The author's thank the referees for their insightful and detailed comments. We have provided our responses in blue.

Response to Anonymous Referee #1

On line 778, they say "Without these observations, it is difficult to gauge mismatches between model and [data]". It is true that more in situ measurements would enable refinement of this component of the model, but there are two additional aspects that warrant mention. (a) The range of realistic values of C:Chl is high (e.g. 25-250). Therefore, a poorly modeled C:Chl ratio could result in a large misfit even when the modeled nitrogen/carbon-based ecosystem dynamics was excellent. (b) Quantifying uncertainty can be addressed in the absence of observations simply by conducting sensitivity tests on the components of the C:Chl submodel.

We thank the reviewer for the additional suggestions. We have included these points in our discussion of C:Chl ratios where we state "In addition to validation, these measurements are needed to avoid erroneous model tuning. For instance, a model that exhibits significant mismatch with respect to surface Chl may in fact accurately estimate carbon based phytoplankton biomass while using unrealistic C:Chl ratios. One could arrive at incorrect conclusions about regional ecosystem dynamics as a result of modifying model parameters or structure in an effort to better fit Chl observations. Given the importance of C:Chl ratios in PBMs, future studies should quantify uncertainty in modeled Chl through sensitivity experiments focused on C:Chl model parameters and formulation, with explicit comparison to direct field measurements of phytoplankton C:Chl." (lines 790-797).

The authors should be good role models and examine whether their modeled phytoplankton biomass (i.e. in carbon) is high.

We agree that examining phytoplankton biomass in carbon (or phytoplankton C:Chl ratios) is important to conclusively determine if a model is accurately representing phytoplankton biomass. Unfortunately, there is a distinct paucity of true phytoplankton biomass (C or N) measurements in the Gulf of Mexico. In particular, we are not aware of any carbon based measurements in places like the Campeche Bank where the model shows its greatest Chl-based model-data mismatch. We are working with collaborators to try to get microscopy and flow-cytometry derived field measurements of phytoplankton carbon in the Gulf, but these are not yet available.

I suggest they reference some of the problems with accuracy in modelled vertical diffusivities and currents that I mentioned in my original review.

We agree that errors due to both biology and physics could be contributing to model-data mismatch at the DCM. We have now included this point where we state "Future PBM studies need to focus more effort on resolving ecological dynamics responsible for the formation of the DCM. Errors originating from the hydrodynamic model or use of temporally averaged velocity fields used in offline models may also contribute to model-data mismatch at the DCM. Vertical mixing is particularly important to PBMs, but is often poorly validated in hydrodynamic models. Greater coordination between physical modelers and biologists to constrain vertical fluxes should be considered an important avenue for improving PBM simulations moving forward." (lines 868-874).

They could add a couple of sentences reminding the reader that modeling diet composition and secondary production relies entirely on the modeled prey selectivity formulation. They may also want to expand on this point to say that there exist a wide array of published multiple prey responses that have the potential to predict very different consumption rates. They should advocate for modeling studies that assess the implications of these different formulae to simulating secondary production.

To further address this point we now have the following text "Uncertainties in model grazing formulations could also contribute to model-data mismatch (Gentleman et al., 2003a; Sailley et al., 2015). Future in situ grazing measurements are needed to enable an objective selection of grazing formulations and parameter values. In particular, field studies that shed light on prey selectivity would be useful for parameterizing PBMs with multiple mesozooplankton functional groups, such as NEMURO-GoM. Such studies are challenging, however, because the difficulty of making in situ grazing measurements on mesozooplankton, combined with the inherent uncertainty of these measurements can make it challenging to differentiate between, for instance Ivlev and Holling's disk grazing formulations (e.g., Fig. 4 in Morrow et al. (2018)). Nevertheless, differences in parameterizing grazing can lead to substantially different model behavior (Anderson et al., 2010; Sailley et al., 2015; Wainwright et al., 2007). In NEMURO-GoM, secondary production and dietary preferences of the mesozooplankton community are both strongly influenced by model grazing formulation. While we carefully chose parameterizations that gave reasonable fits to extensive field datasets of zooplankton biomass and grazing rates, this does not preclude the possibility that other functional forms would have more accurately simulated zooplankton dynamics. Hence, future PBMs should investigate how different grazing formulations impact zooplankton dynamics in the region. We especially recommend collaborations between experimentalists (potentially using new techniques such as DNA metabarcoding of gut contents) and modelers to develop synthetic approaches with the potential to quantitatively assess the realism of different grazing formulations."(lines 894-912).

