Responses to Comments by Referee #1

We thank Referee#1 by the detailed reviews he/she provided. Please find our detailed response below (*referee comments in Italic*).

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: We acknowledge that new findings were not well presented in the first draft. For this study, the novelty of our model includes the incorporation, for the first time, of the silicate cycle in hypoxia simulation in the La-Tex shelf. Another novelty is the inclusion of multiple phytoplankton and zooplankton functional types. As the study region is dominated by the diatom community, we deem these two features of our model would provide new knowledge to the hypoxia research. We have been adding some sensitivity tests to further assess the importance of the silicate cycle in hypoxia development. Some previous results are discussed below and we will incorporate more relevant discussion during the revision stage.

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. The authors need to validate this part of the model (SOC versus water column respiration).

Response: To explore the sensitivity of sinking velocity in hypoxia development, we added two sensitivity tests with different sinking velocities, 1 m/day and 5 m/day, respectively. A most ideal selection of sinking velocity will be determined by the validation of SOC, the ratio of SOC and overlaying water respiration, bottom hypoxic area, and bottom hypoxic extent. Measured SOC and overlaying water respiration was reported by McCarthy et al., (2013), while the measured hypoxic area and extents are based on the Shelf-wide cruise. Following McCarthy et al., (2013), we extract the daily SOC, and overlaying water respiration at sites F5, C6, B7, and MRM (Fig. 1 below) and averaged the observations by months.

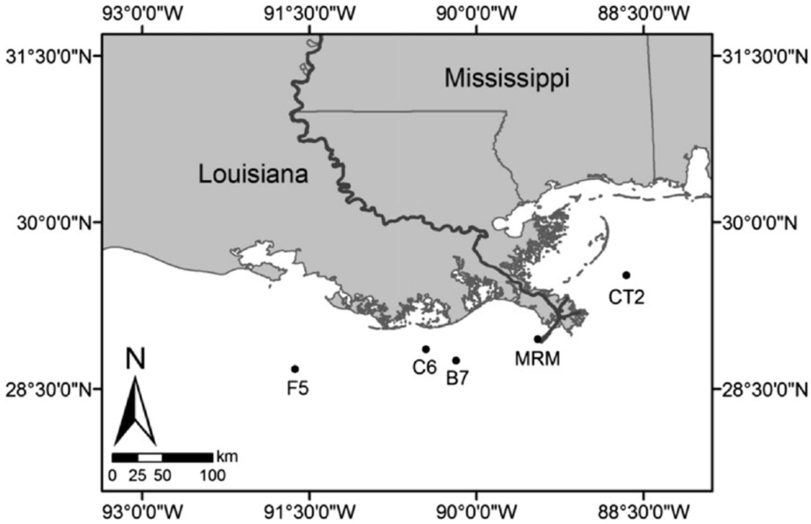

Fig. 1 Map showing the location of the sampling site in the northern Gulf of Mexico in McCarthy et al., 2013.

Fig.2 indicates that a sinking velocity of 5 m/day provides the best estimate of SOC. 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 (used in the first draft), 1 m/day, and 5 m/day, respectively. The simulated (5 m/day) and observed SOC are generally in the same order of magnitude. The model results in general overestimate the SOC at sites F5 and C6 except for January 2009 and May 2010 at site C6, and underestimate SOC at sites B7 and MRM. Times series also reveals that the magnitude of simulated SOC by tests with a sinking velocity of 5 m/day is generally within the measured range (Fig. 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 always yields a SOC below the measured ones.

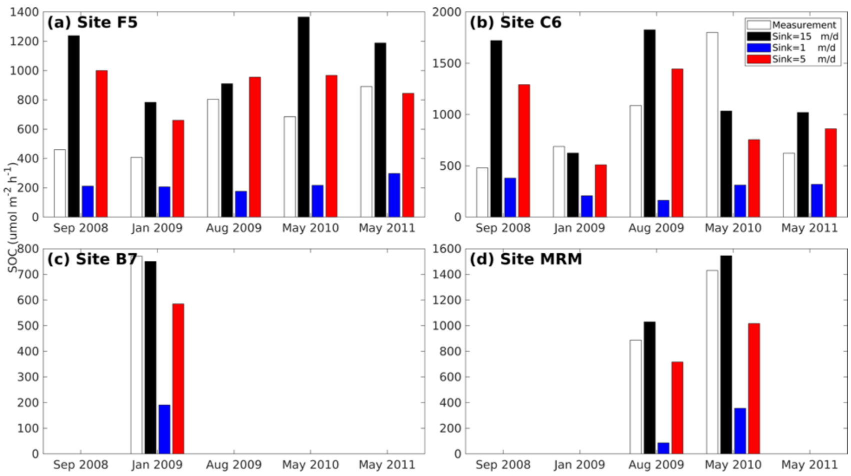

Fig. 2 Comparison of observed SOC (in μmol m^{-2} h^{-1}) by McCarthy et al., (2013) and simulated SOC by different sensitivity tests.

