Revision Letter

Dear Editor,

We are very grateful to the two anonymous reviewers for their constructive comments and corrections. We revised our manuscript following their suggestions to improve the quality of our study. Below is the summary of our responses to the comments (in italic) from the two reviewers. Specifically, we include more details about our model setup and re-designed the sensitivity tests to better reveal the role of sediment in underwater light attenuation. The vertical structures of water density and nutrient are also discussed. We also did another round of spelling and grammar checks.

We hope our responses adequately address the concerns from the reviewers and the revised manuscript is acceptable for publication in Biogeosciences.

Best Regards,

Z. George Xue & Zhengchen Zang, on behalf of all co-authors

Response to Reviewer 1

1. *The experiment design is not flawless. It is not explained why a 20year run was necessary to achieve a stable initial condition, which itself is not shown or validated against data at all. Then, the model starts a few days before the hurricane makes landfall, but it would have made more sense to start a week earlier to show what a "normal" state of the ecosystem in that time period would have looked like. Also, to show whether the model actually recovers to said "normal" state, the model should have been run a week longer after it actually ends. While this would have taken more time, a 20y run to spin up the ecosystem was strictly speaking not necessary. 2-5 years are common in literature, and setting this up is literally an issue of an hour, if the forcing is already available.*

   The 20-yr simulation is originally designed for another paper discussing the long-term variation of biogeochemical cycling in the nGoM. We believe that it is reasonable to extract its result as the initial condition for our hurricane simulation since all the variables are stable in the long-term run. To validate the initial condition, we add one more panel in Fig. 3 and compare with satellite result (details see our response to Question 7).

   Starting the model one week earlier is very challenging in hurricane simulation because the low-pressure vortex of hurricane is not formed yet. We tried to extend our simulation one week longer, but the water did not recover to normal condition because another hurricane Ike approached the nGoM, and vertical mixing turned to be strong again before stratification fully recovered to normal condition.

2. *In line 72, parameterisation approaches to sediment specific attenuation are discussed. One might mention the strong correlation between sediment content and bathymetry, as employed or discussed e.g. in Zhou et al. 2017 (https://doi.org/10.1016/j.pocean.2017.10.008) and Thewes et al. 2020 (https://doi.org/10.3389/fmars.2019.00816). Their approach has the disadvantage of not being time variant, making a stronger case for online modelling of sediment.*

   The two studies use a simplified scheme $\frac{i}{N}$ (i represents the vertical level and N is the total number of vertical layer) to parametrize sediment-induced light attenuation. We cite the two new papers in our introduction (see lines 75-76).

3. *Section 2.1 This section heavily relies on previous publications, which is fine. However, this way, a few bits of information fall under the radar, like how many different sediment types there are. Also, it is possible and not too difficult to program a specific coefficient for each sediment type. An explanation as to why this was not done is needed.*

   We add more information about sediment setup in section 2.1 (Lines 114-119).

   Applying a specific coefficient for each sediment type is not difficult. However, that cannot help us better understand the role of sediment in the underwater optic environment because: 1) sediment light attenuation coefficient is not only determined by size (see details in section 2.2), and we don't have any measurements about light attenuation coefficients for different types of sediment, so the selection of attenuation coefficients might be too arbitrary; 2) resolving the fraction of different types of sediment in the water column during the hurricane is still a challenge due to the complexity of sediment dynamics (flocculation and deflocculation; hindered settling effect, sea floor consolidation and liquefaction, etc.). In addition, sediment fraction is not calibrated in our sediment model paper (Zang et al., 2018) because no observation is available. Thus, using different coefficient for each type of sediment might potentially introduce more uncertainties but little useful information. Here we still stick to our original scheme, which uses one constant coefficient for all sediments. Of course, this coefficient only represents the "mean" value of attenuation coefficient for all types of sediment.

4. *In line 110, it says the model was "largely built on" NEMURO. Other than the addition of a sediment specific attenuation term, has anything else been changed?*

   The original NEMURO model does not have chlorophyll, which is not in favor of model validation. We include chlorophyll in our model following Fennel et al. (2006), and details are described in lines 133-136.

