

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

Response to Anonymous Referee #1

1. Model development

Replacing quadratic mortality with linear mortality for all biological variables except PZ is a sensible simplification, but should be supported by references to model sensitivity studies about this issue (e.g. Anderson et al., 2015)

We thank referee #1 for the suggestion and will include the citation in the revised manuscript.

It would be beneficial to have a sentence explaining what they mean by “monotonic”, so that readers appreciate the problem and know to follow them in making the change.

We reference the “non-monotonic” behavior of the default ammonium inhibition term to describe the result that at high nutrient concentrations total nutrient uptake by phytoplankton (i.e. uptake of NO_3 + uptake of NH_4) can actually decrease despite increases in the total nitrogen available to phytoplankton. This is unrealistic and hence was replaced with a more commonly used ammonium inhibition term. We will make this clearer in the revised manuscript.

Use of the Platt light limitation is commonplace, but it would be useful to state whether the original NEMURO model did or did not include photoinhibition.

The original NEMURO light limitation functional form does includes photoinhibition implicitly. Platt functional form was used because it allows one to explicitly control the amount of photoinhibition. Additionally, the Platt functional form is now more commonly used and thus parameter values were easier to find for comparison (e.g. initial slope of the PI curve (α)). We will acknowledge this in revised manuscript

Replacing constant C:Chl with variable C:Chl is potentially the most critical of all their changes, since this variable dictates the modeled Chl levels, which are used to compare the model and data. While they do provide the equations in the appendix, a brief overview of how those equations work would be beneficial.

In addition to having the description in the supplemental, we will include text in the revised manuscript dedicated to describing how the Li et al. (2008) C:Chl model equations work and the basis for them. For example, we will include a brief sentence on the original Geider et al. (1998) C:Chl model which is foundational to the Li et al. (2008) model.

2. Model validation

The model-data comparisons state the different values, but this does not give the reader any sense on whether the differences actually matter. Any text they could add about potential ecological significance of these differences would be helpful.

We acknowledge that there could be more discussion on the potential ecological significances of model-data mismatch to give the reader a better sense of how the model-data mismatch may influence dynamics in the model. We thank review #1 for this suggestion and will address the ecological significance of the main model-data mismatch in our discussion (e.g. surface Chl on the shelf and grazing mismatch).

As noted above, model-data comparison of Chl will be impacted by modeled C:Chl ratios. It would be useful to have a quantitative sense of the model uncertainty related to this model component. At the very least, the issue should be discussed somewhere.

We agree that quantifying uncertainty in C:Chl ratios is important moving forward as future models will likely continue to depend heavily on satellite Chl for the bulk of the model validation. Unfortunately, there is a paucity of in situ C:Chl data from the Gulf of Mexico that can be used for validation. We have ensured that our estimated C:Chl values are reasonable given patterns found in similar ecosystems, but in the revised manuscript we will also emphasize the need for more in situ measurements of this property in the GoM.

This paper's use of a constant vertical mixing coefficient is not representative of higher mixing at surface as compared to interior. Vertical profiles of light depend on attenuation coefficients, which in this model appear to use Beers's Law, and thus are oversimplifications (e.g. see Anderson et al., 1993, 2015). None of these sources of error are mentioned in the text, but it would be helpful to identify them and therefore guide future investigations.

The model does not use a constant vertical diffusivity but one that varies in space and time and is calculated in the mixed layer package of MITgcm using the KPP mixed layer parameterization which is commonly used in hydrodynamic ocean models. Vertical diffusivity was constant in the one-dimensional model during our tuning process. We acknowledge that Beer's law is an oversimplification for the attenuation of light particularly in highly riverine influenced areas which may have high turbidity associated with suspended material and not with high Chl. We will note this in the revised manuscript.

Their simulation always resulted in modeled DCM being collocated with the maximum growth rate, which is not supported by observations. On pg 34 line 841, they argue that growth rates at a DCM must be high to balance the high mortality at a DCM, a conjecture they based on their (unsupported) assertion that zooplankton abundance covaries with phytoplankton abundance. I question this assertion.