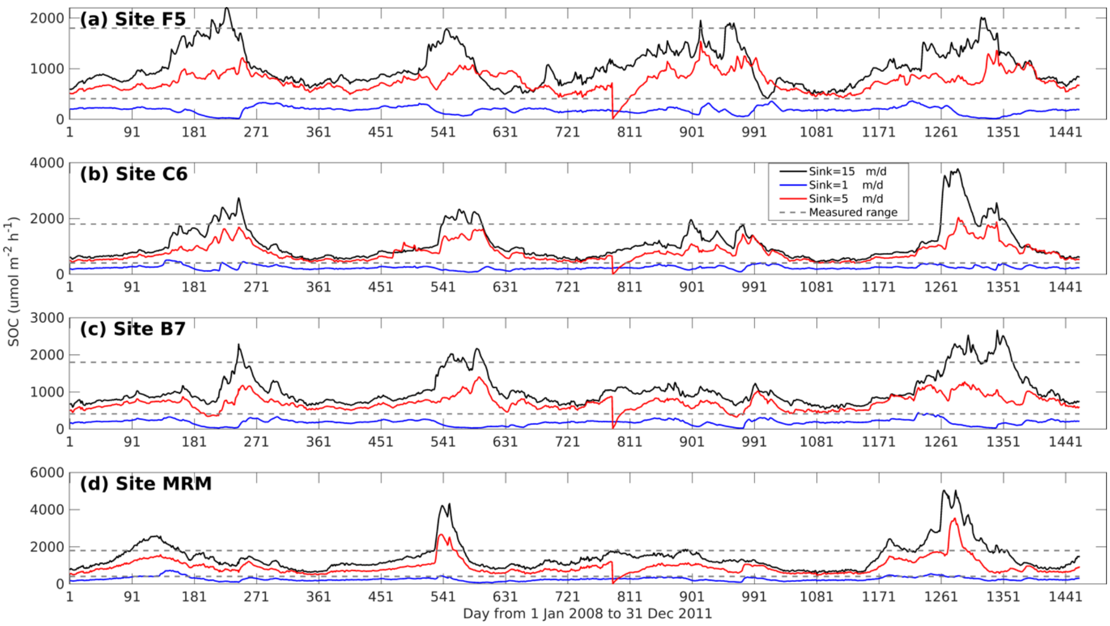

Fig. 3 Daily average of simulated SOC with different PON sinking velocities

We further compared the model-simulated ratio of SOC/overlaying water respiration against that based on available measurements (Fig. 4). The test run with a sinking velocity of 5 m/day shows most agreement with observations with a low averaged RMSE of 4.23 over site F5, C6, and B7, compared with an RMSE value of 4.58 generated by experiment using a sinking velocity of 15 m/day and a value of 6.51 by experiment using a sinking velocity of 1 m/day. At the site near the Mississippi river month (MRM), the two experiments with a sinking velocity of 5 and 15 m/day highly overestimate the ratio observed in August 2009. A possible reason for such bias is that point sources are applied in the model for diverting momentum and concentration tracers from the river to the rest of 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 a shorter residence time for organic matter and plankton.

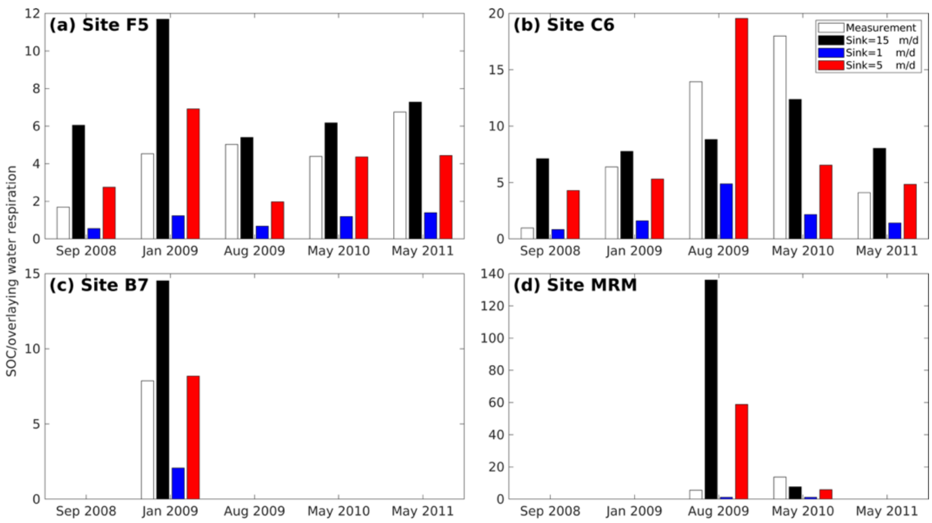

Fig. 4 Comparison of the observed ratio of SOC/overlaying water by McCarthy et al., (2013) and simulated ratio using different settling velocity

Following this comment, we changed the coverage of the model grids used for hypoxic area estimation (Fig. 5). Since the Shelf-wide cruise surveys did not reach the west of 95°W during most of the summers, and the surveys could reach the water with a depth of around 6 m near the Atchafalaya River mouth, we restricted the region from the west side of the Mississippi Delta to 95°W with a water depth ranging from 6 to 50 m. We then compared the model-estimated hypoxic area with different sinking velocities against the Shelf-wide cruise in Fig. 6. Estimations by the two tests with faster sinking velocity (5 and 15 m/day) are close to each other during the cruise periods, while the estimation by the other test (1 m/day) is generally greater than the former two. A sinking velocity of either 5 and 15 m/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. 7–9) by the test run with a 5 m/day sinking rate seems to produce less bias.

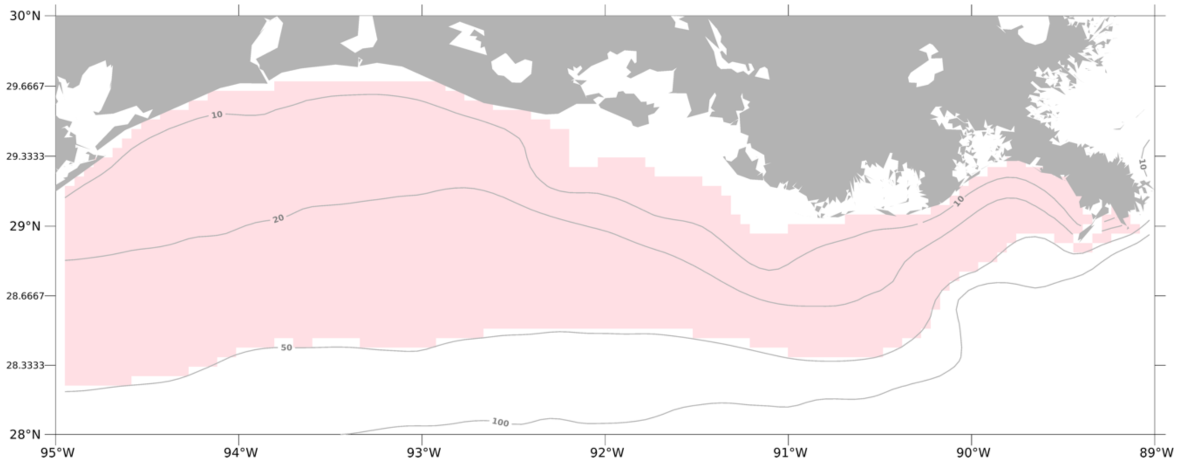

Fig. 5 A distribution of model grids used for hypoxic area estimation.