Response to Anonymous Referee #2

The model greatly overestimates chlorophyll in the coastal region, but the authors did not recognize this problem… How does affect the simulated zooplankton biomass and dynamics? Maybe they are having the right spatial patterns for zooplankton biomass but wrong underlying dynamics.

We respectfully disagree with Referee #2 in regards to not recognizing that the model overestimates Chl in the coastal region. To draw attention to this issue we state "While a clear shelf signature is well resolved in NEMURO-GoM, the model-data mismatch is greater on the shelf compared to oligotrophic region"(lines 761-762) and "In our model, the most noticeable surface Chl model-data mismatch occurs on the southern GoM shelf (Campeche Bank (CB)), where the model consistently overestimates surface Chl" (lines 787-788). We also note that we intentionally included Fig. 2F to draw readers' attention to our models imperfect simulation of coastal Chl. We intentionally included model-data validation plots that highlighted the weaknesses of our model, rather than only showing the strengths. In the revised manuscript we have also included a specific section in our Discussion section dedicated to addressing the issue of surface chlorophyll discrepancies in the model.

We also believe that the reviewer is overstating the model-data mismatch on the shelf when he/she states that our model "greatly overestimates chlorophyll in the coastal region". Many other models of the GoM also struggle to accurately capture variability in phytoplankton biomass on the shelf. We believe we have made a significant step forward in this regard by developing a model that reproduces the dominant spatial patterns on the shelf. Our model even accurately resolves the small shelf signature in the western GoM. On average, the model typically overestimates Chl by anywhere from a very slight overestimate to a factor of 2 overestimates (and occasionally a factor of 3 overestimate). This is not a huge discrepancy compared to other GoM models (see, for instance Fig. 3 of Gomez et al. 2018 which shows that their model consistently underestimates phytoplankton biomass in the Mississippi Delta and Texas Shelf regions, with almost 5-fold underestimates common in the summer months). Furthermore, as we note, the shelf model-data discrepancy is mostly driven by a substantial overestimate of Chl on the Campeche Bank. Overall, we actually consider our model to have relatively decent agreement with surface chl measurements on the shelf, considering that it was a model specifically focused on open ocean regions. Nevertheless, there are clear areas for improvement on the shelf (e.g. more realistic C:Chl models and including (non-climatological) river nutrient data). We have highlighted these in the revised manuscript (lines 784-797).

The authors should include a good validation for nutrients and primary production in the coastal region; otherwise, we do not know if the model is reproducing well cross-shore patterns. The authors should also include a validation for phytoplankton structure (e.g., diatom to total phytoplankton ratio).

We do not believe that substantial validation of coastal region nutrients, primary production, and phytoplankton structure in the coastal region is justified in this (already long) manuscript for two reasons:

1) Our stated focus is on zooplankton dynamics in the oligotrophic GoM, rather than on phytoplankton dynamics in the coastal region.

2) There is a paucity of in situ *data to compare the model to: Spatiotemporal patterns in nutrients and primary production are highly variable in the coastal region, and unfortunately the publicly available* in situ *measurements (at least that we have access to) are not sufficient to resolve this spatiotemporal variability sufficiently to enable detailed quantitative comparison. This does not mean that we have not considered such things as nutrient concentrations, primary productivity, and SP:LP ratios in our model. For instance in the supplement, we noted that LP:SP ratios are an important metric for evaluating the simulated ecosystem which is why we considered this metric during the model tuning process. We have now added an LP:SP ratio map to the supplemental material, so that the reviewer can see this metric. However, there are few true* in situ *measurements of LP:SP biomass ratio in the Gulf to use as true "validation" rather than a subjective validation metric. We hope that future in situ research programs will produce more of such measurements, or alternately, that the planned NASA PACE project (and accompanying* in situ *validation studies) will enable accurate determination of phytoplankton taxonomic groups from satellite.*