions for concentrations of NO_3, PO_4, and $Si(OH)_4$ were interpolated from measurements provided by the World Ocean Database (WOD, Boyer et al., 2018). Initial conditions for DO concentration were given by World Ocean Atlas (WOA, Garcia et al., 2018). Other biochemical tracers were initialized as 0.1 mmol m^{-3} due to the lack of observations. With more realistic hydrodynamic initial conditions than previous ROMS applications, we are confident that a 5-month period for spin-ups is long enough. In addition, model validation for nutrient profiles between 2007 and 2009 illustrated our model was capable of reproducing the WOD profiles and was well spun up.

L226-230: do you do any nudging toward HYCOM or any other climatological product?
Response: We only nudged the salinity and water temperature toward HYCOM with inverse nudging coefficients decreasing from 1 day^{-1} to 1/60 day^{-1} from the open boundary to the interior. The inverse nudging coefficients for the interior (61% of grid cells including land cells) were set identically 1/60 day^{-1}.

L240: can you also show the other rivers for completeness?
Response: We have updated Figure 2a with the locations of all river point sources shown in the updated manuscript.

L245-246: can you elaborate on this assumption?
Response: The river discharges used as forcing in the model were measured not exactly at the river mouths, but at the lower reach of the main channel usually a few hundred kilometers from the mouths. Along the lowermost of the river channel, river discharges shall be greater than those where the measurements were conducted due to the water supplies from the adjacent watershed and lateral inflow of tributaries. Warner et al. (2005) took into account of such contributions by multiplying the discharges of measurements by 1.4 in their model study on the Hudson River. It is almost impossible to determine this factor for different rivers along the Gulf of Mexico since this factor is related to the distances from where the river gauges are deployed to the river mouths and lateral supplies by watershed and lateral inflow of tributaries. Such conditions vary

greatly among the rivers along the gulf. Factor 1.4 was not quite appropriate to all river systems but shall be considered in river systems like the Mississippi River and the Atchafalaya River systems which cover a large area of the watershed. For these two river systems, we chose 1.4 due to the lack of evidence showing the discharge ratio between the measurement spots and the river mouths.

L250: it is indeed highly oversaturated. can you provide some context?
Response: The riverine DO concentration was changed to be a constant of 258 mmol m^{-3} assuming that the riverine DO was saturated at 25 °C under 1 atm.

Figure 2c: The shelfwide surveys were not available prior to 2012? see here: https://coastalscience.noaa.gov/project/integrated-ecosystem-modeling-causes-hypoxia/
Response: We have downloaded the shelf-wide observation prior to 2012 according to the link provided and have expanded the validation accordingly. Plots were therefore updated (Fig. 2, 6 and, 7) in in the updated manuscript

L296: this is not a good comparison, you should provide histograms for surface data is spring, summer, winter. A 1:1 comparison would also be more meaningful because it would show where the mismatch occur (at low, high concentrations? in the bottom, at the surface?)
Figure 3c,f,i: this pair comparison is a bit misleading because you mix all data. Subsurface NO3 should be relatively small, resulting in a good agreement, but there could be significant mismatch at the surface. It is at the surface that a good representation of NO3 is important because that is where primary production occur
L301: Same comment here, I don't think this is a proper way to validate the model. Also, what about chlorophyll?
L283: I don't understand your choice of model data comparison. Are you binning the profiles by bathymetry? This assumes that the variability occurs from shallow (north) to deep (south) regions whereas the variability should be from upstream (east) to downstream (west). Also looking at Figure 3b it looks like vertical profiles of nitrate are uniform even though high nitrate at the surface (within the plume) is expected. Another issue is that you are mixing all times together. Your observed nutrient dataset is relatively short so you could make a better comparison, surface and bottom maps for example at key periods of the year

Response: We have updated all figures for profile validation. Please refer to the above responses. We did not provide validation for chlorophyll, instead, we provided validation for diatom (section "3.3 Diatom ratio" in the updated manuscript).

Regarding PO4, high values are mainly found near the bottom, which suggest that the main source of PO4 is from resuspension events rather than from the river. Can you justify these patterns?
Response: There is a lack of discussion of this issue in our manuscript. High PO_4 concentration at the near bottom was found both in the simulated and WOD profiles. We did not include the sediment module in our model, therefore, the high PO_4 concentration near the bottom should come from the recycling of DOP, POP, and sedimental organic matter (measured as PONsed). A similar phenomenon is found in the $Si(OH)_4$ profiles.

L315: the data are available, see earlier comments. These data also include nutrients which could be used in complement of WOD
Response: Comparisons of DO profiles between simulation and shelf-wide observations have been extended for the period prior to 2012 (Fig. 6 in the updated manuscript).

Figure 4c,f,i: I assume that some differences are much larger than 50% because if the model is normoxic and the observation hypoxic (or inversely) the bias could be several hundred percent

Response: Yes. We did have percentage differences greater than 50%. But there are already 72 % of WOD profiles, 66 % of shelf-wide profiles, and 92 % of SEAMAP profiles being misestimated by within 50%. Our model was capable of reproducing the DO profiles well.

Figure 4h: Aren't SEAMAP cruise occurring in late spring rather than summer?

Response: SEAMAP cruise occurred from June to July. But the measurements were not always carried out within the LaTex Shelf as shown in Figure 2c (red dots) in the updated manuscript.

L335: I don't know why the model data <10m are not shown in Figure 6, these data should be available to the reader

L336: this is not true for the area off the Atchafalaya, observations are available there

Response: The spatial distribution of bottom DO concentration was shown over depth from 6 to 50 m (see Fig. 7 in the updated manuscript).

L337: 2017 as well. Can you comment on the occurrence of hypoxia around 100m (near the slope). Is that an issue in the model, i.e. does that influence hypoxia on the shelf?

Response: The hypoxia around 100 m is due to the overestimation of the sinking velocity of the organic matter. We have rerun the model with a new parameter (sinking velocity=5 m day^{-1}). In this submission, we focused on the validation over the waters where shelf-wide measurements were conducted. According to the new results, we rewrote this part (L389–403).

L349: why 10m? I agree that you should exclude the Atchafalaya Bay but you should include the coastal area. Also, you should have a more restrictive longitudinal extent because the observations are always <94.5W

Response: The modeled hypoxic area (Fig. 8 in the updated manuscript) was re-estimated over the shelf with depth from 6 to 50 m (Fig. 2b in the updated manuscript).

L349-353: In some years the model simulates a relatively large hypoxic area in June, sometimes also in May, do you think this is realistic? Are the SEAMAP data showing similar conditions?

Also, bottom waters don't always get fully reoxygenated in July-August in years with tropical storms/hurricanes, e.g. 2018-2020. Can you comment?

Response: The simulated hypoxic area is affected by the studied area we chose (here waters with depths from 6 to 50 m). However, no matter for which area is selected, we still cannot provide a comparison for days without observations. We showed the daily time series of the hypoxic area here to emphasize that the hypoxic area can change during a short time period, which suggests that more cruise observations are needed to depict the whole picture of the shelf hypoxia.

We do not have much hypoxia observed by the SEAMAP, since the SEAMAP cruises cover a larger spatial area with less observation in the LaTex Shelf during summer. Bottom waters don't always get fully reoxygenated in July-August. It may result from our relatively large study area and features of hurricanes (e.g., tracks, intensity, and translation speeds). In mid-July 2019, hurricane Barry stroke coastal Louisiana as a category 1 level hurricane. It was a fast-moving and relatively weak hurricane which may not lead to fully reoxygenated bottom waters, especially after a massive hypoxia event in early July. However, more discussions are needed on hurricane influences which are out of the scope of this study.

Figure 6: 1) Another way to make this comparison would be to overlay the observations as scatter points over the model maps
Response: We have updated this plot (Fig. 7 in the updated manuscript) accordingly.

2) hypoxia varies rapidly and it might be better to show a mid-cruise map from the model rather than a ~1 week average
Response: The model bottom DO shown for the comparison purpose was not a ~1-week average, but a ~1-week composite according to the cruise locations on different days. We added a description of the composite plot in the first paragraph of section 3.6 (Line 390–393).

