Responses to Mara Freilich (Reviewer 3)

We thank the reviewer for the constructive comments and suggestions which will be very helpful as we revise the manuscript. Below the complete reviewer comments are shown along with detailed responses to each comment (reviewer comments in black, responses in blue font)

Review:

This manuscript presents a very well developed and well described parameter optimization process for biogeochemical models using available sea surface and profile observations. The authors show that a model parameterized with both surface chlorophyll from satellite and BGC-Argo profiles of chlorophyll and POC best represents the available observations of ecosystem state and fluxes. The authors demonstrate that the parameter choice has important implications for carbon cycling and export.

Comments:

2. Study Region

This section is very useful as an orientation to the region. Since one of the major objectives of the study is to analyze carbon export, it would be useful to include more in this section about what is known about carbon export in the Gulf of Mexico.

Response: We may include some key references in our revision, but would like to note that the major objective of this study is to evaluate the trade-offs between different observation types for biological model validation and parameter optimization. Our model-data comparison of carbon export is to illustrate the model performance and additional value brought by BGC-Argo profiles to parameter optimization. Studying the carbon export in the Gulf of Mexico is not within the intended scope of this study.

3. Methods

p. 6, Line 147: Earlier in this section, the satellite chlorophyll was adjusted to the float chlorophyll, but here the float backscattering is adjusted to the satellite backscattering. Why is this? These adjustments may have important implications for the POC profiles and partitioning of POC between phytoplankton, zooplankton, and POC. The reasoning for and implications of this choice should be explained.

Response: Chlorophyll and backscatter, regardless of whether they are obtained by satellite or in-situ sensors, are proxy measurements. That means they have to be converted and calibrated before they can be used. Here the BGC-Argo floats provided measurements of chlorophyll fluorescence and the volume scattering function at a centroid angle of 140^o and a wavelength of 700nm ($\beta(140^o, 700nm)$ m-1 sr-1), i.e. a measure of backscatter. The conversion from fluorescence into chlorophyll concentration was based on the sensor manufacturer's calibration. The conversion into bbp700 was performed by us following Green et al. (2014) and by cross-calibration against existing relationships with POC and phytoplankton biomass (Martinez-Vicente et al., 2013; Rasse et al., 2017). The resulting concentrations of phytoplankton biomass and POC as well as the ratio of chlorophyll to phytoplankton biomass are reasonable (please see figures 4 and 10). We will include a more in-depth explanation of this process in our revised manuscript.

Since this comment is similar to comments by Reviewer #1 under 3.1, we point also to those responses.

Green, R. E., Bower, A. S. and Lugo-Fernandez, A.: First Autonomous Bio-Optical Profiling Float in the Gulf of Mexico Reveals Dynamic Biogeochemistry in Deep Waters, PLoS ONE, 9(7), 1–9, doi:10.1371/journal.pone.0101658, 2014.

Martinez-Vicente, V., Dall'Olmo, G., Tarran, G., Boss, E. and Sathyendranath, S.: Optical backscattering is correlated with phytoplankton carbon across the Atlantic Ocean, Geophysical Research Letters, 40, 1154–1158, doi:10.1002/grl.50252, 2013.

Rasse, R., Dall'Olmo, G., Graff, J., Westberry, T. K., van Dongen-Vogels, V. and Behrenfeld, M. J.: Evaluating Optical Proxies of Particulate Organic Carbon across the Surface Atlantic Ocean, Frontiers in Marine Science, 4(November), 1–18, doi:10.3389/fmars.2017.00367, 2017.

p. 7, Line 167: What is the range of the vertical resolution of the ROMS model in the upper 200 meters? This information will be useful for comparison with the 1D model.

Response: The 3D model had 36 terrain-following sigma levels which means that the vertical resolution varies with the bathymetry. The vertical resolution in the 3D model ranges from a few meters near the surface to ~50 m near the depth of 200 m around the 1D site. We will add this during the revision.