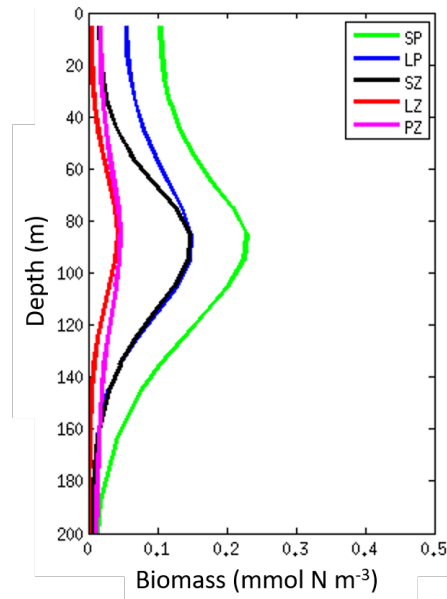


Figure: Average model biomass profiles from a spatial average over a 2° x 2° sample box located in the oligotrophic GoM with an origin at 24° N, 91° W that shows high correlation of SZ and phytoplankton biomass.

We found that in all of our NEMURO parameter sets small zooplankton (SZ) biomass closely varied with phytoplankton biomass, and this follows fairly clearly from the differential equations used for protistan growth. At the low biomass conditions existing in the open GoM, protistan growth rates are almost a linear function of prey (small phytoplankton) biomass, while mortality scales with the biomass of large and predatory zooplankton. Increased phytoplankton biomass thus leads to increased small zooplankton biomass, until the increase SZ growth is balanced by increased SZ mortality. This leads to a strong correlation between SP and SZ with depth in the model. Whether there is a similar correlation in the field is unknown, because of the paucity of field studies that differentiate phagotrophic and non-phagotrophic protists. The attached figure shows typical profiles of SP, LP, SZ, LP, and LZ found in the center of the oligotrophic GoM. As a response to referee #2 we included a figure showing the higher specific rates at the DCM in the model while measured specific phytoplankton growth rates are lowest in the DCM.

They found that grazing by small mesozooplankton was overestimated, whereas grazing by large mesozooplankton was underestimated, both by a factor of about 2 (pg. 22 and pg 33). I am skeptical about their explanation that this is due to modeled grazing reflecting functional groups as opposed to size classes, especially given that the two groups' biomasses were reasonably simulated. It seems far more rational to assume the problem lies with the characterization of grazing, especially since grazing formulations are known to be highly uncertain, particularly when considering ingestion of multiple prey types.

We agree with referee #1 that errors in the grazing functional form could also be a source of error leading to the model-data mismatch. We will include this point in the revised manuscript and discuss the need for future grazing measurements in the field to help select among a group of grazing functions. However, we do believe that most of the issue with the balance in grazing between PZ and LZ is due to the fact that PZ in NEMURO is explicitly defined and parameterized

as a group of higher trophic level zooplankton that can feed on LZ. In reality, while there is a correlation between size and trophic level in the ocean, there are many predatory zooplankton that are <1-mm and many suspension-feeding zooplankton that are >1-mm.

3. Parameter sensitivity

Sections 3.4 and S5 are dedicated to what they call their “Parameter Sensitivity” study.

This is not a sensitivity study in the traditional sense, which assess the robustness of a model output to variations in parameters (e.g. +/- 10%). Rather it is more an examination of how the model output changes if some subset of the original NEMURO formulations/values were used... I think this analysis could easily be removed without any loss to the paper.

We used this analysis to demonstrate the relative importance of changing parameters from default NEMURO, but now agree with referee #1 that this section is not necessary in the main text of the manuscript. In the revised manuscript we plan to move the analysis to the supplemental where we already have a table outlining the 18 parameter sets used in the analysis. Removal of this section will also allow us to elaborate on some of the points put forth while not increasing the length of the manuscript further.