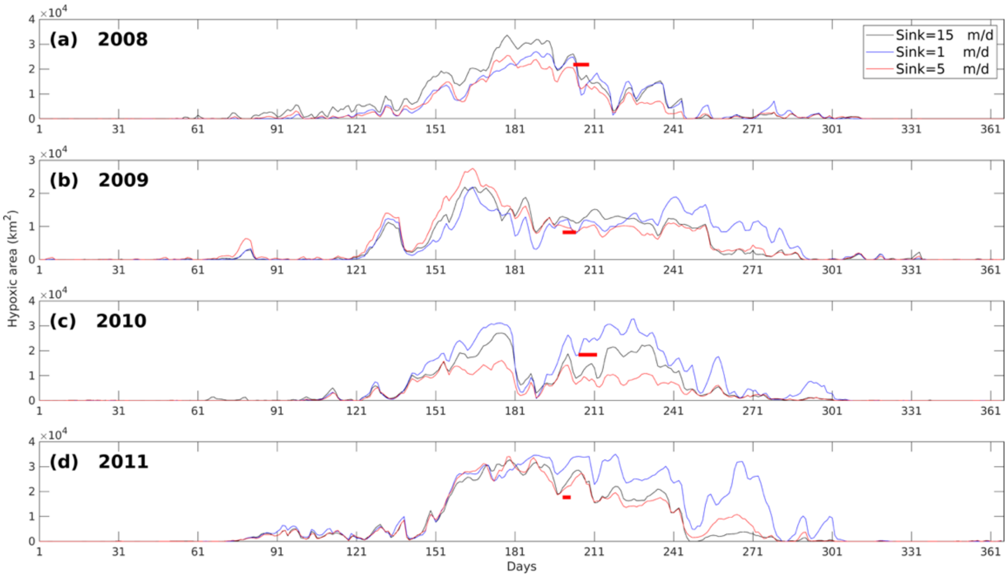

Fig. 6 Comparison of observed and simulated hypoxic area. Note that the horizontal red bars denote the magnitude and temporal coverage of the Shelf-wide cruise measurements.

According to the above comparisons, we will change the sinking velocity of PON from 15 m day^{-1} to 5 m day^{-1} in all experiments and will update the relevant results and discussion.

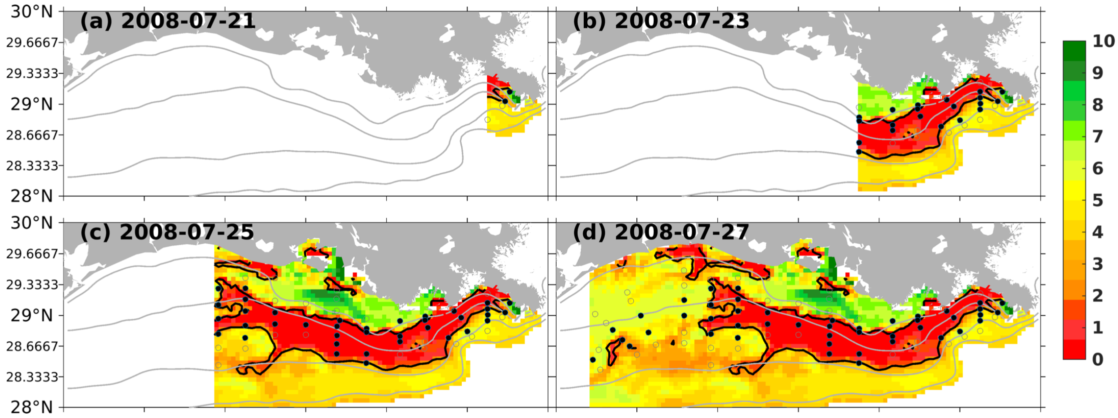

Fig. 7 Evolution of simulated bottom water dissolved oxygen concentration (unit mg l^{-1}) with a sinking velocity of 15 m day^{-1}. The black-filled circles and open circles indicate the hypoxic site and non-hypoxic site, respectively, according to the Shelf-wide cruise observations. The grey curves denote bathymetry of 5, 10, 20, and 50 m.

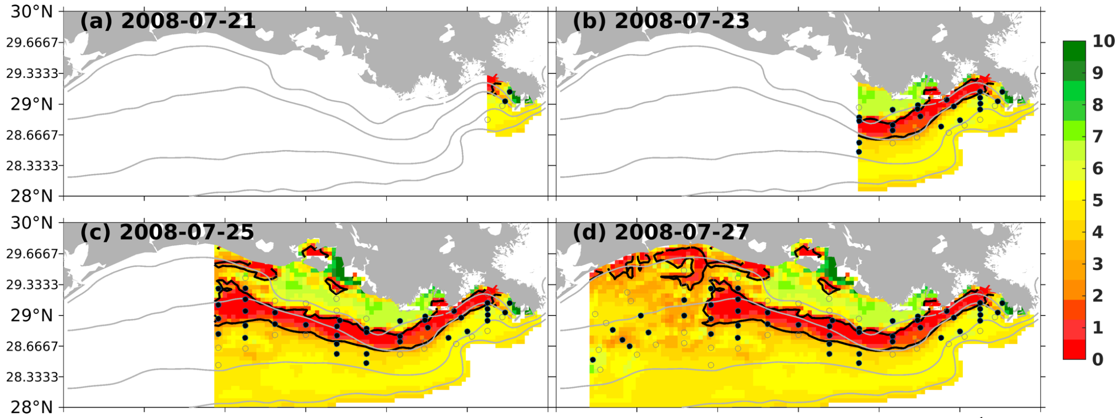

Fig. 8 Same as Fig. 7 but for the test run with a sinking velocity of 1 m day^{-1}.