We also point to our response to comments by Reviewer #1 under 3.2.

p. 8, Line 187: Since the biogeochemical model functional forms are essential to evaluating the model performance and the parameter optimization results, the biogeochemical model equations should be reproduced in this manuscript or in an appendix, rather than referring the reader to a different paper.

Response: The model code is published and freely available and the model equations, originally published in 2006, have recently been republished in the Appendix of Laurent et al. (2017). The model has been used in well over a dozen publications by our group and by hundreds of other researchers (the model is part of the widely used ROMS distribution). We do not agree that it is necessary or frankly appropriate to republish the same equations which each manuscript. This is also not common for other widely used models. We will make a specific reference to the equation in the Appendix of Laurent et al. (2017) when we revise.

Laurent, A., K. Fennel, W.-J. Cai, W.-J. Huang, L. Barbero, and R. Wanninkhof (2017), Eutrophication-induced acidification of coastal waters in the northern Gulf of Mexico: Insights into origin and processes from a coupled physical-biogeochemical model, Geophys. Res. Lett., 44, 946 – 956, doi:10.1002/2016GL071881.

p. 8, Line 199: "Zooplankton and small detritus were assumed to amount to 10% of phytoplankton biomass and the remaining fractions of POC attributed to large detritus." Why is this assumption made? Is this assumption only employed to define the initial condition?

Response: This assumption is also used for the open boundary conditions. For the biological model, the open boundary conditions are usually unavailable. Here we assumed the zooplankton and small detritus accounted for 10% of phytoplankton following Gomez et al., 2018 in which 20% was assumed for a different biological model in the Gulf of Mexico. We also do some sensitivity tests in the first year (2010) by perturbing this fraction to be 5%, 20%, and 40%. As shown in Figure 1, the choice of this fraction has little impact on the model results, e.g. phytoplankton, zooplankton, and POC. In the revised manuscript, we will add a short explanation of this assumption.

Gomez, F. A., Lee, S.-K., Liu, Y., Hernandez Jr., F. J., Muller-Karger, F. E., and Lamkin, J. T.: Seasonal patterns in phytoplankton biomass across the northern and deep Gulf of Mexico: a numerical model study, Biogeosciences, 15, 3561–3576, https://doi.org/10.5194/bg-15-3561-2018, 2018