3) can you show the other years for completeness?
Response: We have added comparisons in other years for completeness.

L364: you use a mixed format for Results and Discussion but then you do not discuss much your results with respect to the literature
Response: We have updated the section on Results and Discussion. Previous contents were replaced by the following section "4.1 Factors controlling subregion bottom DO variability", "4.2 Stratification and Bottom DO Advection/Diffusion", and "4.3 Riverine nutrient reductions". Please see our response above.

L375: I don't quite follow this analysis, what does it mean?
It looks like there is a long term negative trend in the Atchafalaya plume and offshore. The 2 signals could be problematic: the Atchafalaya plume signal indicate that PONsed accumulates there during the simulation and the offshore signal seems to indicate that there is a drift in offshore subsurface O2 or that the offshore part of the model is still adjusting
Note: you don't have resuspension in your model. Can you justify your choice? this feature would be easy to implement and would provide a realistic distribution of SOC over the shelf. This may also prevent the accumulation of PONsed near the Atchafalaya.
L408: see earlier comment about the long term trend
Response: We have removed this part. PONsed did not accumulate near the Atchafalaya nearshore regions. Please see the responses above for the model spin-up issue.

L380-390: can you compare these patterns with the literature?
Response: We have removed this part.

L385: This is surprising that you find substantial hypoxia in a monthly climatology. This means that 1) hypoxia almost always occur at this location during that month (as shown on the right panels) and/or 2) bottom O2 concentrations are low at these locations, well below the hypoxia threshold.
Response: We have removed this part.

L450: also vertical diffusion and possibly horizontal advection, as well as SOC
Response: See the updated section "4.1 Factors controlling subregion bottom DO variability" where we compared the contribution of bottom DO advection, bottom DO diffusions, and biochemical terms on the bottom DO changes.

L456-457: you should compare your results with these. For that you should integrate respiration over the subsurface layer (or lower 4m for instance). You could also discuss your results with respect to other budgets, e.g. Yu el al (2015)

Response: We have added the validation of SOC and ratio of SOC/ overlaying water respiration. We discussed the contribution of different controlling factors, in the section "4.1 Factors controlling subregion bottom DO variability".

L477-478: this is not obvious

Figure 12a,c,e,g: I think the time series in Figure 10 were enough. I don't find these PEA maps very useful

Response: We have removed this part.

L498-511: this paragraph should go in the Methods section

Response: We have removed this part.

L513-527: other authors found that water column respiration is not dominant but not negligible either (Lehrter et al, Yu et al), can you comment on that? Is the large dominance of SOC in your model due to the set up of your model, high settling rate for instance?

Response: We have removed this part. The high settling rate of organic matter indeed affects the results. We reduced the sinking velocity to 5 m day^{-1} and provided a reasonable range of SOC and SOC/overlaying water column respiration.

L517: yes, this is where you find persistent hypoxia

Response: PONsed did not accumulate near the Atchafalaya nearshore regions. Please see the responses above for the model spin-up issue.

L522: where is this shown? you speculate here

Response: We have removed this part.

L524: +10% would be a more conservative value and used for climate projections in the region, e.g. Lehrter et al 2017.

Response: We have removed this part.

L525: you speculate here

Response: We have removed this part.

L543: ah yes, that explains the very low water column respiration, see earlier comment.

In the Atchafalaya nearshore, PON settles instantly to the bottom and accumulate which explains SOC and hypoxia there. I think this is problematic as your model setup drives your conclusions. This brings up two points: 1) you should validate your choice of high settling rate. For instance if surface nutrients, surface chlorophyll, water column respiration and SOC compare well with the observations/literature then your choice is fine. If not then you may want to recalibrate your model. 2) if PON sinks rapidly to the bottom and water column respiration is not significant, then why do you have 3 functional types of zooplankton?

Note: with this type of model setup the predatory zooplankton tend to have a top-down control over primary producers, is this the case in your system and is this why the sinking rate is so high, to escape this control?

Response: By sensitivity experiments, we have changed the sinking velocity to 5 m day^{-1} as we have shown in this document. We introduced a complex ecosystem in this study aiming to test its impacts on hypoxia dynamics.

L543-555: I don't get the point of this paragraph
Figure 15: I don't get the point of this figure
Response: We have removed this part.

L569-570 (see also earlier comment): Given your fast sinking environment it seems that a single functional group for phytoplankton (diatom) and zooplankton was enough in you study of the LATEX shelf. A more convincing argument for your model choice would be that it is needed for the open ocean part of your domain (if indeed it is)
Response: In this study, we prescribed different phytoplankton and zooplankton functional groups to capture the system sensitivity to different nutrient limitation scenarios, and the benefit of utilizing different zooplankton groups has already been illustrated by Shropshire et al. (2020). Their study highlighted the NEMURO with a complex zooplankton community structure can reproduce well the zooplankton dynamic in the Gulf of Mexico. According to the updated simulation and discussion, the complexity of the ecosystem does affect the hypoxia responses to the different nutrient reduction scenarios suggesting that changes in hypoxia are not straightforward as what has been shown in previous studies.

We have changed our sinking velocity to 5 m day^{-1} which provides a more reasonable range of SOC and SOC/overlaying water respiration.

Reference:
Shropshire, T. A., Morey, S. L., Chassignet, E. P., Bozec, A., Coles, V. J., Landry, M. R., ... & Stukel, M. R. (2020). Quantifying spatiotemporal variability in zooplankton dynamics in the Gulf of Mexico with a physical–biogeochemical model. *Biogeosciences*, *17*(13), 3385-3407.

L571: P limitation: you did not show that either
Response: We did not include the P limitation discussion in our study since P limitation has already been discussed and deemed to be important to the shelf hypoxia by Laurent and Fennel (2014) and Fennel and Laurent (2018). However, we provided sensitivity tests for a nutrient reduction on P and found that the percentage of reduction in the hypoxic area was close to the finding by Laurent and Fennel (2014). Please see the updated section 4.3 (L571–580).

L572-573: this was the main novelty of this work. However, model tuning may be necessary to properly reproduce water column respiration (see also earlier comment)
L573: you did not show that, see earlier comments
Response: We have validated the SOC and SOC/overlaying water respiration for the updated results with the sinking velocity of 5 m day^{-1}.

L627-628: The model does not include a light attenuation factor for terrigenous material near the river (dependent on salinity for instance)? Light limitation is strong near the Mississippi and Atchafalaya River mouths but this light limitation effect is not included in your model. This lack of light limitation would result in high primary production near the river mouths and less production downstream, thereby influencing the timing and distribution of phytoplankton, respiration and bottom oxygen over the shelf Also (and L638), why is PAR different for small and large phytoplankton? shouldn't it be the same, each functional type having a different sensitivity to light? Looking at your parameter table I see that you are using the same value for both so effectively there is no difference in PAR
Response: The inclusion of light limitation due to riverine sediments shall be an improvement of our model. We have validated the surface nutrients (0–5 m) concentration and found the observations can be reproduced well by our model (see above responses). For the light limitation due to phytoplankton self-shading effects, the

original codes in NEMURO were written with the AttPS and AttPL separated. We did not change this part of the codes but set the AttPS and AttPL the same assuming each functional type has the same sensitivity to light.

L650: L650: did you mention how these parameters were chosen? were they calibrated to the Gulf of Mexico?
Response: The parameters largely followed the set-ups by Shropshire et al. (2020), Laurent et al. (2012), Fennel et al. (2006), and Fennel et al. (2013). Details can be found in the updated Table B4 in the updated manuscript.

Minor comments/typos:

L1: "impact" is not the right wording

L30: shrinking is not the right word, reduction is better. Please rephrase the sentence accordingly

L34: "shrinkage" is not the right word, may be "reduction"?