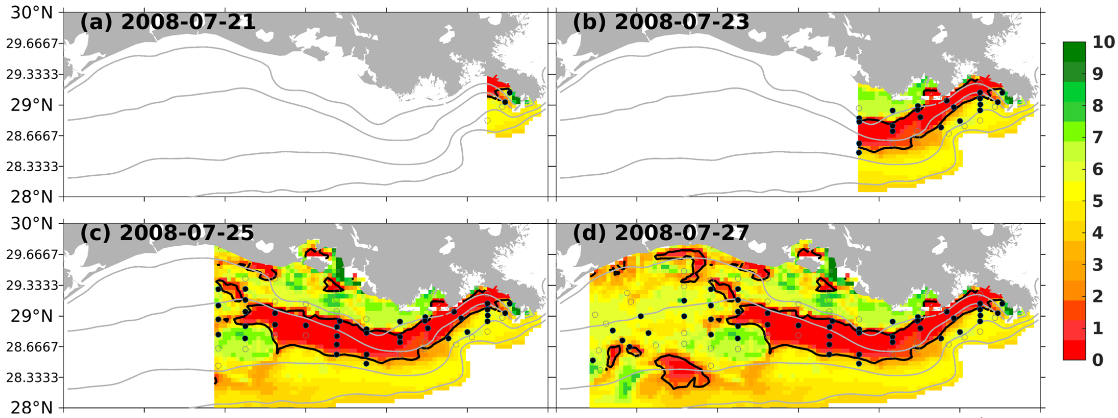

Fig. 9 Same as Fig. 7 but for the test run with a sinking velocity of 5 m day^{-1}.

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 fields 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 will carry out a series of sensitivity tests to make sure the model reached equilibrium before any analysis.

Spatially, the sedimental PON is the largest on the nearshore near the Atchafalaya River mouth as a result of the material delivered by the along-shore currents. The plume of sedimental PON exhibit a seasonal pattern that can be explained by the coastal current systems. In spring and early summer, as the dominated currents are westward, sedimental PON can be transported westward reaching 95.5°W. Such westward coastal current systems stimulate coastal downwelling, restricting the offshore transports of nutrient and sedimental PON as water column PON sinks. As the prevailing currents turn to be eastward in July and August, sedimental PON stretches southeastward from the Atchafalaya River mouths. Due to the consequent coastal upwellings, more offshore transports of sedimental PON can be expected. Temporally, the sedimental PON did not exhibit a continuously increasing pattern during the studied period, instead, it fluctuated between a peak and a trough throughout a year. We will provide a more relevant discussion in the revision.

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 for the given limited computational resources, which is the practice of most of the hypoxia modeling efforts in this region. One objective of this study is to achieve a long-term simulation to train a hypoxic area prediction model presented in Part II of the paper. Thus, although our research group published a sediment-induced light attenuation algorithm (see Zang et al., 2020), for model efficiency we did not consider the light attenuation induced by sediments. . For a 15-year hindcast, it cost 170 hours (~1 a week) using 500 CPU cores. 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 the hypoxia waters.

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 will provide more quantitative comparisons. The model's robustness in reproducing surface nutrient concentration will also be added.

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: We see this point is similar to the reviewer's first comments. We will add more discussion about silicate limitation's impact on hypoxia evolution and the benefits of incorporating a multiple phytoplankton and zooplankton group. In detail, we plan to discuss silicate's impacts on hypoxia in two aspects: 1) the contribution of large phytoplankton (Pl, diatom); and 2) riverine silicate inputs. We were able to present some new model results with different riverine silicate inputs. The control run is symbolized as exp0 with the same setups as that in the first submission but with an updated sinking velocity changed to 5 m day^{-1} based on the above discussion. We then changed the riverine SiOH4 concentration from 0.2 to 2.0 with an increment of 0.2 in each experiment (exp1 through exp 9). For exp10 (ongoing), river silicate inputs were the same as that of the exp0, however, we removed the silicate limitation on large phytoplankton growth. More detail on these new experiment tests is listed in Table 1.

Table 1. Model setups of different sensitivity tests. Simulation of exp10 has not yet finished and thus updates will focus on results by testing exp0 through exp9.

	Simulation	Sinking velocity (m day^{-1})	Scale of Riverine SiOH4 concentration	Silicate limitation on Pl	# phytoplankton group	# zooplankton group
exp0 (control)	1 Aug 2017–25 Aug 2020	5	1	Yes	2	3
exp1	1 Aug 2017–25 Aug 2020	5	0.2	Yes	2	3
exp2	1 Aug 2017–25 Aug 2020	5	0.4	Yes	2	3
exp3	1 Aug 2017–25 Aug 2020	5	0.6	Yes	2	3
exp4	1 Aug 2017–25 Aug 2020	5	0.8	Yes	2	3
exp5	1 Aug 2017–25 Aug 2020	5	1.2	Yes	2	3
exp6	1 Aug 2017–25 Aug 2020	5	1.4	Yes	2	3
exp7	1 Aug 2017–25 Aug 2020	5	1.6	Yes	2	3
exp8	1 Aug 2017–25 Aug 2020	5	1.8	Yes	2	3
exp9	1 Aug 2017–25 Aug 2020	5	2.0	Yes	2	3
exp10	1 Aug 2017–25 Aug 2020	5	1.0	No	2	3

The hypoxic area is estimated as the sum of the area of model grids defined in Fig.5 when the bottom DO is less than 2 mg l^{-1}. Percentage changes (Fig. 10) of the May–September hypoxic area are calculated between the sensitivity run (i.e., exp1–9) and the control run (exp0). The results of exp1 through exp9 indicate that the hypoxia area is mostly positively correlated with riverine silicate inputs. It is worth noting that the impact on the hypoxic area due to the changing riverine loads is not linear. The hypoxic area is more sensitive to elevated silicate supply especially when riverine input increases by more than 1.2. The average percentage increment of the hypoxic area ranges from ~35% to ~72% as the riverine silicate supply increase by 40% to 80%. In contrast, the hypoxic area decreases by ~19% to ~53% as the silicate supply decreases by 40% to 80%. When the changes in riverine silicate supply are less than 20%, the changes in the hypoxic area would be expected to be greater when riverine silicate is reduced (-19%) than increased (+12%).

We found three points that are new to coastal managers. Firstly, decreases in riverine silicate loads by 20% and 40% do not lead to significant differences in terms of hypoxic area reduction (by ~19% for both cases). Secondly, it is hard to meet the Gulf Hypoxia action plan goal of a 5,000 km^2 hypoxic area by reducing the riverine silicate loads solely. The average summer hypoxic area from 2018 to 2020 is 15000 km^2 which is comparable to that of 1985–2010 (14000 km^2). Thus, to meet the Gulf Hypoxia action plan goal, the average percentage reduction of the hypoxic area should be ~67%. More discussion of the combined silicate and nitrogen reduction is needed and will be provided in the revision. Thirdly, as the range of 25th–75th percentile of hypoxic area changes enlarges as the riverine silicate load increases, an elevated riverine silicate input is likely to introduce much worse hypoxia events.