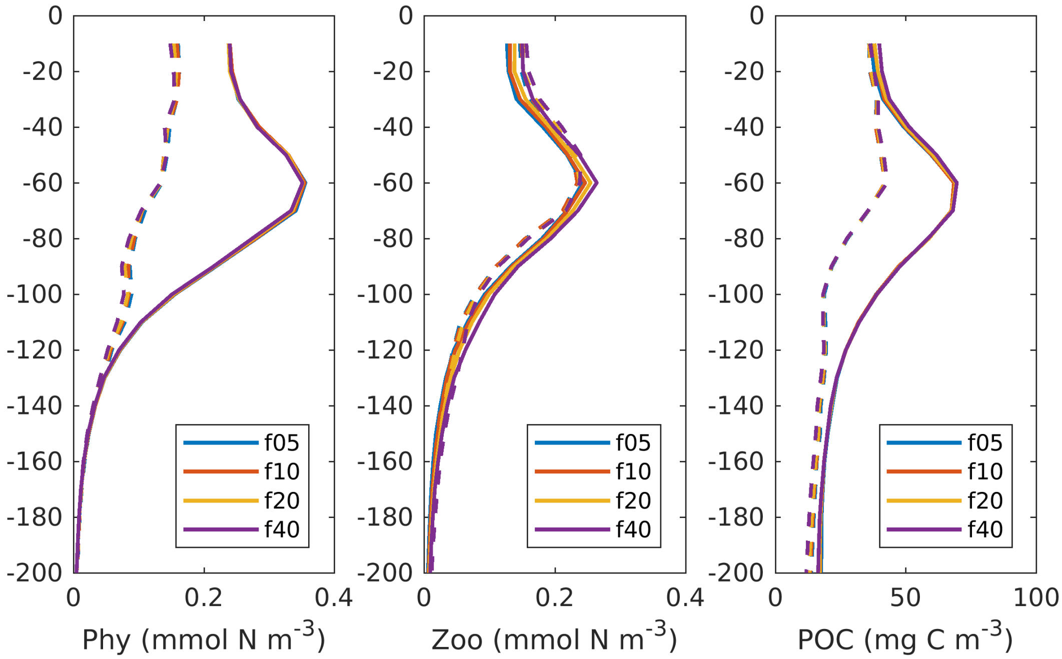

Figure 1. The simulated vertical profiles of phytoplankton, zooplankton, and POC from different sensitivity tests. The solid lines represent median profiles while the dash lines represent interquartile range.

p. 8, Line 210: The 1D model setup is sensible, but more information is needed to understand, evaluate, and reproduce the 1D model - Why is a 5 meter vertical resolution chosen in 1D and is the model sensitive to this choice? - What are the units for the diffusion coefficient and why are these values chosen? They are substantially higher than typical vertical diffusion coefficients in 3D models. - Are all of the parameters that are taken from the 3D model seasonally varying (the mixed layer depth, temperature, solar radiation, and NO3 below 100 meters)? If so, what the temporal frequency at which they are updated? - Why is the temperature and mixed layer depth determined using values from the 3D model rather than the floats, which also have that data?

Response:

There are two main reasons why we chose 5 m. First, the vertical resolution in our BGC-Argo floats is 4-6 meters in the upper 200 m. Second, a 5-m vertical resolution in the 1D model keeps the computational cost reasonable. As stated in sections 3.4 and 3.5 of the manuscript, one parameter optimization experiment ran 36,000 model simulations (30 simulations/generation * 300 generations*4 replications). It took us about one day to run a single parameter optimization experiment and we performed 13 1D optimization experiments in total (A1-A4, B1-B4, C1-C5). This is in addition to all the 3D model runs we performed.

We are confident the vertical resolution is appropriate because we performed tests of the effect by using different vertical resolutions including 10 m, 5 m, 3 m, and 1 m for optimization C4. As shown in the Figure 2 and 3 below, the choice of vertical resolution has little impact on model results except for some small jumps in the DCM depth (Figure 2e) and DCM magnitude (Figure 2) when the vertical resolution of 10 m is used.

The unit of the diffusion coefficient is m^2 day^{-1} and its value was tuned such that our 1D and 3D models behave similarly. We will add the unit of diffusion coefficient in our revised manuscript.

Finally, the physical input was obtained from daily output of the 3D model. Float data were not used as physical inputs because of their low temporal resolution (~2 weeks).

We also point to our responses to comments by Reviewer #1 under 3.2.