5. *Because passages in later sections are on the subject of hypoxia, NEMURO might not have been an ideal choice. Perhaps it should be explained why NEMURO was chosen over models that have been coupled to ROMS and published before, which calculate oxygen as a state variable.*

   We use NEMURO because it includes diatom, which accounts for 20-60% of total phytoplankton biomass (Zhao and Quigg, 2014) in the northern Gulf of Mexico (details see lines 120-121). Since simulating hypoxia is not the objective of this study, without oxygen dynamics in NEMURO does not harm to our conclusion.

6. *There is a very important error in the equation in line 132. Because plankton biomass and sediment content are functions of depth, the equation must include integrals of PSn + PLn and SSC over depth. Section 2.2 The authors should explain why they selected the specific values of a_sed that they chose. Are the studies to which they refer conducted in the same region? Are they at all comparable to the model situation?*

   We update the equation in section 2.1 (line 150).

   In the benchmark run, $\alpha_{sed} = 0.059$ is based on the study in the Delaware Bay (McSweeney et al., 2017) because no studies provide a specific light attenuation coefficient of sediment in the nGoM. Several studies estimated α_{sed} in the nGoM but only gave a wide range of α_{sed} from 10^{-3} to 10^{0}, which cannot be applied to our numerical study because we need a specific number rather than a range. Here we re-design the sensitivity tests following reviewer's suggestion in Question 17.

7. *Section 3 The authors have conducted a 20y run to obtain an initial condition, yet the initial condition is never shown or validated against data. Because the 20y run is not the object of the study, it needs no validation, but the initial condition certainly does. This could be an extra panel in figure 3.*

We add an extra panel in Fig. 3 showing the initial condition (chlorophyll) of our model. The spatial distribution and value of chlorophyll are quite similar to the pre-hurricane satellite image (Fig. 3b) although the area of simulated plume around the Mississippi bird-foot delta is larger (details see lines 179-184).

8. *RMSE and R were computed. For this to be done, one of the data sources would have to be regridded to match the other. Was satellite data or model data regridded?*

   We interpolate satellite data to model grid to make the comparison.

9. *Figure 4 shows logarithmic values for SeaWiFS and model data. Were RMSE and R calculated using the actual data or the logarithmic data?*

   Both RMSE and R are calculated using the logarithmic data.

10. *In line 174 it says that the "model's performance was significantly improved in high productivity waters where chlorophyll concentration is >1mg/m3". Does this refer to satellite or model data?*

    This refers to both model and observed data (see line 199).

11. *Section 4.1 NPP is not defined in the text. Although there are literature definitions, this should be explained.*

    We define net primary production (NPP) at the very beginning of section 4.1 (lines 208-209).

12. *Why is there a 3-4hour delay in chlorophyll with respect to NPP? Are biomass and chlorophyll uncoupled? Do different species have different C:CHL ratios?*

    The peak of NPP suggests the maximum growth rate of chlorophyll, not the peak of chlorophyll. Chlorophyll concentration reached its maximum when NPP went down and was compensated by other loss terms (e.g., grazing and mortality). In this study, we estimate chlorophyll concentration following Fennel et al. (2006), and the variation of chlorophyll due to the growth of small and large phytoplankton are simulated individually, so the C:Chl ratio of different types of phytoplankton are not the same in the model (details see section 2.1 lines 134-136).

13. *In line 197, it says that surface cooling and decreased light contributed to reductions of chlorophyll and NPP. Can you identify the individual contributions? From fig. 5b, it seems that surface cooling limits by about a factor of 0.5 and light limits by up to 0.3 at the maximum. This should be calculable just by putting in representative values in the respective equations for light and temperature limitation. Also, it says reduced temperature and light availability "contributed" to the reductions. What else might have contributed? The first peak in figure 5a is significantly broader than most of the following peaks. After that, they seem to have an almost bimodal quality. This is not addressed in the text. Is this due to the different species? It is true for both runs. The NPP in test 1 recovers almost immediately to the same peak value, albeit narrower. Chlorophyll does not really recover to pre-storm values at all, but the benchmark run does. Again, is this due to speciation?*