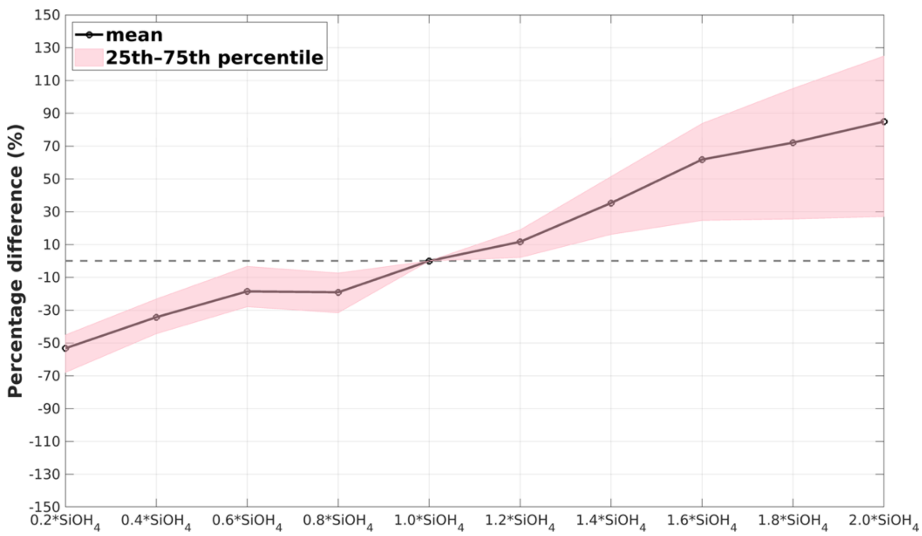

Fig. 10 Percentage differences of the simulated hypoxic area between sensitivity tests (exp1–9) and the control run (exp0). Statistics are based on the simulations in May–September from 2008 to 2020.

The differences in the spatial DO distribution between the riverine inputs sensitivity test (i.e., exp1–9) and the control run in August (2018–2020 average) are shown in Fig. 11. When riverine silicate is reduced, the low slope west shelf is more sensitive to the changing silicate supply than the east shelf. In the cross-shelf direction, bottom DO between 10–50 m isobaths is more sensitive to the reduction of riverine silicate inputs than the rest regions. A slight decrease (by 20%; Fig. 11d) of silicate supply would lead to a maximum bottom DO increase of 2 mg O_2 l^{-1} in this region. A 20% decrease in silicate supply can therefore easily induce a change from hypoxic to normoxic bottom waters in such regions. When riverine silicate inputs are increased, not much difference in the spatial distribution of DO reduction between the west and east part of the La-Tex shelf until the increase is more than 80%—then DO drops more in the west part of the shelf than the east part.

The above results show the impacts of riverine silicate loads on the bottom hypoxic area and bottom hypoxic water extent. We found that 1) the distribution of the hypoxic waters is more sensitive to the elevated riverine silicate loads with greater uncertainties than the reduced inputs; 2) a dual or triple silicate reduction is needed to meet the goal of the Gulf Hypoxia action plan; 3) the responses of bottom DO concentration is not spatially homogeneous along the shelf when riverine silicate loads are adjusted, and 4) the west shelf will suffer more from hypoxia conditions when riverine silicate is increased by more than 60%. We will provide further analysis and discussion on this topic in the revision including a recommendation of combined nitrogen and silicate reduction to meet the goal of the Gulf Action Plan.

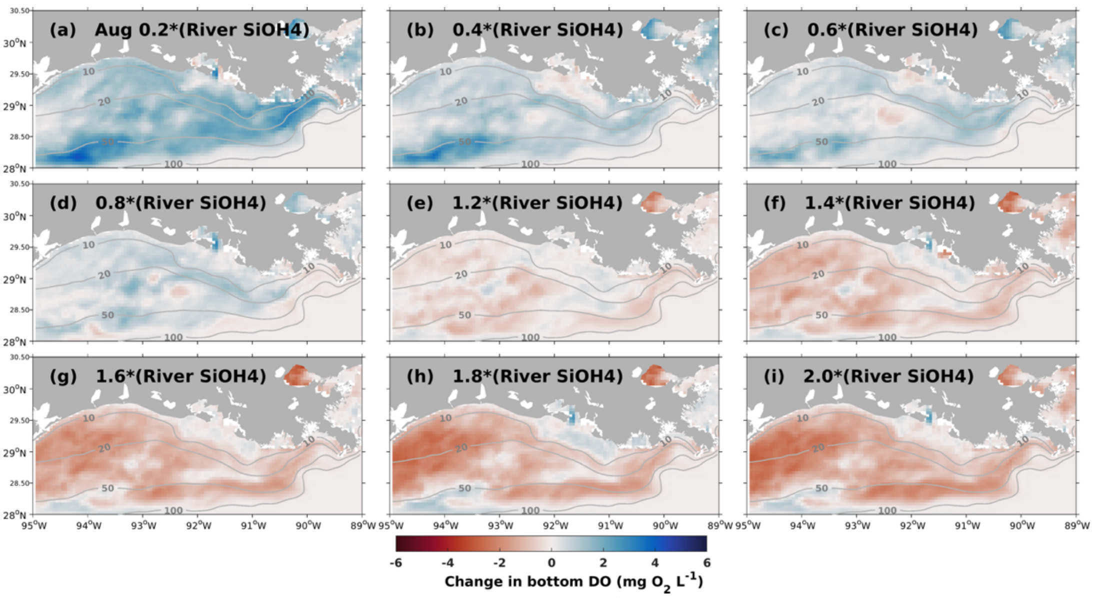

Fig. 11 Differences in bottom DO (in mg O_2 l^{-1}) between experiments with different riverine inputs and the control run (August mean of 2018–2020).

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: Arguments will be provided to address the objective of this study. The arguments will be centered on the importance of silicate limitation on hypoxia development in the LaTex Shelf.

L33-34: this is true only in a dual reduction strategy

Response: Yes. L30-34 already addressed it.