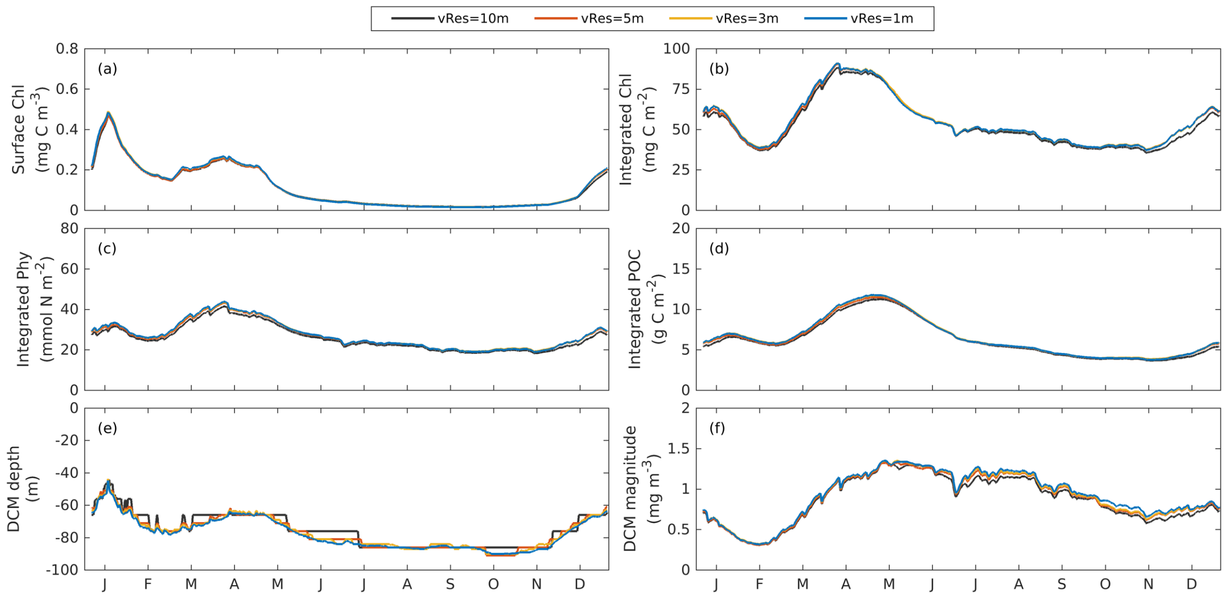

Figure 2. The simulated annual cycle of surface chlorophyll (a),vertically integrated chlorophyll (b), vertically integrated phytoplankton (c), vertically integrated POC(d), and the depth (e) and magnitude (f) of the DCM by using different vertical resolutions.

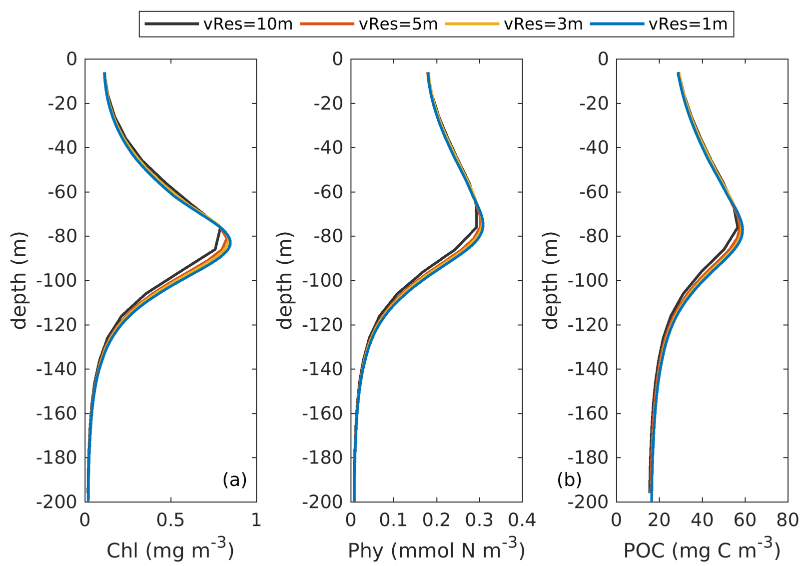

Figure 3. The simulated (colored lines) vertical profiles of chlorophyll, phytoplankton, and POC by using different vertical resolutions.

p. 11, Line 281: Why is the ratio of 0.1 between the sinking speed of the small and large detritus selected? How sensitive are the results to this choice?

Response: We refer to our response to Reviewer #1, comment on Pg 11 line 81-84, which we paste again here for sake of convenience:

These two parameters cannot be constrained simultaneously because they can compensate each other: an increase in the one parameter can be counteracted by a decrease in the other. This is a well-known and much discussed issue in the parameter optimization literature. Basically, there is no information in the available observations that allows to distinguish between the two. Including both parameters in the optimization will degrade the model's predictive skill. To solve this problem, one can either introduce more independent observations or prior knowledge by, e.g. fixing these parameters to their prior values or defining a link between these parameters. Generally, and also in this case, observations are insufficient to constrain all parameters and prior knowledge about the parameters has to be included. This is the common approach to addressing the underdetermination problem of parameter optimization in biological models.

In this study, the ratio of 0.1 was selected based on the prior values of the two parameters. Perturbing and estimating the ratio between two sinking velocities are equivalent to letting both parameters (wSDet and wLDet) into optimization and may degrade model's predictive skills. Without additional information about the depth distribution of Sdet and Ldet, which is not available, the ratio cannot be constrained.