In the model, the growth of phytoplankton (Gpp) is determined by temperature ($T_{limit}=exp(k_{GPP}*temp)$), light ($L_{limit}=[1-exp(-\alpha*I/V_{max})]*exp(-\beta*I/V_{max})$), and nutrient ($N_{limit}=NO_3/(NO_3+K_{NO3})exp(-\Psi NH_4)+NH_4/(NH_4+K_{NH4})$):

$$\text{Gpp}=\text{Gpp}_{\text{max}}*\ T_{limit}*\ L_{limit}*\ N_{limit}$$

It is noteworthy that temperature-limit factor (T_{limit}) is different from light- and nutrient-limit factors (L_{limit} and N_{limit}): T_{limit} is much larger than 1, while L_{limit} and N_{limit} are between 0 and 1. Thus, T_{limit}, L_{limit} and N_{limit} are not comparable quantitatively. However, T_{limit} declines with lower temperature, and L_{limit} decreases with lower light availability (I), so it is safe to conclude that reduced temperature and light contribute to the lower primary production.

Although NEMURO model is able to resolve phytoplankton size structure, we have no field data to convince the readers that our model is able to well reproduce the fraction of different types of phytoplankton. Therefore, we decide not to analyze small or large phytoplankton separately in this study.

14. *I suggest to separate contributions of the individual plankton species to both NPP and CHL. It looks to me as though one of the two phytoplankton species is more susceptible to light or temperature limitation than the other. If the run had been longer by a week in both directions, one might have seen a full recovery to a "usual" state after the hurricane (i.e. the broad peak in NPP on the 30th of August). It looks like the benchmark run recovers faster to that broad, supposedly "normal" peak, while test 1 shows a quicker recovery in the leading peak. It may, as stated in the manuscript, be largely due to the boost in NO3. However, temperature and light sensitivity might be different for the two phytoplankton species. This should be disclosed.*

    As mentioned above (Question 13), we are very prudent about simulated phytoplankton size structure due to the lack of observation, so detailed analysis regarding different types of phytoplankton is not included in our updated manuscript.

15. *NPP is a depth integrated quantity, but only surface chlorophyll is shown. There is no info on lower layer productivity. What is the vertical structure of phytoplankton, or rather, does it change when switching on sediment specific attenuation? Section 4.2 Figure 6 is very illustrative. It might be helpful to have two more panels showing NO3. In that case, perhaps rotate orientation to columns <-> time and rows <-> parameter.*

    We include NO3 profiles of the benchmark run and test 1 in Fig. 6. The orientation of this figure is also rotated following reviewer's suggestion.

16. *In the second paragraph of this section, the authors discuss hypoxia. However, NEMURO does not provide oxygen output. Then, the situation in the nGoM is compared to the Delaware estuary. It would be prudent to show that these areas are at all comparable, which they are likely not, because an estuary is usually bounded horizontally and is characterised by strong lateral salinity gradients. The Delaware estuary experiences hypoxia due to density stratification. There is no figure representing temperature or salinity stratification in the nGoM model, but only chlorophyll*

and SSC. The latter strongly influences the former and SSC stratification is perhaps purely due to sediment settling. What is implied in the text is that the chlorophyll stratification is due to a density stratification, and that phytoplankton does not reach the lower layers, because it is physically bound. There needs to be a figure showing temperature and salinity, or density and stability frequency, to be able to imply a similar situation as in the Delaware estuary. In line 237 of the text it says that "post-hurricane stratification recovery" prevented oxygen ventilation to the bottom. Is this with reference to the model in the study? Please show stratification along the transect. Perhaps show some in situ data of oxygen from that time period, to show that there actually was a hypoxia event after Gustav. Otherwise, consider removing the paragraph.

We add water density anomaly profile in Fig. 7 to illustrate the variation of vertical structure: water column went through a cycle of stratification (strong)-well mixed-stratification (weak) in 11 days. In our sediment paper (Zang et al. 2018), the analysis of Brunt Väisälä frequency (BVF) also demonstrate the recovery of stratification after hurricane. In addition, both previous field measurements (McCarthy et al., 2013) and model results (Moriarty et al., 2018) suggest that water stratification after strong resuspension event is favorable to the formation of hypoxia, so we still prefer to keep this paragraph in the updated manuscript.