"Sensitivity experiments of hypoxia area reduction to different 30 nutrient shrinking strategies by Fennel and Laurent (2018) suggested that to meet the hypoxic area reduction goal (reduce to < 5,000 km2 in a 5-year running average) set by the Hypoxia Task Force (2008), a dual nutrient strategy with a reduction of 48 % of total nitrogen and inorganic phosphorus would be the most effective way. Although nitrogen is the ultimate limiting nutrient, phosphorus load reduction would also lead to a significant shrinkage of the hypoxia (Fennel and Laurent, 2018)."

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 will revise this part. Justic and Wang (2014) used both deposited OM and DO for the estimate of SOC. We will revise our statement.

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: We mean "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: We want to address that we did not couple a sediment model for consideration of model efficiency. The SOC scheme we adopted is highly based on Fennel et al., (2006, 2011) but with an additional sediment PON term to correct the misestimations by the instantaneous remineralization parameterization. Indeed, our SOC scheme is somewhat similar to that of Justic and Wang (2014). But, in their model, silicate is not included.

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: Yes, we agree that more is not always better. Yet here we demonstrated the importance of riverine silicate loads to the hypoxia development. Below, we present the diatom validation to further address the significance of the inclusion of the diatom function group.

We updated our baseline simulations with the sinking velocity of 5 m day^{-1}. The simulated ratio of diatom concentration (i.e., large phytoplankton or Pl in the manuscript) over total phytoplankton concentration is therefore updated. We compared our simulated ratio (Table 2) to the only two, to our best knowledge, available in-field studies (Schaeffer et al. , 2012; Chakraborty and Lohrenz, 2015; see Figs 12 and 13 for smapling location). 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. Thus, the measurements by Schaeffer et al. (2012) may not be a good reference for validation. However, we can still make the comparison with an assumption that the densities of all phytoplankton groups are similar.

The simulated diatom percentages (from 67% to 90%) are within the range of measurement (from ~40% to ~99%) by Schaeffer et al. (2012) (Table 2) during February, April, May, and June 2008. Our simulations also well reproduced the Chakraborty and Lohrenz's (2015) measurements except for the summer and fall of 2009. In the inner shelf area, modeled (observed) diatom percentages are 60% (65%), 57% (63%), and 56% (50%) in January 2009, April 2009, and March 2010, respectively. In the mid shelf, overestimations are found in January 2009 (61% vs 48%) and April 2009 (61% vs 30%) while a slight underestimation is found in March 2010 (57% vs 64%). Averaged diatom percentages over the inner and mid shelf are close in January 2009 (61% vs 57%), April 2009 (59% vs 46%), and March 2010 (57% vs 57%) between the simulations and observations. The simulated percentages overestimate the observation in both summer and fall of 2009 over both inner and mid shelf. Nevertheless, the measurements were carried out by only two cruises conducted in summer and fall 2009, respectively

The high contribution of diatom group to the shelf phytoplankton community emphasizes that the inclusion of diatom group in numerical model is critical understand phytoplankton dynamics and the associated hypoxia events in the LaTex Shelf.

Table 2 Comparison of simulated diatom percentage of the total phytoplankton. Note that the simulated percentages are conducted based on concentration values and averaged over the corresponding months. The measured percentages by Schaeffer et al. (2012) (denoted as superscript *) are calculated based on biovolume values. The measured percentages by Chakraborty and Lohrenz (2015) (denoted as superscript $^\wedge$) are derived by chlorophyll *a* attributed to different phytoplankton groups.

	Diatom/total phytoplankton × 100%		
	Inner shelf	Mid shelf	Inner and mid shelf average
February 2008	88 (~65*– ~99*)	59	
April 2008	95 (~55* – ~85*)	62	
May 2008	90 (~40* – ~99*)	63	
June 2008	67 (~80* – ~99*)	61	
January 2009	60 (65$^\wedge$)	61 (48$^\wedge$)	61 (57$^\wedge$)
April 2009	57 (63$^\wedge$)	61 (30$^\wedge$)	59 (46$^\wedge$)
July 2009	68 (30$^\wedge$)	51 (13$^\wedge$)	60 (22$^\wedge$)
October–November 2009	63 (46$^\wedge$)	51 (19$^\wedge$)	57 (33$^\wedge$)
March 2010	56 (50$^\wedge$)	57 (64$^\wedge$)	57 (57$^\wedge$)

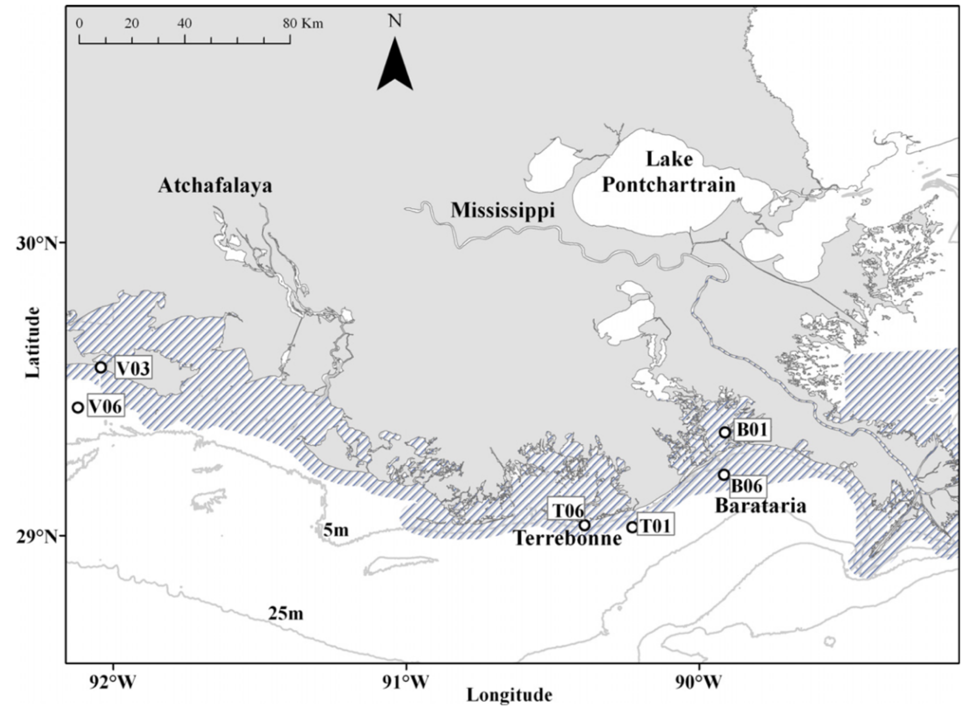

Fig. 12 Sampling sites (denoted by white circles) in Schaeffer et al. (2012).