The underdetermination problem and the practice of incorporating prior knowledges is also discussed on P20, L10-28 of our original manuscript.

4. Optimization of 1D model
This section is well written and clear. The discussion of the differences between the surface chlorophyll and vertical profiles is particularly clear and interesting.

Response: Thank you for this comments.

p. 13, Line 329: "Unlike phytoplankton, the observations show that the POC concentrations are 19 mg C m-3 at about 200 m depth because of the existence of detritus (Figure 4c)." What is the evidence that the POC is in the detritus class rather than zooplankton? This point is not supported, but is used later to discriminate between models.

Response: We would like to revise this as follows: _"… the POC concentrations are 19 mg C m-3 at about 200 m depth because of the existence of detritus, or zooplankton, or both"_. We would also like to revise on P15 L70-71 of original manuscript to _"In contrast to the observations where the phytoplankton's contribution is neglectable, the simulated POC at 200 m is dominated by phytoplankton (49%)"_.

Section 4.2: In the section for each experiment, remind the reader what data is used the optimization and which parameters are included in the parameter optimization

Response: As stated already in response to Reviewer 2, we will revise as suggested.

Could you plot the mixed layer depth of the 1D model for comparison to the DCM depth? In the 1D model, the mixed layer depth is the main physical control. One option would be to plot the mixed layer depth in figure 3e.

Response: Since the Figure 3e will be busy with a plot of the mixed layer, we will plot the seasonal mean levels of the mixed layer in Figure 4 and hope that will convey the information this reviewer is looking for.

The supplemental figure S1 shows that even the corrected satellite data does not capture the seasonality observed by the floats. One sentence here describing the differences that remain between the floats and corrected OC-CCI could help us to better assess why experiment A in particular does not capture the seasonality.

Response: In the experiment A, the satellite estimates of surface chlorophyll was used in parameter optimization and the model results were compared with the satellite data in the Figure 3a. The differences between the satellite and float data would not be a reason for experiment A not capturing the peak of satellite surface chlorophyll. It could be a result of many factors and might be improved by assigning a higher weight on the peaks of surface chlorophyll. However, studying the peaks of surface chlorophyll was not the purpose in this study. Nonetheless, the parameter optimization resulted in a large improvement in surface chlorophyll in the experiment A.

p. 15, Line 375: "Although a slight increase in the misfit occurs for the surface chlorophyll (_5%)," Is the increase of 5% relative to experiment B?

Response: The slight increase here was relative to the base case. The full sentence in the original manuscript is *"Although a slight increase in the misfit occurs for the surface chlorophyll (~5%), the total misfit is reduced by 75% compared to the base case."* We hope this is clear.

Table 2: In the caption, explain what the dashes in the table mean. In the cases where the parameters are not included in the optimization, are the values that are presented in table 1 used? It would be helpful to the reader if the experiment that is discussed in the text is highlighted. Since different parameters are used in the 3D experiment, include additional lines in this table for the parameters that are used in the 3D experiment.

Response: Thanks for this comment. We will revise as suggested.

Figure 3: Include what data is used for each experiment in the figure caption. Could error bars be added to the observational points in this figure?

Response: Yes, we will revise as suggested.

Figure 4: What do the error bars represent? Are they the interquartile range?
Response: Yes, they are interquartile range. We will state that explicitly.

5. 3D biogeochemical model
p. 16, Line 421: How were the parameters that were modified manually chosen? It would be useful to provide more discussion of this choice.

Response: Same as comment "5. 3D biogeochemilca model" by Reviewer 1 and same response. We paste it again here for sake of convenience:

We did manual modifications because when the resulting parameters were directly applied, the model-data agreement in 3D models was not as well as in 1D models in some aspects, but the most important features were well preserved. In experiment C, chlorophyll concentrations in the upper layer of 3D model were lower than the 1D model and farther away from observations. This might be a result of differences between 1D and 3D models and has been also reported in other studies (e.g. Kane et al., 2011; Hoshiba et al., 2018). However, the depth of the DCM and the non-zero POC concentrations at 200 m with appropriate contributions from each component were well preserved. We therefore did some tests manually by tuning one or two parameters and final set the KNH4 to 0.01 in order to have a better agreement with respect to observations. All of these modifications were based on our tests.