17. *Section 4.4 Although it is clear from a modeler's perspective and from a methodical point of view why the authors decide to do a sensitivity study with regards to sediment specific attenuation, the explanation that this is to accommodate for the way different types of sediment attenuate light is vague. It is true that different types attenuate differently, but then it appears more reasonable to simply compute the model with one coefficient for each sediment type. The programming effort is a day's work at best. The researching effort to get plausible values might be a little more work, but all in all it is not clear why this has not been done. It might be worth an explanation or at least an elaboration as to what prevented the authors from doing that. Again, it is a perfectly reasonable approach to perform a sensitivity study over a varying singular attenuation coefficient, yet the explanation lacks context. Why were the specific values chosen? They were taken from referenced studies, but are the sediments in these studies comparable in their make up (see comments to section 2)? Why not linearly vary around a reasonable value by 20% (this is almost the case anyway)? Why not do 5 tests instead of 3?*

    We re-designed four sensitivity tests (tests 2-5) by increasing (decreasing) attenuation coefficient by 20 and 40%, respectively (lines 169-171).

18. *Figure 7 hints at the chaotic nature of the ecosystem by showing how a small change in an initial state can alter the following development. Even though after the 6th of September, SSC at the surface was almost zero, the benchmark and test 2 deviate more from each other after that day than they do before. This is perhaps an interesting point to make.*

    The difference between the benchmark run and sensitivity tests is mainly caused by different light attenuation coefficients. Our discussion in section 4.4 explains why our model is more sensitive to α_{sed} when SSC is in a certain range (lines 302-312): The contribution of high SSC overwhelms that of α_{sed} to the variation of sediment-induced light attenuation term ($\alpha_{sed}SSC$), while light attenuation term is very small when SSC is almost zero. Therefore, the primary production and nutrient dynamics are only sensitive to α_{sed} when SSC is moderately high, and that explains why we conclude there is a range for SSC in which α_{sed} plays a vital role in photosynthesis (see our response to Question 20).

19. *Ideally, for all runs with sediment attenuation, the light limitation should be identical over time. However, given the chaotic nature of turbulence, an elaboration as to whether or not the individual run's SSC deviate from each other in any way is missing. It is not expected to deviate heavily, but a note on whether or not there is deviation is appropriate.*

    We add one sentence at the end of section 2.2 "sensitivity tests" to elaborate that the derivation due to turbulence was not considered in this study. Details see lines 172-173.

20. *In line 290 it says that the study does not provide a widely accepted SSC range to determine whether a_sed plays a vital role in photosynthesis and primary production [...]. This is somewhat confusing, as a_sed clearly plays a vital role. If it is 0, the model output looks drastically different in some regards. I am not sure what is meant by a "widely accepted SSC range". Does that mean that a certain value of a_sed may be valid for a certain range of SSC? If so, please rephrase the sentence to make this clearer.*

    As shown in the sensitivity tests, primary production was not sensitive to α_{sed} when SSC is too low (pre-hurricane) or too high (hurricane landing), which means there is a range of SSC in which the ecosystem is sensitive to α_{sed}. Due to the differences (hydrodynamics, phytoplankton species, nutrient sources, optic feature of sediment particles, etc.) between coastal ecosystems, this range might vary greatly. Therefore, we state that "this study does not provide a widely accepted SSC range in which α_{sed} plays a vital role in photosynthesis and primary production". Here we rephrase the sentence to make it clearer (lines 313-314).

21. *The authors further speculate in line 296ff that there might be variation on possibly annual to decadal scales. To emphasise the validity of the claim, as well as the importance of the subject, the following studies on the North Sea might be of interest and inspiration: Dupont & Aksnes 2013 (https://doi.org/10.1016/j.ecss.2013.08.010), Capuzzo et al. 2015 (https://doi.org/10.1111/gcb.12854) and Wilson & Heath 2019 (https://doi.org/10.5194/os-15-1615-2019). It might also be noteworthy that hurricane events are predicted to occur more frequently under the changing climate. Implications of this could be lowered primary production.*

    These studies confirmed the importance of long-term sediment dynamics in marine ecosystem and biogeochemical cycling, so we included these papers in our manuscript to support our speculation (see lines 332).

    The frequency of hurricane events under the changing climate in the future is still controversial because different prediction models show opposite results: the numerical study of Bender et al. (2010) found the overall frequency of Atlantic hurricanes decrease due to global warming, while Knutson and Tuleya (2004) suggested the increase frequency of hurricane due to CO2-induced warming. Thus, we decide not to discuss the influence of hurricane frequency on primary production in this study.