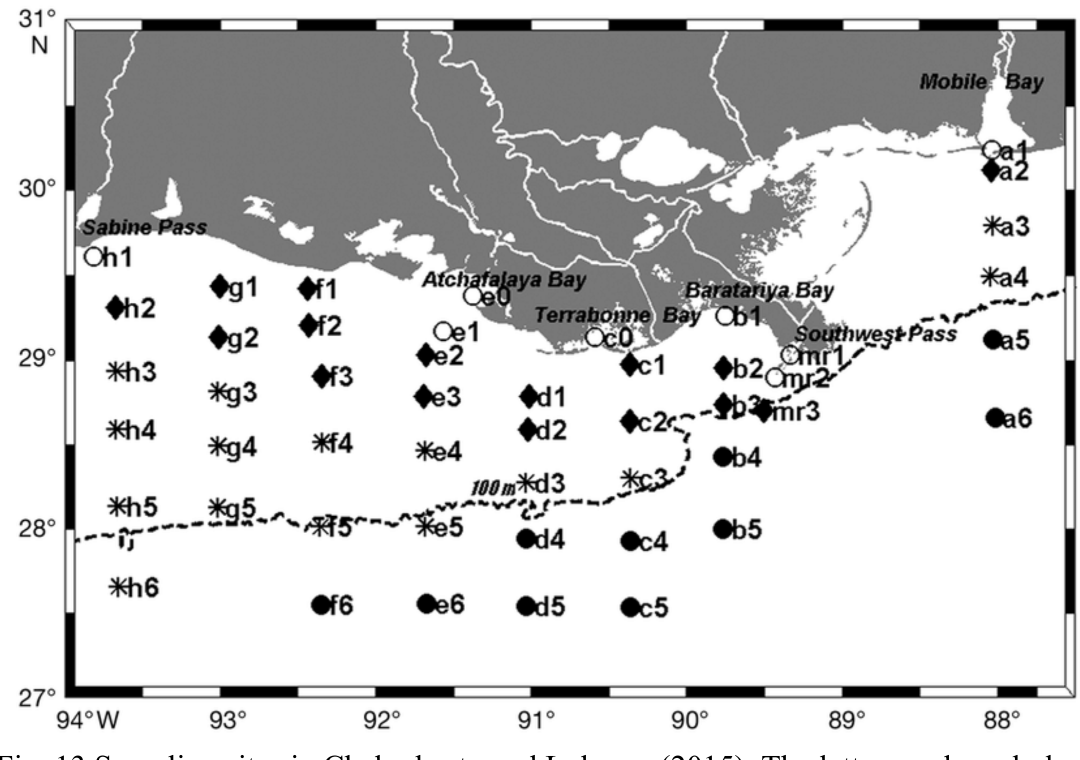

Fig. 13 Sampling sites in Chakraborty and Lohrenz (2015). The letters and symbols denote the geographical locations of stations demarcating the different water masses found in the area: estuaries (open circle), inner shelf (filled diamond), mid-shelf (asterisk), and offshore/slope (filled circle) waters.

References:
Schaeffer, B. A., Kurtz, J. C., & Hein, M. K. (2012). Phytoplankton community composition in nearshore coastal waters of Louisiana. *Marine pollution bulletin*, *64*(8), 1705-1712.

Chakraborty, S., & Lohrenz, S. E. (2015). Phytoplankton community structure in the river-influenced continental margin of the northern Gulf of Mexico. *Marine Ecology Progress Series*, *521*, 31-47.

L79-80: there are lots of discussions about the factors controlling bottom O2 in the papers you cited above.
Response: Yes, there are lots of discussions on the dominated factors of bottom DO. But we may want to compare the spatial difference of these factors. We may need to revise this sentence to make it more clear to readers.

L85: you did not discuss silicate above
Response: Will add a discussion of silicate in the revision.

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 the hypoxic area in the LaTex Shelf. The predicted hypoxic area showed a high agreement with the ROMS hindcast time series (RMSE=4,571 km^2, R^2=0.8178) even when the prediction model was applied to the HYCOM global dataset. (The prediction model was trained using the daily output by the coupled model, and then was applied to the HYCOM dataset for prediction.) We will post the results of the accompanying paper here.

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 sediment transport model with a consideration of model efficiency. For a long-term hindcast purpose, we simplify the sedimental processes just like many previous researchers did. We want to demonstrate the advances of a simplified treatment of sediment model in a long-term hypoxia hindcast/forecast.

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 will add more discussion on the inclusion of both phosphorus and silica in the introduction part.

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: We will provide more information on the riverine nutrient time series including download links and supporting figures and will also update the plot of the computational domain with the positions of the selected river gages plotted.

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. On the other hand, 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 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 will clarify this paragraph.

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 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 the reality. Thus, we prescribe 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 will restate the calculation in the revision. 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 all 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 only a portion of 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.

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. But we assumed that the rate of oxygen consumption due to sedimental biochemical processes was identical at the bottom most water layer. The M here is the production rate of NH_4 (released to the overlaying water) due to aerobic decomposition occurs in the sediment. Thus, M is in mmol m^{-3} day^{-1}. As we did not couple any sediment models, the calculation of SOC was a kind of simplification even we considered a sedimental PON pool which was an imaginary 2-D pool rather than a 3-D sediment pool.

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 oxygen level is low enough, however, the replenishment of DO in water column is usually faster than in the sediment. Thus, although water column could reach below hypoxia condition during some period, the fast replenishment would weaken the dentification in 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: A time series of PONsed will be provided. The 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. To further assess the model stability from 2007, we will add more evidence including time series of water level, water temperature, water salinity, PONsed, NO_3, PO_4, $Si(OH)_4$, two phytoplankton functional groups, three zooplankton functional groups, and bottom DO in different regions of the shelf.