L34-35: replace with: "Transient phosphorus limitation on the shelf (Laurent et al 2012; Sylvan et al 2007) was deemed..."
Sylvan et al: 10.4319/lo.2007.52.6.2679

L35: "with the delayed onset and reduction of the hypoxic area"

L39: Conley et al 2009 is not related to the LATEX shelf

L56: "coupled"

L93: you could mention your main results here.

L123: I don't think you need this reference as this formulation is wide spread. However, you could mention that you use the same formulation as for the other nutrients

L162: please rephrase, the sentence is not complete

L332: I agree but you could mention the underestimation of the hypoxic layer

L345-346: you did not introduce Figure 7 yet

L377: this makes sense, the STDs are larger in the plume region where hypoxia occurs

L381: that seems normal since the hypoxic area is calculated from bottom O2
Figure 8e: the DO scale is a bit misleading
Why do you show bottom oxygen up to 100m in Figure 6 but then limit the output to 50m in Figures 8-9, 11-12?

L400: yes because the extent is a climatology (see comment above)

L414/446 (and elsewhere): "trough": minimum may be a better word (elsewhere as well)
Figure 9: can you show the results for the coastal area when you show maps?

L448: "also water stratification (Figure 10)."

L450: be more accurate, here you talk about water column processes
Figure 11: Since you don't compare modeled SOC with observations it would be easier to keep the original units

L468: Note that the maps show a nearshore/offshore gradient in PEA, following the bathymetry. This is due to the multiplier z in the PEA equation, which increases PEA with increasing bathymetry

L471: may be 1 reference is sufficient here?

L537: replace "low" by small

L568: "the NEMURO model"

Response: We have corrected all the typos and inappropriate usage of words according to the comments.

Responses to Comments by Referee #2

General comments:
This manuscript by Ou et al. utilized a coupled physical-biogeochemical model to investigate the controlling factor of bottom hypoxia on the northern Gulf of Mexico and Louisiana-Texas Shelf. The authors added the phosphorus cycle and modified the sediment oxygen consumption module in an existing biogeochemical model NEMURO and coupled it with ROMS model. The coupled model was validated with observational data and then used to implement a 15-year hindcast simulation during 2006-2020. Then the authors explored the spatial variation of hypoxia development in the study area and found sediment oxygen consumption (SOC) and water column stratification are main factors to control the bottom oxygen in nearshore and offshore area respectively. Their model results also indicated separate hypoxia development schemes on the west and east Louisiana-Texas Shelf. Coastal deoxygenation is one of the most prominent environmental issues with important implications for marine ecosystem services. Although this paper made efforts to adopt a more sophisticated biogeochemical model with added phosphorus cycle and improved sediment oxygen consumption module, making contributions to investigate the spatial differences of dominant processes on hypoxia, it lacks original and novel aspects to explore the well-studied topic in this region, as well as comprehensive comparison with previous modeling study on the model performance, simulation results and conclusions, and address the question that how this new model stands out. The manuscript missed an advanced understanding and deep insight on the research topic of coastal hypoxia in a well-organized discussion section, thus this paper is a little thin on content. Although I see the value of this work, I perceive that the publication is premature at this time.

Response: We thank Reviewer #2's comments. Compared with existing modeling efforts, our model, for the first time, included a silicate cycle as well as multiple plankton functional groups in the modified biogeochemical model, the importance of which has already been addressed in previous studies yet not included in hypoxia modeling efforts. The extensive model validation against nutrient and dissolved oxygen profiles confirmed the robustness of our model. In addition, another purpose of this model study is to provide the needed numerical solution for the development of a novel statistical model presented in paper Part 2.

We agree that our findings presented in the first submission need some improvement. In this round, we performed a comprehensive revision of the manuscript. The main changes in the manuscript are listed below.

Changes in model set-ups:
1. We performed a series of sensitivity runs to evaluate the model's robustness regarding different parameterizations of the sinking velocity of the organic matter (i.e., PON and Opal). And we conclude that a sinking velocity of 5 m/day is a reasonable prescription and updated relevant model results.
2. The riverine DO concentration was changed to be a constant of 258 mmol m$^{-3,}$ assuming that the riverine DO was saturated at 25 °C under 1 atm.

Content removed:
1. The model validation for hypoxic thickness was removed since we focused more on the bottom DO dynamic.

2. We rewrote the entire “Result and Discussion” section. Both reviewers pointed out that the manuscript lacked originality. In this revision, we focus on 1) the contributions of different biogeochemical and hydrodynamic terms on bottom DO variability in different subregions, 2) the impacts of diatom and the complexity of lower-trophic community on the hypoxia dynamics, and 3) bottom DO’s responses to the reduction of riverine nutrient loads with different reduction combination (percentage and nutrient type).

<u>Content added:</u>
1. The model validations of diatom ratios, sediment oxygen consumption (SOC), and ratios of SOC/ overlaying water respiration were added following the reviewers’ suggestions.
2. Results and discussion on the factors to bottom DO variability was added (see section “4.1 Factors controlling subregion bottom DO variability” and “4.2 Stratification and Bottom DO Advection/Diffusion”). All terms that directly contribute to bottom DO changes were calculated and evaluated, which include horizontal advection, vertical advection, horizontal diffusion, vertical diffusion, the local rate of change (bottom DO), SOC, and DO changes due to water column biochemistry at the bottom layer. The summation of these terms contributes to the total changes in the bottom DO. A comparison of their contributions was given in different subregions of the LaTex shelf. We found that the most prevailing factors to the bottom DO changes are the two advection terms, vertical diffusion term and SOC. The former three terms were associated with the changes in water column stratification. The strong linear correlations between PEA and the advection terms suggest that increased water stability in summer leads to fewer DO exchanges from advection processes. Nevertheless, the relationship between PEA and vertical diffusion of DO across the bottom layer appears to be non-linear.
3. Sensitivity experiments for riverine nutrient reduction strategy were added (see section “4.3 Riverine nutrient reductions”). We found the responses of summer hypoxia to the changing nutrient loads are not linear due to the impacts of the complexity of the lower-trophic community. Nutrient reductions do not always guarantee a reduction in summer hypoxic area, instead, due to the interactions among different plankton groups (e.g., competition on nutrients, grazing, and predation behaviors), the hypoxic area could even increase under some nutrient reduction conditions. The most effective strategy is to reduce the nitrogen, phosphorus, and silicon loads simultaneously. A triple riverine nutrient reduction of 80% can help to fulfill the hypoxia reduction goal of 5000 km^2.

Major comments:
The hypoxia at the northern Gulf of Mexico has been well studied since the 1990s with
increasing model studies in recent years. It ranged from a simple oxygen respiration model (Hetland&DiMarco, 2008) to a sophisticated coupled biogeochemical model (Laurent et al. 2012; Fennel et al. 2013). Including this study, they all generated similar conclusions that SOC is the controlling factor for hypoxia. In this sense, the improvement of complexity in the biogeochemical model does not make much sense. Also, the authors mentioned the additional work done on the NEMURO-based model filled gaps in phosphorus cycling and improved SOC representation. It’s better to prove

the advancement of the new model by validating with important variables, such as DO,
Chla, PO4, NO3, with other model simulation studies.
Response: In the validation part, we added validation of SOC (L349–373, Fig. 4 in the updated manuscript), the ratio of SOC/overlaying water respiration (Fig. 5 in the updated manuscript), and the ratio of diatom/total phytoplankton against measurements (L329–348, Table 1).
We updated our results and focus on the direct impacts of biochemistry and hydrodynamics on the bottom DO variability (section "4.1 Factors controlling subregion bottom DO variability" and "4.2 Stratification and Bottom DO Advection/Diffusion" in the updated manuscript), the impacts of diatom and the complexity of lower-trophic community on the hypoxia dynamics including the impacts on the responses of hypoxia changes to the decreasing riverine nutrient loads (section "4.3 Riverine nutrient reductions")

The oxygen balance analysis is confusing and questionable. Although SOC is the dominant process in the bottom hypoxia generation (Yu et al. 2015), water column respiration (WCR) should not be orders of magnitude smaller than SOC, especially in the whole water column, as shown in Figure 15 and L455-456. Observational studies still showed varying evidence on SOC contribution (Murrell&Lehrter, 2011; Quiñones-
Rivera, et al. 2010). More importantly, the reviewer has a sense that the authors did not understand and explain the oxygen dynamics well (Figure 10 and 15, section 4.2). What is oxygen balance in the text? Based on L450-452, it should be water column respiration plus phytoplankton photosynthesis. This is a very confusing term and the physical transport of oxygen was totally missing. A lot of oxygen studies utilized standard oxygen budget analysis to separate dynamic terms in oxygen change (Li et al.
2014; Scully 2013; Yu et al. 2015). Please refer to those studies on the analysis and consider recalculating/rewriting this part.
Response: We agree with Reviewer#2 that the impact of WCR should not be neglected. In the revision we validated the SOC and ratio of SOC/overlaying water respiration in the section "3.4 SOC and overlaying water respiration". For the factors controlling bottom DO variability (sections 4.1 and 4.2 in the updated manuscript), contributions of five hydrodynamic-related terms and two biochemical-related terms were discussed.