22. *Section 4.5 The authors are right in addressing the lack of CDOM representation, and they correctly point out that many models share the same weakness. However, as they also point out, salinity is often used as a proxy for CDOM via an inverse relationship, which is well documented and easy to implement. It does make the paper more complex, but it might have been worth the effort, specifically when doing a sensitivity analysis with respect to when an ecosystem shifts from a nutrient limited to a light limited regime. The CDOM contribution would undoubtedly have biased the results of the sensitivity test. It is therefore appropriate to include a sentence or two in the discussion, explaining that CDOM specific attenuation would likely shift a hypothetical threshold a_sed downwards, below which the regime may be considered nutrient limited.*

    Although the Mississippi-Atchafalaya River is one of the major sources of CDOM to the nGoM due to its high discharge, previous field measurements found that coastal wetlands provide important CDOM contributions to nearshore shelf waters, and the relationship between CDOM concentration and salinity off the wetlands was totally different from that in the river plume (Chen and Gardner, 2004; Shank and Evans, 2011). Moreover, our model does not include freshwater discharge from wetlands. Therefore, using an inverse relationship between salinity and CDOM to estimate its light attenuation is not applicable in our study region. In our model uncertainty discussion (section 4.5 lines 330-337), we add the potential impacts of neglecting CDOM light attenuation on our model results following reviewer's suggestion.

23. *Section 5 The authors say in the first sentence of this section that they "introduced a sediment-induced light attenuation algorithm to the coupled physical-biogeochemical model on the platform of ROMS", which is, strictly speaking, not true. They did introduce it to NEMURO, however, ROMS still uses a different light attenuation scheme. See e.g. Mobley et al. 2015 (https://doi.org/10.1002/2014JC010588) or Cahill et al. 2008 (https://doi.org/10.1029/2008GL033595) to find what the not insignificant difference is. There is no coupling between sediments and light absorption in the physical model.*

    We change the sentence to "We introduced a sediment-induced light attenuation algorithm to ROMS' biogeochemical model" (line 382). In addition, we cite these two papers at the end of Section 4 and highlight that our model does not modify water heating due to light absoprtion in the hydrodynamic model (see line 378).

24. *In line 367, the authors again claim that the post hurricane situation might have caused a hypoxia. Without any data, this is purely speculative and does not belong in the conclusions chapter.*

    We removed the statement about hypoxia in the conclusion chapter. Details see line 403.

25. *Minor comments see attached documents for grammar errors and other minor comments.*

    We corrected the grammar errors and updated our manuscript following the reviewer's comments.

Response to Reviewer 2

1. *The authors reference the models (papers) which their model is based on (in some cases the same or very similar model), including a hydrodynamic, sediment transport and biogeochemical model. They need to give a short description of the models in the "Methods section", especially the sediment transport and biogeochemical models.*

   We add more details about sediment transport and biogeochemical models in section "Model Setup". Details see lines and lines 110-131.

2. *Authors should mention why they did not include nutrient river loadings (and show values), boundary conditions (and show values) and provide values of initial conditions. Values can be averages, ranges etc.*

   Our model include fluvial nutrient discharge (data was extracted from USGS Water Data website; details see lines 127-129). The concentrations of biogeochemical tracers along the open boundaries are zero due to the lack of observations and the limited impact on our study area. We add a new panel (a) in Fig. 3 to show the initial condition of chlorophyll concentration.

3. *I am no expert with satellite data, but my understanding is that SeaWiFS is no longer in use? Regardless the authors need to provide information (reference) about the algorithms used to calculated satellite-based chlorophyll-a. Did they use an in-house algorithm? Perhaps also mention why newer satellite data were not used?*

   SeaWiFS is active from 1997 to 2010, and its chlorophyll products have been widely used to calibrate numerical model results. In this study, we used SeaWiFS chlorophyll concentration based on OC4 algorithm (line 180), which is a fourth-order polynomial relationship between a ratio of visible bands reflectance and chlorophyll concentration. We also tried to use MODIS satellite data, but the quality of satellite image is unsatisfactory due to thick clouds.