4. Simulated mesozooplankton diet

Their use of trophic level was a clever way to summarize zooplankton diet information. However, their findings that, on the shelf, LZ is more herbivorous and PZ is more carnivorous, and that protists are more important diet source in oligotrophic are not unexpected since they prescribed their model to have that output

We agree that these results follow from the ways in which the zooplankton are parameterized, however these results are not immediately obvious a priori. While we certainly would have predicted a priori that LZ would feed predominantly on protists in oligotrophic regions and on large phytoplankton in coastal regions (and hence have higher trophic levels in the open GoM), it was not readily apparent that PZ would have lower trophic levels in the open ocean. Indeed, one might have expected a priori that PZ trophic level would follow that of LZ.

5. Simulated secondary production

Despite this being a stated objective, secondary production was only given a paragraph in the text, and is not mentioned in the discussion/conclusions. This is a missed opportunity to gain some more insight into the GoM function. If the authors don't want to flesh out the analysis, I suggest removing it from the objectives and possibly also from the results.

We understand the reviewer's interest in secondary production and plan to elaborate on this discussion in the revised manuscript. We also note that we are preparing a second manuscript that adds an individual-based model of larval fish to explicitly investigate the relationship between secondary production and food availability and ultimately starvation for Atlantic Bluefin tuna larvae in the Gulf of Mexico.

6. Technical Corrections

We appreciate the detailed corrections provided and will implement them accordingly.

Response to Anonymous Referee #2

1. Specific Comments

The tuning is fairly qualitative and does not use statistics to minimize the model-data mismatch or at least this was not mentioned. For instance it is indicated that DCM magnitude is used as a benchmark during the tuning but observed DCM biomass was not available. This suggests that there was no systematic comparisons with observations during the process. Clarifications on the procedure could be added.

While some studies have used quantitative cost functions as an approach for objective model tuning, such approaches are often limited to a very restricted number of benchmark measurements (e.g., satellite-derived chlorophyll and NPP). It is not immediately clear how to implement such approaches when a wide array of expected ecosystem properties are being considered (e.g., we considered sea surface chlorophyll, NPP, relative ratios of large:small phytoplankton, DCM depth, DCM biomass, ratio of phyto growth to zooplankton grazing, mesozooplankton biomass, mesozooplankton size structure, protistan grazing, and mesozooplankton grazing). These measurements come with different distinctly different levels of confidence, number of measurements available, and in some cases (e.g., relative contribution of large phytoplankton) had to be based on intuition from other similar regions. For this reason, we relied on a semi-quantitative approach – we compared results of multiple simulations to all of our expected benchmarks to search for simulations that gave a reasonable fit in all ways, rather than searching for one that minimized a specific cost function. We will add text to the supplemental and briefly in the revised manuscript which is similar to our response here to further clarify our tuning process.