Although this study employed sophisticated machine learning techniques to determine the controlling mechanisms on hypoxia in different regions. It could be actually achieved by oxygen budget analysis, with much clear representation in physical terms (advection and diffusion), rather than relying on stratification indicators. In addition, compared to the manipulating force on DO variability on a seasonal scale, the interannual variability is more of interest and worthy to look into.
Response: We removed the machine learning part and rewrote the sections which discuss the controlling factors of hypoxia (please see sections 4.1 and 4.2 in the updated manuscript).

The manuscript missed a comprehensive discussion section of advanced understanding *of the study topic in-depth and in breadth. The overview of previous observational and model studies in this region, comparison with the current study, what are the agreements and differences, what are the causes, what are the defective aspects of this study, etc. are all important points to include. Expanding the implication to the global context is also valuable to discuss.*

Response: In the updated section "4 Results and discussion", we moved our focus to 1) the contributions of different factors in hypoxia evolution in different parts of the LaTex shelf; 2) the impacts of different riverine nutrient reduction scenarios on the variability of bottom DO and shelf hypoxia within a complex lower-trophic ecosystem.

In the section "4.1 Factors controlling subregion bottom DO variability", our results indicated that variability of bottom DO on the LaTex shelf was mostly controlled by four processes: horizontal advection, vertical advection, vertical diffusion, and SOC. Their impacts on bottom DO variability varied over different subregions over the year. Such results have not yet been pointed out by previous studies. On the other hand, although the importance of DO advection and SOC on bottom DO balance was also documented by Ruiz Xomchuk et al. (2021), vertical diffusion was proposed as a minor contributor in their study. Such a disagreement could result from the water layers investigated. Vertical diffusion of DO across the layer 10 m above the bottom was discussed in Ruiz Xomchuk et al. (2021), while here we estimated vertical diffusion of DO across the bottom layer.

In the section "4.2 Stratification and Bottom DO Advection/Diffusion", significant negative correlations were found between the PEA and the two absolute advection terms of bottom DO, while bottom DO flux due to vertical diffusion was found positively and moderately correlated to the PEA (Fig. 10). Previous studies pointed out that water stratification affect the DO replenishment at the bottom layer, but did not provided evidence on how the stratification was correlated to the DO transport processes at bottom layer especially for the non-linear relationship between PEA and DO vertical diffusion across the bottom layer (e.g., Hetland and DiMarco, 2008; Bianchi et al., 2010; Fennel et al., 2011, 2013, 2016; Justić and Wang, 2014; Wang and Justić, 2009; Feng et al., 2014; Yu et al., 2015; Laurent et al., 2018).

In the section "4.3 Riverine nutrient reductions", we provided 16 sensitivity experiments (Table 3, also see the above responses) which suggested that reductions on riverine nutrients loads would not guarantee a reduction of hypoxia area due to the impacts of interactions among different species (e.g., competition on nutrients, grazing, and predation behaviors) in a complex lower-trophic community. Scenarios of the nitrogen reduction even illustrated an increase in hypoxic area. The results were opposite to previous coupled numerical studies which were built upon a highly simplified lower-trophic community model (e.g., Fennel et al., 2006, 2011, 2013; Fennel and Laurent, 2018; Justić and Wang, 2014) and statistic models which were developed based on the relationship between hypoxia and total nitrogen loads (e.g., Scavia et al., 2013; Obenour et al., 2015; Turner et al., 2012; Laurent and Fennel, 2019). According to our simulations, it is expected that the 3-year mean hypoxic area can reach the hypoxia goal of 5000 km^2 if all nutrients (nitrogen, phosphorus, and silicon) are reduced by nearly 80% which is much more demanded than the recommended percentage indicated by previous model studies.

Detailed comments:
Method

L105-106: are the new features of this biogeochemical model suitable for NGoM?
Response: For the zooplankton modeling configurations, we followed Shropshire et al. (2020)'s works focusing on the zooplankton dynamics in the Gulf of Mexico, which use a similar model setup. According to that study, the zooplankton community could substantially affect the primary production in the study area.

Reference:
Shropshire, T. A., Morey, S. L., Chassignet, E. P., Bozec, A., Coles, V. J., Landry, M. R., ... & Stukel, M. R. (2020). Quantifying spatiotemporal variability in zooplankton dynamics in the Gulf of Mexico with a physical–biogeochemical model. Biogeosciences, 17(13), 3385-3407.

L108: what is PL? should it be LP (large phytoplankton)?
Response: We have corrected all typos related to abbreviations of plankton group, that is small phytoplankton (PS), large phytoplankton (PL), small zooplankton (ZS), large zooplankton (ZL), and predatory zooplankton (ZP).

L120-122: no reactive, labile and refractory category in organic matter pool? In other words, is a single reaction rate enough?
Response: We included a burial PON pool in the conceptual sedimental layer. PON settles down at the conceptual sedimental layer fuels the PONsed pool, which is a reactive labile organic sediment pool. After a portion of PONsed is decomposed during aerobic and anaerobic processes in sediment (see the SOC scheme), a certain portion of PONsed is burial and fueling the PONburial pool, which will be removed from the system.

L156: What are ExcZS, ExcZL and ExcZP represented (I could not find those in the Appendix, and guess they should be zooplankton excretion rate to NH4?)? Why not include the zooplankton respiration term?
Response: ExcZS, ExcZL and ExcZP are typos in the equation. They should be ExcZSn, ExcZLn, and ExcZPn, respectively as shown in Table B2. In our model, we combined zooplankton excretion and respiration. Thus, during excretion, zooplankton consumes oxygen.

L158-159: How did oxygen inhibition on nitrification and aerobic decomposition rates were calculated? Using Michaelis–Menten formula?
Response: The oxygen inhibition (Fennel et al., 2006; 2013) is considered as the maximum of 0 and an oxygen-dependent unitless term. It uses Michaelis–Menten formula. The inhibition term ($\hat{r}$) is described in A2 and A3 with the related parameter description in Table B4.

$$\hat{r} = max\left[\frac{max(0, Oxyg - Oxyg_{th})}{K_{Oxyg} + Oxyg - Oxyg_{th}}, 0\right]$$

Where

Parameter	Description	Units	Values
K_{Oxyg}	Oxygen concentration at which inhibition of nitrification and aerobic respiration are half-saturated	mmolO$_2$ m^{-3}	3.0

$Oxyg_{th}$	Oxygen concentration threshold below which no aerobic respiration or nitrification occurs	$mmolO_2\ m^{-3}$	6.0

Reference:
Fennel, K., Wilkin, J., Levin, J., Moisan, J., O'Reilly, J., and Haidvogel, D.: Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget, Global Biogeochem. Cycles, 20, 1–14, https://doi.org/10.1029/2005GB002456, 2006.

Fennel, K., Hu, J., Laurent, A., Marta-Almeida, M., and Hetland, R.: Sensitivity of hypoxia predictions for the northern Gulf of Mexico to sediment oxygen consumption and model nesting, J. Geophys. Res. Ocean., 118, 990–1002, https://doi.org/10.1002/jgrc.20077, 2013.