4. *It will be useful if the authors could compare model results with actual observations or ranges. E.g. Figure 5 show pre and post hurricane simulations – perhaps the authors can compare the pre-hurricane values to typical Gulf values and the post hurricane to highest values measured during "windy" times (if there are data available).*

   We searched World Ocean Database (WOD) and found no data available 1 month before/after the landing of hurricane Gustav in the northern Gulf of Mexico.

5. *So much is different in the shelf waters during a hurricane – sediments stirred up, high levels of solids and nutrients in the water (including the surface water), likely breakup of stratification, impact on river loadings and discharge into the shelf and more. I therefore do not think comparing model results, where only differences in sediment enhanced light attenuation is accounted, with satellite data prove that one light attenuation formulation is better than the other. I think it is fine to show the comparison and speculating that it might be better, but I do not think the satellite data prove it one way or another. I think the authors almost make this point themselves by pointing out the limitations and uncertainty of the model.*

As we mentioned in section “Model Uncertainties”, there are many factors might potentially impact our model results. Compared with these factors, however, sediment-induced light attenuation is more important during hurricane because of the dramatic increase of SSC. Thus, the uncertainties associated with other terms can be overwhelmed by sediment term, and that is the major reason why we believe better resolving light attenuation by sediment might improve the performance of biogeochemical model during hurricane. Owning to the lack of field measurements during hurricane, satellite image is the only source which can help us evaluate model results.

General comments

6. *Abstract: The abstract seems reasonable. I believe the authors can say a little more about model uncertainty since they make a good point in the paper about all the uncertainties in the model. The authors mention "episodic hurricanes" in line 57, but I do not think the authors should rather mention "wind events" or another term when discussing the impact of high winds including tropical storms etc. The abstract also need to be changed based once changes is made to the paper to reflect any changes in the paper.*

   We add more information about model uncertainties in the abstract (see lines 42-43), and we use “hurricane event(s)” to replace other similar terms in the manuscript to avoid confusion (details see lines 100).

7. *Introduction: Line 44: Light is one of the primary agents for photosynthesis (also nutrients, temperature) Line 70: Since light attenuation is an important part of the paper, I think the authors should dig a bit deeper in what has been done, perhaps show their equations (or discuss conceptually) etc. I believe some models/papers have discussed CDOM and other influences on light attenuation.*

   For Line 44, we include nutrients and temperature (details see line 46). Line 70: since we show the equation in the following “Model Setup” section, here we add more background information about how these models estimate light attenuation due to CDOM and other light absorbers (lines 70-76).

8. *Model Description: Line 115: Why nitrogen and silica and not phosphorus? I believe some studies in the Gulf have shown that phosphorus can be important at certain times of the year. Perhaps a sentence why it was not included?*

   Primary production in the northern Gulf of Mexico is limited by phosphorus from May to July, so without phosphorus does not impact our model results (August 30th-September 10th). Details see lines 131-133.

9. *Line 116: Please expand on how chlorophyll-a was estimated. Fennel reference is fine, but perhaps add a sentence or two.*

   We add more information about how to estimate chlorophyll concentration in our model (134-136).

10. *Line 132: Expanding on my "main points": Provide additional details in the Introduction or this section about the light attenuation formulation used in the paper, what others have done in terms of sediment attenuation, the section of the sediment attenuation coefficients (0.059, 0.025 and 0.075). Are these values based on some median/percentile values? How much faith should we have in these values?*

    Additional information is added in Introduction. For sediment-induced light attenuation term in previous studies, we mentioned that "Justić and Wang (2014) tentatively employed a new scheme by connecting sediment-induced light attenuation with river discharge and hydrodynamics" (line 78-80). In fact, most biogeochemical models do not have sediment-induced light attenuation term because they are not originally developed for highly turbid water.

    In the updated manuscript, we re-design the sensitivity tests by increasing (decreasing) attenuation coefficient by 20 and 40%, respectively (lines 170-172). Although these coefficients in the sensitivity tests are artificially assigned, the results can provide valuable information about the response of primary production and nutrient to different light attenuation efficiency.

11. *Line 171: Should this part not be the Results and Discussion section?*

    The part quantitatively evaluates the performance of our model and confirms the readers that our model is well calibrated, so we still put this part in the "Model Validation" section.