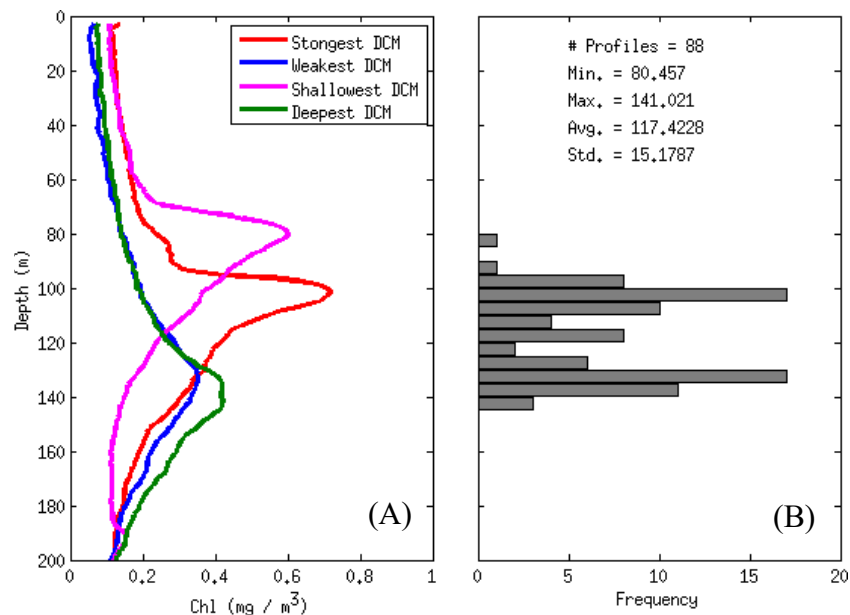


Figure (A & B): Observed Chl profiles from calibrated fluorimeter obtained during May 2017 cruise (A). Frequency of observed DCM depth (B).

With respect to the DCM depth and magnitude that the referee enquired about, 88 calibrated fluorescence profiles were obtained from our Gulf of Mexico cruise in 2017 and these were used for reference in tuning (i.e these profiles provided a target value with chlorophyll concentrations at the DCM of $\sim 0.5 \text{ mg m}^{-3}$). We will clarify this in the supplemental. No measurements of LP abundance were available but was used as a baseline from what the authors would expect to find. All other target metrics used in the one-dimensional model tuning are averages from synoptic data sources (SeaWiFS – Surface Chl, SEAMAP – MZB, SEAMAP – DCM depth). We will add text in the revised draft to clarify this.

Although Text S3 is too long, some of the tuning explanations could be incorporated to the manuscript as it seems relevant to the reader. For instance a schematic of NEMURO including the default and modified pathways and parameters would be useful for the reader to get a general sense of the modifications. Other technical information could also be added to the manuscript.

We will include more detail of the model tuning procedure and also include some text giving the reader an understanding of the sources from which we obtained our initial conditions and forcing.

Vertical profiles of chlorophyll concentration would have been very useful for the validation... A comparison of monthly averaged vertical profiles of chlorophyll in WOD and in the model would be informative since the magnitude of the DCM cannot be validated with the SEAMAP data.

We agree with referee #2 that vertical profiles of Chl are useful to increase confidence in the model solution. When writing the manuscript we did not have access to CTD data from our 2018 cruise and so we relied on the extensive SEAMAP fluorescence profiles for validation of the DCM. Since then we have processed the 2018 CTD data. In our revised manuscript we plan on including an additional validation figure devoted to show point-to-point model-data comparisons of Chl profiles based on the samples collected during both May 207 and 2018 lagrangian process study cruises. This figure will also include a histogram of DCM depths from SEAMAP fluorescence profiles and simulated DCM depth at corresponding times and locations.

As mentioned above, the validation is thorough but there are some mismatches that are not properly acknowledged. For instance there is a large discrepancy between observed and simulated surface chlorophyll on the shelf (Fig. 2F) but it is referred to as a good agreement in the manuscript (L419, L755-756). The model clearly overestimates surface chlorophyll on the shelf, which may have implications for the secondary production. This should be discussed.

We agree that the model overestimates Chl on some portions of the shelf. However, the model does a significantly better job overall at resolving the coastal signature in comparisons to past studies. We acknowledge this in the manuscript where we state “While a clear shelf signature is resolved in the NEMURO-GoM, we find greater model-data mismatch on the shelf compared to oligotrophic regions. This is an expected finding when considering the model incorporates climatological river forcing while actual variability is in reality much more complex” – line 790.

We acknowledge some text could be added discussing how this model-data mismatch will influence secondary production. Text will be devoted to this topic in the revised manuscript.

Also, regarding the model-data mismatch in grazing rate (L546-549), the statement that it is due to the model functional types not reflecting the size classes would be true only if the total grazing was similar in the model and observations, which is not the case for mesozooplankton. A better explanation should be provided.