L164-166: how was the portion of sinking PON buried (PONburial) determined? How the initial sediment PON pool was calculated? Is there also an anaerobic layer? Is there any exchange between PONburial and PONsed?

Response: The burial faction is determined using the scheme embedded in the original NEMURO model, where the burial faction is a function of the vertical flux of particulate organic matter. As the organic matter is buried, it will leave the system without returning to PONsed.

The PONsed is initialized as 0 due to a lack of available data to initialize the model. Our model does not include a sediment module, thus, the sedimentary PON pool is in an imaginary or conceptual sediment layer. In this layer, aerobic decomposition, nitrification, and denitrification occur simultaneously following the linear relationship between denitrification rate and total oxygen consumption rate (Eq. (5)). So, there is no specified anaerobic layer.

L193: the description of THKbot is confusing. Is it the thickness of overlying water, or sediment layer?

Response: THKbot is the thickness of overlying water or the thickness of the bottom layer of the ocean model. In our model, we do not separate overlaying water and bottom water. We consider THKbot since we assumed that oxygen consumption at the conceptual sediment layer directly contributes to decreases in oxygen concentration at the bottom water layer.

L195: SOC/THKbot is basically the oxygen consumption rate in the sediment. Why not integrate SOC in the hypoxic area and get an overall integrated SOC?
Any observational data validation on the newly added sediment and phosphorus module? In addition to the oxygen concentration validation?

Response: We assumed that oxygen consumption at the conceptual sediment layer directly contributes to decreases in oxygen concentration (only) at the bottom water layer. It also implies that the oxygen consumed in the sediment is from the bottom water layer. For further comparison with other terms of DO changes (e.g., advection and diffusion) expressed in a volumetric unit, we transformed the areal SOC rate ($mmolO_2\ m^{-2}\ day^{-1}$) to a volumetric rate by dividing the THKbot.

For the newly added phosphorus module, we have added validation for PO_4 profiles against WOD observation in section "3.2 Nutrients concentration profiles". Comparisons showed that our model was capable of reproducing the measured PO_4.

For the validation of SOC, we added the section "3.4 SOC and overlaying water respiration" in the manuscript.

L211: is 5 months enough for spin-up in this area? What is the initial condition (cold start or hot start)?
Response: We are confident that our model spin-up period is long enough. In previous ROMS applications to the shelf, the spin-up period was usually as short as a few months even though the model was initialized with the averaged climatological profiles (e.g., Hetland and DiMarco, 2008; Fennel et al., 2011, 2013, 2016). For comparisons, in our model, initial conditions for the physical terms were derived from the Hybrid Coordinate Ocean Model (HYCOM) global analysis products. Initial conditions for concentrations of NO_3, PO_4, and $Si(OH)_4$ were interpolated from measurements provided by the World Ocean Database (WOD, Boyer et al., 2018). Initial conditions for DO concentration were given by World Ocean Atlas (WOA, Garcia et al., 2018). Other biochemical tracers were initialized as 0.1 mmol m^{-3} due to the lack of observations. With more realistic hydrodynamic initial conditions than previous ROMS applications, we are confident that a 5-month period for spin-ups is long enough. In addition, model validation for nutrient profiles between 2007 and 2009 illustrated our model was capable of reproducing the WOD profiles and was well spun up.
Biogeochemical model validations
The entire validation is qualitative rather than quantitative. Need statistic metrics to assess the overall model performance, i.e. taylor and target diagram.
Response: We replotted all the figures for model validation of nutrient and oxygen profiles (Fig. 3 and Fig. 6 in the updated manuscript). Probability histograms of nutrient differences between modeled and observed results were given for the upper 50 meters and upper 5 meters, respectively. Probability histograms of DO differences were given for the upper 50 meters.

For the validation of diatom ratio, we compared the mean±1SD between simulation and observation (Table 1 in the updated manuscript).

For the validation of SOC and the overlaying water respiration, we compared the mean SOC and the mean SOC/overlaying water respiration between simulations and observations using bar plots (Fig. 4 and 5 in the updated manuscript).

For the validation of time series of hypoxic area, we provided the averaged modeled and measured value in the time series plot (Fig. 8) and also the 5-year running R^2 in Table 2.

Figure 3: which cross-section was compared in Figure 2b? The difference histogram in (c)(f)(i) is vertically averaged or bottom value?
Response: We have replotted Fig. 3. Please see the above responses.

L287-288: both NO3 and PO4 were overestimated
Response: Yes. But the biases are slight when compared to the riverine supplies.

L295-296: this statement is a bit questionable that the high riverine nutrient concentration may not be the cause for the model-observation bias. Because the high concentration of PO4 and Si(OH)4 is at the bottom which indicates that it is nutrient regeneration, rather than the allochthonous source.
What are the causes for the hot points (with bottom high nutrient concentration) of PO4
and Si(OH)4?

Response: We did not attribute the nutrient bias to the high riverine nutrient concentration, instead, we wanted to emphasize that such bias was acceptable regarding to the strong influences of high riverine nutrient loads.

We did not include the sediment module in our model, therefore, the high PO_4 and $Si(OH)_4$ concentration in the bottom layers could be a result of the recycled DOP, POP, sedimental organic matter (measured as PONsed), and opal. Specifically, the high peak of $Si(OH)_4$ concentration occurred at around 35 m depth was consistent with biogenic silica remineralization at lower water columns (Baronas et al., 2016).

L303-304: model overestimates DO while also overestimating the recycled nutrient concentration. Usually, it is the opposite case since nutrient remineralization is associated with oxygen consumption. Any explanations?

Response: Based on the sensitivity experiments, nutrient reductions do not always lead to increases in bottom DO due to the interaction among different species (e.g., competition on nutrients, grazing, and predation behaviors) in a complex lower-trophic community. In other words, the responses of DO changes to nutrient changes are not straightforward as what was suggested by previous numerical studies relying on highly simplified ecosystem models. Therefore, although the model overestimated the recycled nutrient concentration (i.e., NO_3 and PO_4) we would expect to have an overestimation in DO.

L331-332: in section 3.4 model validation of oxygen, the result suggested that the model overestimated DO and hypoxia was more frequent in observed WOD profiles. Why here the modeled hypoxia thickness (<=4m) is greater than observed profiles?

Response: We have removed the validation for hypoxic thickness since we did not provide a discussion on hypoxic volume but focused on the DO at the bottom water layer.

L336-337: the model showed more offshore extension of hypoxia than observation. Any possible causes?

Response: Using the updated model set-ups (sinking velocity of organic matter was changed to 5 m day^{-1} from 15 m day^{-1}), such inconsistency seldom occurred. There is no model that can reproduce exactly the extension of the hypoxia area documented by ship-based observations. The model simulated more offshore extension (e.g., 2011) of hypoxia may result from the overestimated offshore transport of water and materiel due to a relatively coarse spatial model resolution (~5 km), parameterization of advection and diffusion processes, and the coarse spatial resolutions of atmospheric forcings. Compared to the existing modeling studies, we are confident that our model performed is pretty robust.

L346-347: the hypoxia area was separated around 92.5W instead of 91W shown in the
model simulation? This may reveal a certain defect in the dynamics of model simulation in oxygen.

Response: Based on the updated simulations, we would note that the western and eastern hypoxic water were not always merged but were separated at around 91 °W (e.g., 2007, 2010, 2012, 2014, 2017, and 2018; Fig. 7 in the updated manuscript).

L349: why not include hypoxia area in the water depth<10m?

Response: We have replotted the spatial distribution of bottom DO (see Fig. 7 in the updated manuscript) covering depths from 6 to 50 m.

L346: the order of figure citation is a bit messy; the figure should be numbered according to the order of citation, not the other way around. For example, the order of Figure 10 is not optimal for reference.
Response: We have renumbered the figures according to the order how the results were shown.
Figure 7: please adjust the x-axis as the other years for better comparison.
Response: We have adjusted the x-axis according to the comment (see Fig. 8 in the updated manuscript).