In our case, although the 1D parameter sets could not be used in 3D models directly, they were sufficient to reproduce the main features as in 1D models and largely simplified the following subjective tuning of 3D models by limiting the number of parameters to be adjusted. We will include some discussion in our revised manuscript.

p. 17, Line 442: Here the authors point to specific parameters that were inappropriate, but on p. 10, Line 246 the authors say that parameters are not allowed to exit a predefined range (which is shown in table 1). This seems inconsistent given that the method could have excluded inconsistent values. Could this difference be explained in more detail?

Response: We think there might be a misunderstanding here. The predefined range of parameters was a collection of their values that had been used in other studies, models, and ecosystem regimes. That means not all parameter values within this range are appropriate for our model. The parameter optimization is therefore to search for an optimal parameter set within this predefined range by minimizing the misfit between model and observations. However, the parameter optimization cannot constrain all parameters (e.g. producing the inappropriate estimates) because of insufficient observations, which is referred to as the underdetermination problem (see earlier responses here and for Reviewer 1). For instance, in our experiment B, the parameter optimization fitted our model to chlorophyll observations but sacrificed other aspects of the model's performance which were degraded, e.g. the vertical profiles of phytoplankton. That is why we said the estimates of the maximum ratio between chlorophyll and phytoplankton (θmax= 0.0158) in experiment B was inappropriate. An appropriate estimate of this parameter requires more independent observations, e.g. of phytoplankton that we used in the experiment C.

Section 5.1: The authors state that the 3D model does not perform well in coastal regions and therefore choose to exclude those regions from the model evaluation. This may be justified based on the statement in section 2 that there is little cross-shelf exchange in the Gulf of Mexico. However, this point should be discussed in more detail in order to justify ignoring the shelf. In particular, it is important to discuss the extent to which nutrients are supplied to the oligotrophic regions from either the shelf or the open ocean boundary. What is the importance of the boundaries relative to the biogeochemical cycling in the interior? In a 3D model, the boundaries could be very important for setting the primary production in the oligotrophic region.

Response: Firstly, we excluded coastal regions because we have no float data in coastal regions used for parameter optimization and model validation. All parameter optimization experiments in this study were designed for modelling the deep ocean part of the Gulf of Mexico. We will clarify this in our revised manuscript.

Secondly, if we define the boundary between coastal regions and open ocean as 1,000 m, the DIN fluxes (NO3+NH4) across the 1,000 m isobath is a sink for the open ocean and nearly balanced by the inputs from open boundaries. Specifically, for the upper 200 m where the primary production occurs, the amount of DIN transported is 2.7×10^{11} mol N yr^{-1} (sink) from the open ocean into coastal region and 2.1×10^{11} mol N yr^{-1} (source) from open boundaries. The primary production in the deep part of the Gulf of Mexico is 10.6×10^{11} mol N yr^{-1} (sink for DIN) which is mainly supported by the vertical transport and interior nitrogen recycle in the upper 200 m.

p. 18, Line 451: What is meant by "different spatio-temporal scales between the two model versions"? This point seems important and could be clarified. Is it referring to time stepping, resolution, retention in a 1D location, or the presence of a seasonal cycle?

Response: The different spatial-temporal scales refer to a lot of factors here. The 1D model is configured in a single station while the 3D model covers the whole Gulf of Mexico. In the vertical direction, the 1D model only simulates the upper 200 m but the 3D model has the whole water column. For temporal scales, the 1D model is simulated for one year but the 3D model runs for 6 years from 2010 to 2016. The model steps are also different between the 1D and 3D models. As suggested, we will clarify this in our revised manuscript.

p. 19, Line 479: "cannot" should be "can not"

Response: We will revise as suggested.