We believe that there are some subtle points here that need to be addressed, and we will include them in the revised manuscript. In detail: while there is model-data mismatch for total mesozooplankton grazing it is not entirely clear that this model-data mismatch is outside the inherent uncertainty associated with the gut pigment approach to grazing measurements. Specifically, after zooplankton consume chlorophyll the acid in their guts will begin to convert chlorophyll to phaeopigments, although this conversion is neither instantaneous nor complete. Ideally, gut pigment measurements should consider both chlorophyll and phaeopigments, however, the presence of chlorophyll in detrital aggregates that are collected during zooplankton net tows can lead to substantial contamination. For that reason, we have only included phaeopigment when computing mesozooplankton grazing, which leads to a likely underestimate. Furthermore, the approach assumes that any group of mesozooplankton has a constant gut throughput time, which of course is an oversimplification of the biology of complex organisms. Any mesozooplankton grazing measurements thus need to be treated with considerable uncertainty, hence it is not clear that the model is necessarily wrong in terms of its estimate of overall grazing (note that we are certainly not arguing that the model is right and the measurements are wrong, simply that within the uncertainty inherent to the field measurement approach it is not possible to conclusively determine that the model is wrong about total mesozooplankton grazing rates). However, the model is fairly clearly wrong about the proportion of grazing mediated by PZ and LZ. We believe that the error in this ratio is due to the fact that PZ in NEMURO is explicitly defined and parameterized as a group of higher trophic level zooplankton that can feed on LZ. In reality, while there is a correlation between size and trophic level in the ocean, there are many predatory zooplankton that are <1-mm and many suspension-feeding zooplankton that are >1-mm. We will attempt to make this clearer in the revised manuscript. Motivated by this comment and comment from referee #1 we will include some text discussing the potential source of error from the grazing functions.

Specific growth rates at the DCM are discussed during the validation (L556-557) but the data are not presented. A figure of the specific growth rate profiles could be added to the supporting material?

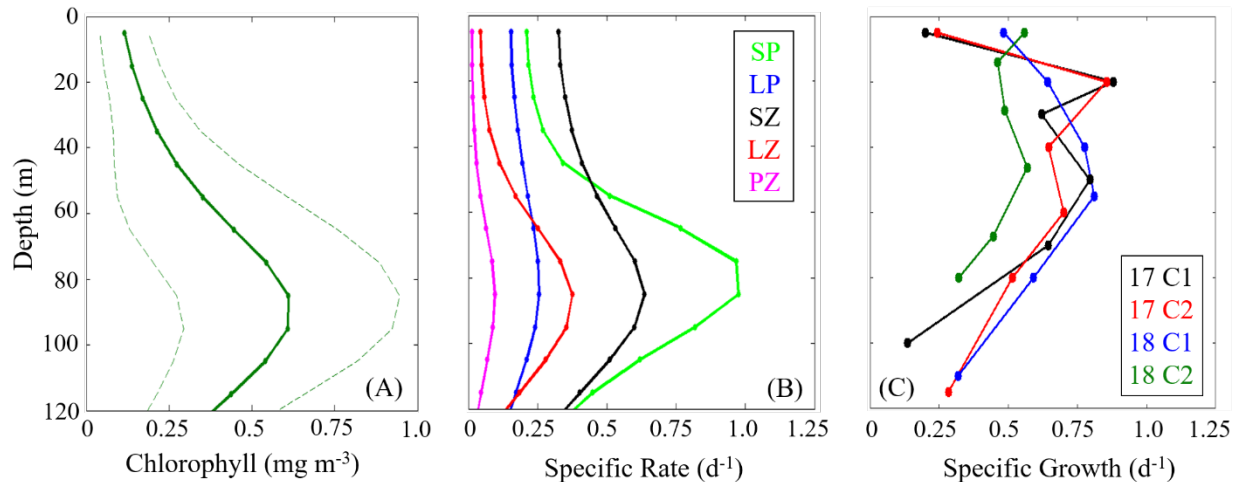


Figure: Simulated average chlorophyll profile in the center of the oligotrophic GoM (A). Comparison of simulated specific growth and grazing rates (B) to observed phytoplankton specific growth rates from dilution experiments conducted during May 2017 and 2018 lagrangian process study cruises (C). Final sample depth was always located at the observed deep chlorophyll maximum.