L351-352: this means no apparent bias of model simulation in the hypoxia area. How is this model performance compared to other model studies in this region?
Response: Previous coupled models did not provide validation of hypoxic areas with a time record as long as ours. So, we did not add a comparison of hypoxic area time series against previous coupled models.

Results
L432: use biogeochemical instead of biochemical throughout the manuscript
Response: We did not use the term biogeochemical when discussing only the biological or chemical processes.

L433: denitrification process should not consume oxygen
Response: The "4 Result and Discussion" has been updated. The denitrification does not consume oxygen (Eq. (R3)). In our model, SOC was estimated considering the aerobic mineralization, nitrification, and denitrification in the sediment as a 1-step process with the linear assumption applied.

L453-454: what does it mean by saying contributions are limited? I suggest showing the contribution in percentage. What is DO balance and how it was calculated?
L450-457: the entire description and calculation is misleading and confusing. Generally, all DO budget terms including physical terms, photosynthesis, SOC and WCR should be calculated. The summary of budget terms should match the change of DO. I think the authors did not understand and explain the oxygen dynamics well. Please refer to the model studies with oxygen budget analysis and rewrite this part.
Response: This part has been removed and updated. In section "4.1 Factors controlling subregion bottom DO variability", we compared the contributions of bottom DO advection (horizontal and vertical), bottom DO diffusion ((horizontal and vertical), SOC, and water column biogeochemistry on the bottom DO variability. Our results indicated that the variability of bottom DO on the LaTex shelf was mostly controlled by four processes: horizontal advection, vertical advection, vertical diffusion, and SOC (Fig. 9b in the updated manuscript). Their impacts on bottom DO variability varied from in different subregions throughout a year.

L455-456: does the biochemical process in this sentence represent water column respiration?
Response: We have removed this part. But in section "4.1 Factors controlling subregion bottom DO variability", we calculated the changes in bottom DO due to water column biogeochemistry compiled processes of phytoplankton photosynthesis,

phytoplankton respiration, zooplankton metabolism, aerobic decomposition of PON and DON, and nitrification.

L475-476: please indicate the change of PEA quantitively (e.g. in percentage).
Response: We have removed this part. Instead, the influences of water stratification on bottom DO variability were addressed in section "4.2 Stratification and Bottom DO Advection/Diffusion" by investigating the relationships between PEA and bottom DO transports. We found strong linear correlations between PEA and the advection terms suggesting that increased water stability in summer lead to less DO exchanges from advection processes. Nevertheless, the relationship between PEA and vertical diffusion of DO across the bottom layer appears to be non-linear.

L480-482: west-Mississippi nearshore did not show a change of current direction from
westward to southward, rather it pointed to northward.
Response: We have removed this part.

L498-499: please justify the choice of GBMs method.
Response: We have removed this part.

L498-511: Move detailed description of GBMs into method section.
Response: We have removed this part.

Figure 13(a) and Figure 10(a) conflicted in PEA contribution in nearshore West Mississippi?
Response: We have removed this part.

L540: what does this statement mean? Please clarify it.
Response: We have removed this part.

L543-544: how does it compare to other model studies? Is this parameterization better or not? Please add a more in-depth discussion here.
Response: Concerning the possibly improper sinking velocity applied in the model, we added two sensitivity experiments with different sinking velocities: 1 m/day and 5 m/day. A comparison of SOC against the observation by McCarthy et al., (2013) (Fig. RC2-1) suggests that the experiment with a sinking velocity of 5 m/day provides the best estimates (Fig. RC2-2). The root-mean-squared errors (RMSEs) are 567 μmol m^{-2} h^{-1}, 713 μmol m^{-2} h^{-1}, and 452 μmol m^{-2} h^{-1} for sensitivity tests with a sinking velocity of 15 m/day, 1 m/day, and 5 m/day, respectively. The model results (tests with sinking velocity = 5 m/day) generally overestimate the SOC at site F5 and C6 except for January 2009 and May 2010 at site C6, but underestimate SOC at site B7 and MRM. However, the simulated and observed SOC are generally in the same order of magnitude. Times series also reveals that the magnitude of simulated SOC simulated with a sinking velocity of 5 m/day is generally within the measured range (Fig. RC2-3) over the entire year. The magnitude of simulated SOC by tests with a sinking velocity of 15 m/day is out of the upper measured bound especially in summers. Modeled SOC by the test with a sinking velocity of 1 m/day is always below the lower measured bound.

We further compared the simulated ratio of SOC/overlaying water respiration against measurements (Fig. RC2-4). The test run with sinking velocity = 5 m/day provides the best-simulated ratio with a low averaged RMSE of 4.23 over site F5, C6, and B7 compared with 4.58 (sinking velocity = 15 m/day) and 6.51 (sinking velocity = 1 m/day) derived by the other two tests. At site MRM, both the two tests with faster (5m and 15m/day) sinking velocity highly overestimate the ratio in August 2009. We ascribe such bias to the relative course bathymetry near the river mouths. Point sources are applied in the model for diverting momentum and concentration tracers from the river to the computational grid cells. The scheme can lead to an overshot of river water at the near-mouth grid cells, which, may further result in shorter residence time of organic matter and plankton.

We also compare the hypoxic area simulated by the three different experiments. The simulation with 5m and 15m/day settling velocity show a similar hypoxia zone, while the using 1m/day is generally greater than the former two (Fig. RC2-5). Both the former two estimations (5m and 15m/day) can reproduce the magnitude and interannual variability of the measured hypoxic area. Compared to the shelf-wide observations, the simulated bottom hypoxic extent (Fig. RC2-6– RC2-8) by the experiment with a sinking velocity of 5m/day produces less bias among the three experiments. Based on these results, we changed the sinking velocity of PON from 15 m day^{-1} to 5 m day^{-1} in the baseline simulations.

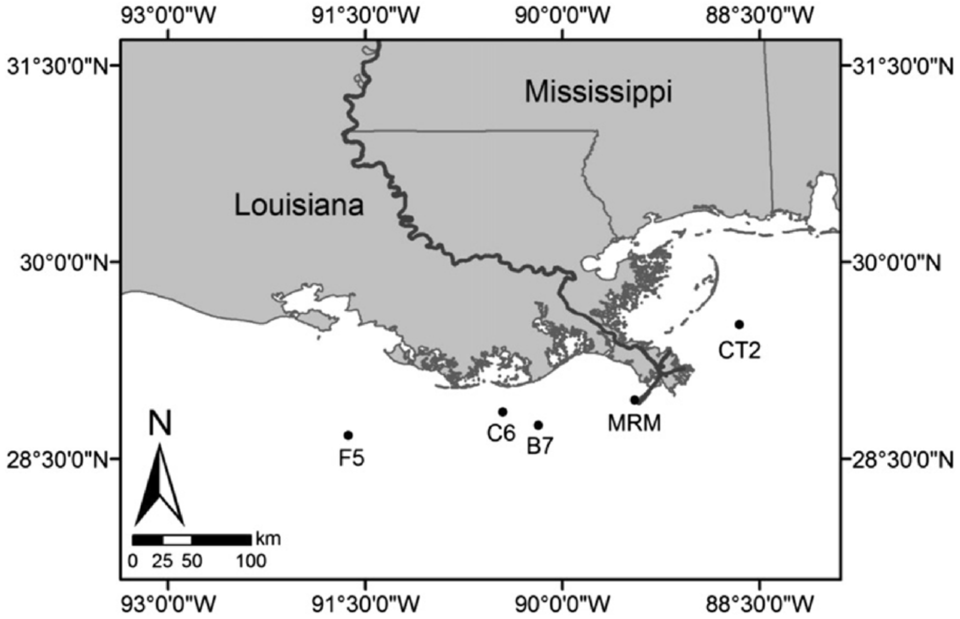

Fig. RC2-1 Map showing the location of sampling site in the northern Gulf of Mexico (McCarthy et al., 2013).