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} on the open boundary toward 1/60 $day^{-}1$ for the interior (61% of computation grid cells).

L240: can you also show the other rivers for completeness?
Response: We will plot out the locations of all river point sources applied in the model.

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 lower reach of the main channel usually few hundred kilometers from the mouths. Along to the lowermost of the river channel, river discharges shall be more 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 in 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 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. The 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 watershed.

L250: it is indeed highly oversaturated. can you provide some context?
Response: The oxygen level in river water should be slightly oversaturated. We will perform another experiment with a reduced riverine oxygen supply to see the response of shelf water DO dynamics.

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 Shelfwide observation prior to 2012 according to the link provided and will expand the validation accordingly.

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
Response: We will provide comparison of surface data in different seasons.

L301: Same comment here, I don't think this is a proper way to validate the model. Also, what about chlorophyll?
Response: We will provide comparison of different profiles in different seasons. And we will also add validation for chlorophyll a.

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 will revise this part accordingly.

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 lack of discussion of this issue in our manuscript. High PO_4 concentration is found both in the simulated and observation-based WOD profiles. We did not include sediment module in our model, therefore, the high PO_4 concentration near bottom may come from recycled DOP, POP, and sedimentorganic matter (measured as PONsed). The resuspension may be important as the high PO_4 concentration is found extending from near bottom to subsurface layers in some observed profiles with the maximum depth around 10 m. 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: We will complete our validation accordingly.

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%. We will revise this part for a more detailed and clear comparison.

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).

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: Indeed, observed bottom DO at a depth around 6 m is available. We will show the model results of other regions with the observed values overlayed in the revsion.

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: Such phenomenon occurs in summers with massive offshore extent of bottom hypoxic waters. We will examine the local water column stratification, local consumption of oxygen, and offshore transport of bottom waters to clarify this issue.

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: We will re-restrict our studied area (like as shown in Fig. 5 in this document) according to the observations covering more inshore waters but less western coastal waters.

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 depth from 10 to 50 m). Thus, as we may need to re-restrict the studied area according to the observations, the bias may change. However, even when the simulated hypoxic area is reconstructed, we still cannot provide comparison for days without observations. We showed the daily time series of hypoxic area here is to emphasize that the 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 (more westward) and from the tracks, intensity, and translation speeds of hurricanes. In mid-July 2019, hurricane Barry stroke the 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 the early July.

Figure 6: 1) Another way to make this comparison would be to overlay the observations as scatter points over the model maps
Response: We will try it.

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 in Figure 6 for comparison purposes was not a ~1-week average, but a ~1 week composite according to the cruise locations on different days (see the hypoxia evolution in Fig. 7–9 in this document). For example, for measurements conducted on July 22 in waters east of 90W, modeled bottom DO on July 22 in waters east of 90W is extracted. For measurements conducted on July 23 in the region between 92W to 90W, modeled bottom DO during this period over this region is extracted. The map of modeled bottom DO is the composite of these segments. We will elaborate on it in the revision.

3) can you show the other years for completeness?
Response: We will add 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 will split out the Result and Discussion and provide model comparisons between our results and previous studies in the Discussion part.

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: The analysis for long-term trend is to dig out characteristics of variations of bottom DO concentration. Long-term trends, multi-yearly mean, standard deviation, and season cycle shall be the general characteristics in geoscience studies. But this analysis may need to be simplified as we will focus more on the silicate limitation rather than general patterns. PONsed did not accumulate near the Atchafalaya nearshore regions. We will provide more details to elaborate it (like time series). We will also provide a more comprehensive decomposition of bottom DO changes considering local rate of changes, advection, and diffusion to address the DO balance in different regions.

L380-390: can you compare these patterns with the literature?
Response: We will carry out more model comparisons.

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 were struggling in finding time series of bottom DO concentration with the LaTex Shelf. However, previous studies showed that bottom DO was usually below the hypoxia threshold at the C6 station (e.g., Justic and Wang, 2014).

L450: also vertical diffusion and possibly horizontal advection, as well as SOC
Response: We will provide a more comprehensive decomposition of bottom DO changes considering local rate of changes, advection, and diffusion to address the DO balance in different regions.

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 will complete the recommended comparisons.

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: The PEA maps may not be useful here since PEA is not only affected by river freshwater but also the adjacent deeper waters. We will revise this part accordingly.

L498-511: this paragraph should go in the Methods section
Response: We will move it to the Methods section.

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: The high settling rate of organic matter indeed affect the results. We decrease the sinking velocity to 5 m day^{-1} and provide a reasonable range of SOC and SOC/overlaying water column respiration (Fig 2–4 in this document).

L517: yes, this is where you find persistent hypoxia
Response: We will provide more figures for PONsed including time series and spatial pattern for further discussion.

L522: where is this shown? you speculate here
Response: This is our speculation since according to the statistical analysis, SOC is more dominated in nearshore regions. We will add some more sensitivity tests to further elaborate it.

L524: +10% would be a more conservative value and used for climate projections in the region, e.g. Lehrter et al 2017.
Response: We will do more literature studies on this issue.

L525: you speculate here
Response: Yes. We will add more sensitivity tests here and move it to the discussion 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 tests results, we have changed the sinking velocity to 5 m day^{-1} as we have shown in this document.

L543-555: I don't get the point of this paragraph
Figure 15: I don't get the point of this figure
Response: We aimed to provide discussion on the relationship between hypoxic volume and hypoxic area and the hypoxia thickness (1–2 m). We agreed that this part may out of the focus of this study and we will revise 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 in the Gulf of Mexico biogeochemical modeling has already been illustrated by Shropshire et al. (2020). 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 discussed and deemed to be important to the shelf hypoxia by Fennel and Laurent (2018). Instead of focusing on the P limitation, we have and will add more sensitivity tests to address the Si limitation on hypoxia dynamics.

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 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 will validate the surface nutrients and surface chlorophyll *a* concentration to assess the bias our model could lead to.

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 having the same sensitivity to light.

L650: L650: did you mention how these parameters were chosen? were they calibrated to the Gulf of Mexico?
Response: We will provide the sources of these parameters.

Minor comments/typos:
Response: We will correct all the typos and inappropriate usage of words according to the comments.

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 shelfL56: "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"