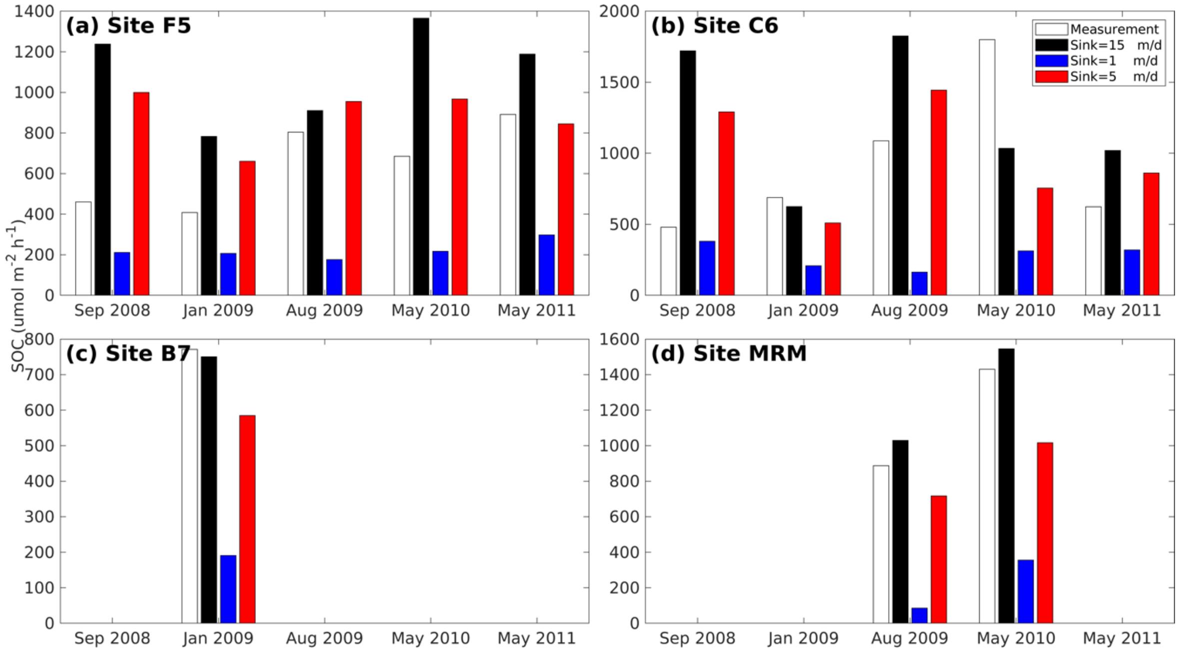

Fig. RC2-2 Comparison of observed SOC (in μmol m^{-2} h^{-1}) by McCarthy et al., (2013) and simulated SOC by different sensitivity tests.

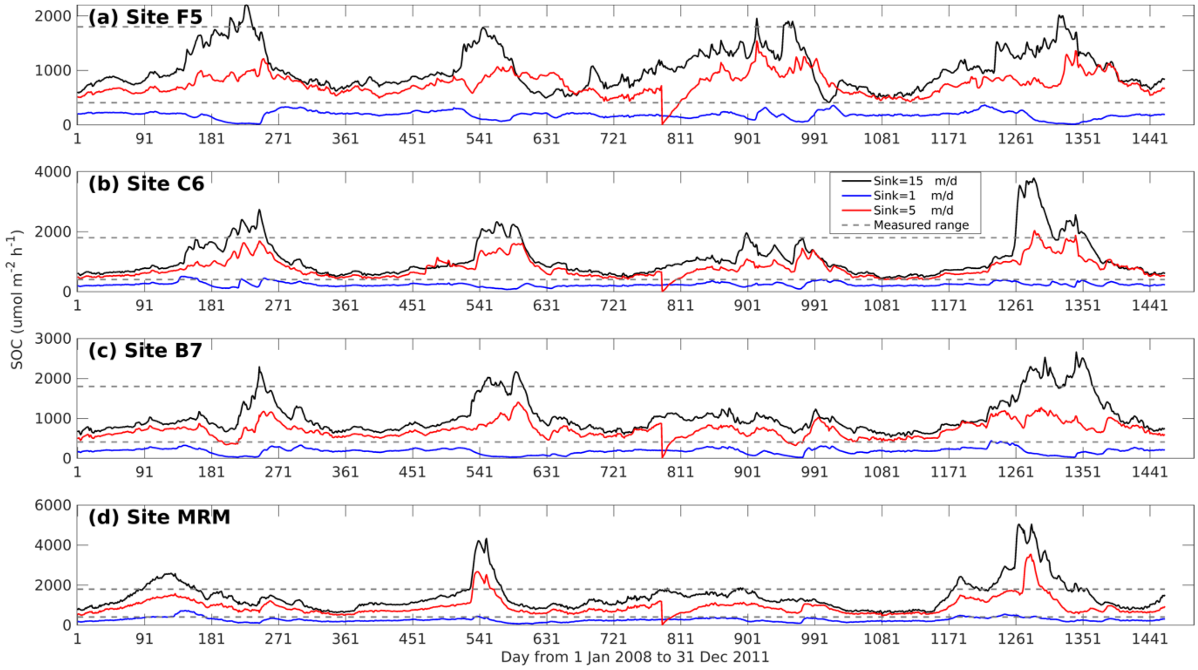

Fig. RC2-3 Daily average of simulated SOC by different sensitivity tests.

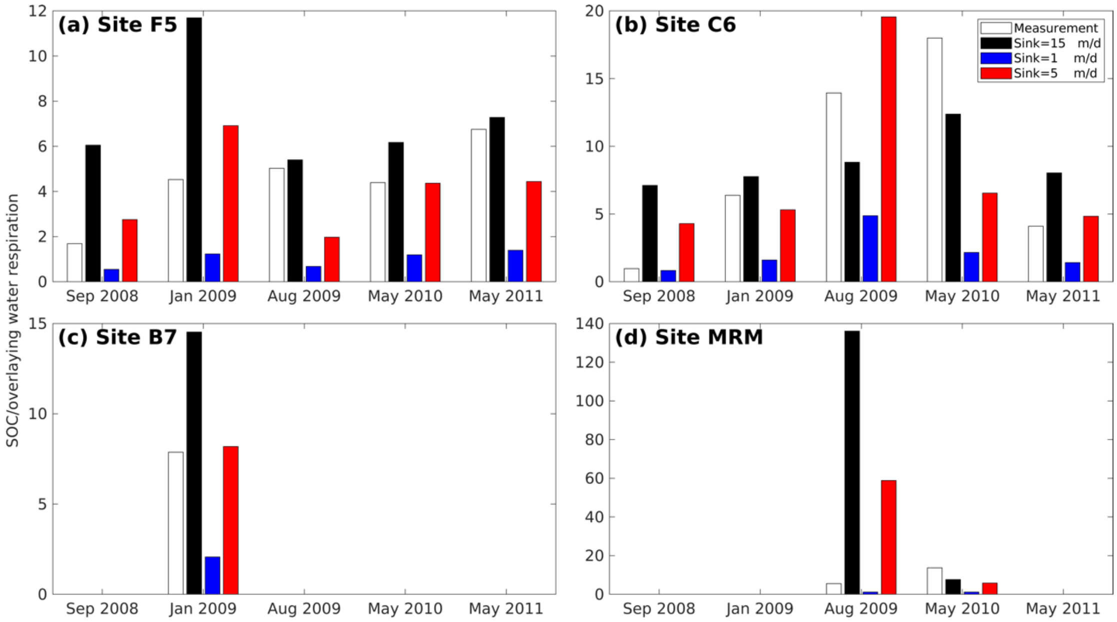

Fig. RC2-4 Comparison of observed ratio of SOC/overlaying water by McCarthy et al., (2013) and simulated ratio by different sensitivity tests.

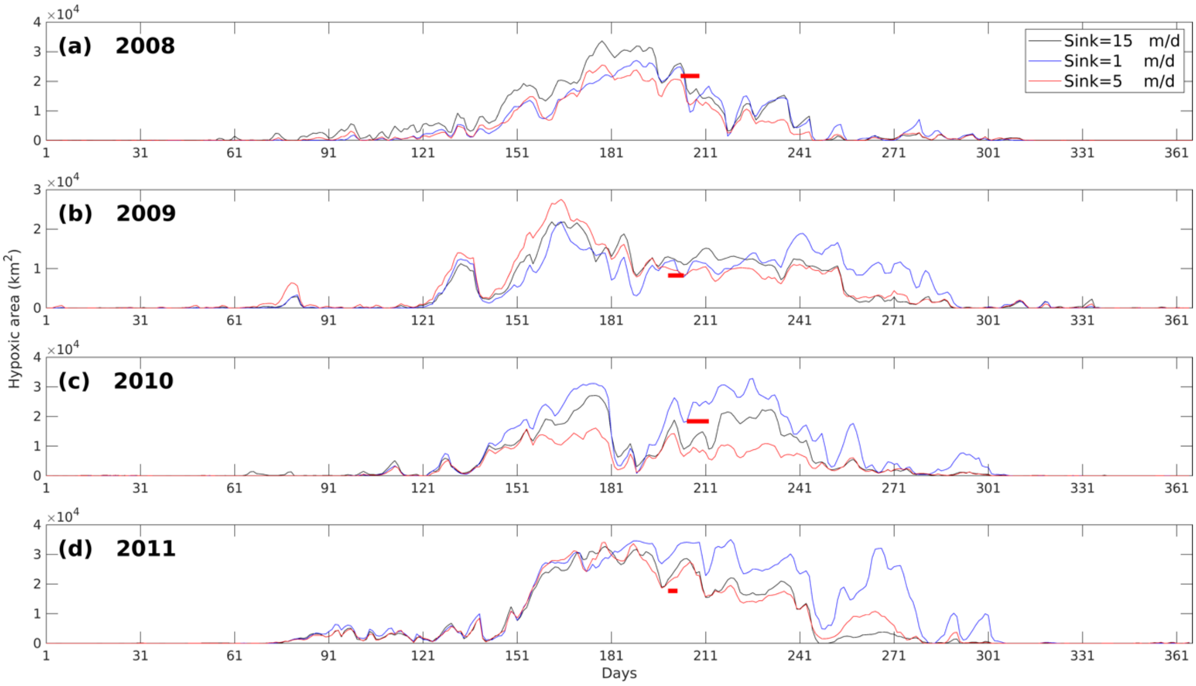

Fig. RC2-5 Comparison of observed and simulated hypoxic area. Note that the horizontal red thick bars denote the shelf-wide cruise measurements.

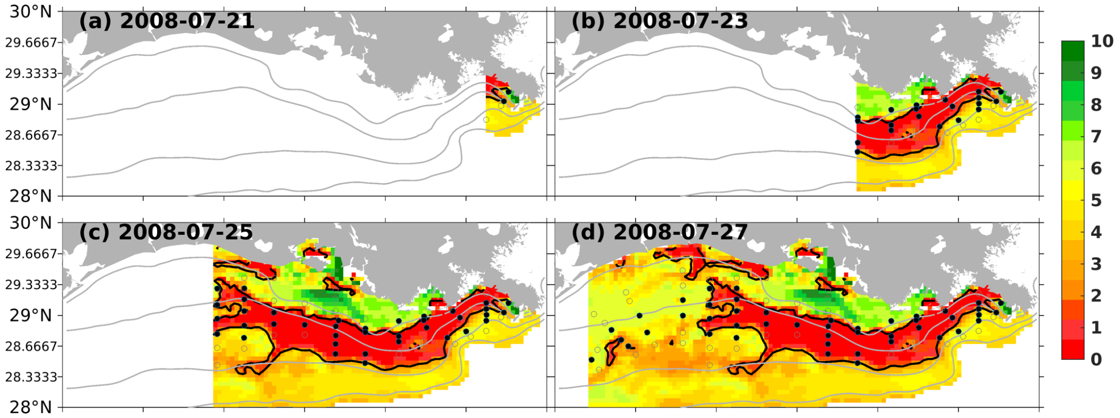

Fig. RC2-6 Evolution of simulated bottom water dissolved oxygen concentration (unit mg l^{-1}) by the sensitivity experiment with a sinking velocity of 15 m day^{-1}. The black filled circles and open circles indicate the hypoxic site and non-hypoxic site, respectively, according to the Shelf-wide cruise observations. The grey curves denote bathymetry of 5, 10, 20, and 50 m.

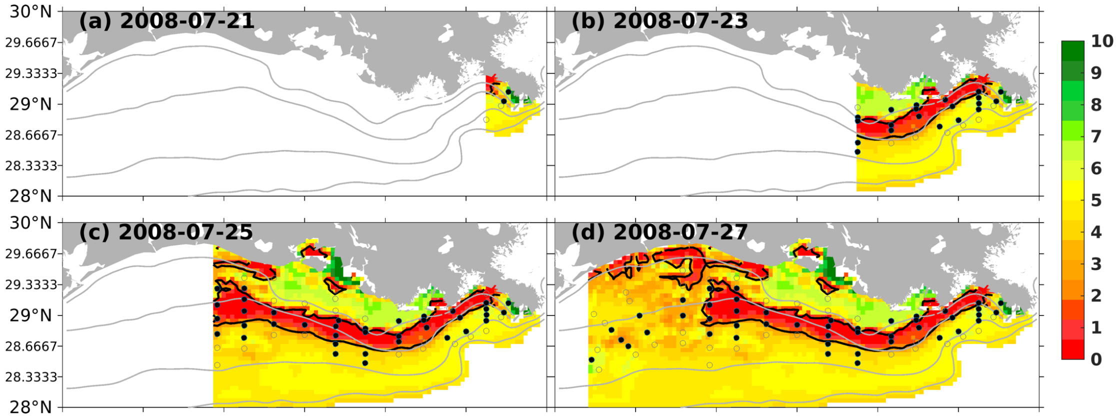

Fig. RC2-7 Same as Fig. RC2-6 but for the sensitivity experiment with a sinking velocity of 1 m day^{-1}.

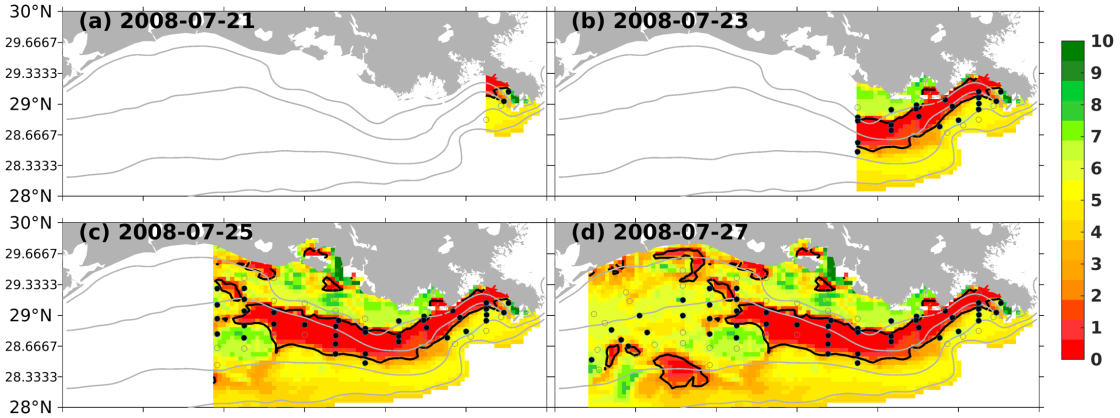

Fig. RC2-8 Same as Fig. RC2-6 but for the sensitivity experiment with a sinking velocity of 5 m day^{-1}.

L544-548: Figure 10 and Figure 15 looks very similar which is questionable to me. The previous studies suggested that sediment oxygen consumption dominated the hypoxia in the study area, while the water column respiration was still notable.

Response: We have removed this part.