We will add a figure of specific growth rates from our 2017 and 2018 cruise in supporting material.

Regarding the "sensitivity" analysis, model results are compared with observations (e.g. Figure 6). Some model-model comparisons could also be presented. For instance, mesozooplankton biomass cannot be evaluated because the observations are depth averaged (L595-597) but this information is available in the model and therefore the effect of parameter/functional changes can be evaluated with the model.

A full parameter sensitivity analysis with parameter changes and functional form changes could be focused on in a future manuscript. Indeed, our offline modeling approach makes this type of detailed analysis possible. Table S3 outlines the results of 18 different parameter sets which can give the reader a sense of how our parameter value modifications influence simulated mesozooplankton biomass. Given the (already long) length of this manuscript, and referee #1's comment that the analysis could be removed entirely, we believe that it is not best to add additional analyses.

Including denitrification would probably result in lower phytoplankton biomass, hence a better agreement with observations. It might also alter the trophic pathways or at least the magnitude of the fluxes. Therefore, I wonder why this important process for shelf regions was not included, as it is relatively easy to implement (e.g. Fennel et al, 2006). The potential impact on the model results should be discussed.

We agree that denitrification can be important on the shelf and that it could be implemented in future versions of the model. However, it is not part of the standard NEMURO model and was not necessary for us to implement it, because our focus in the study was on the open ocean regions of the GoM where zooplankton biomass is likely to be limiting to higher trophic levels. In the revised

manuscript, however, we will highlight the fact that NEMURO does not include denitrification which can be important on the shelf.

The GoM shelf area includes the western Florida shelf, the northern GoM shelf, the western GoM shelf and the northern Yucatan shelf, which have different characteristics. These differences are relevant for the study.

We agree that the Gulf of Mexico shelf has many distinct regions (and any one of these distinct regions could be the focus of another manuscript), however in this (already long) manuscript, we prefer to keep our focus on the open ocean regions. In our own analyses we found that although the model-data mismatch was significantly worse on the Campeche Bank relative to other shelf regions the ability of the model to capture the temporal variability was similar among shelf regions.

2. Technical Corrections

We appreciate the detailed corrections provided and will implement them accordingly.

Response to Anonymous Referee #3

1. General Comments

A validation for nutrient's patterns should be included to gain additional confidence in the model results.

Unfortunately there are surprisingly few nutrient measurements from the GoM available in public databases. When tuning the model, two things that we paid attention to were the expected collocation of the nutricline and deep chlorophyll maxima, and the expected low nutrient concentrations in the surface layers of the open ocean. The model accurately represents these dynamics. However, given the already long length of this manuscript and our specific focus on mesozooplankton dynamics, we do not believe a figure focused on nutrient dynamics is warranted.

The model is not able to reproduce surface chlorophyll patterns in the coastal regions (bottom depth <1000 m). Is this a consequence of the model parameterization or a misrepresentation of river runoff fluxes?

We note that although the model does not do a great job of reconstructing average surface chlorophyll temporal variability in the coastal regions, it actually does a very good job (compared to previously published models and thus a clear step forward) of estimating spatial patterns of chlorophyll, particularly on the shelf. As we have noted in the manuscript, the region where the model performs worst is the Campeche Bank near the Yucatan Peninsula, where the model consistently overestimates phytoplankton biomass. The model's issues with estimating temporal variability on the shelf stem primarily from three issues: 1) we use climatological river forcing, which has a disproportionate impact on nutrient supply to the shelf. 2) because the primary focus of the model was on the deep ocean GoM, when tuning we focused more on this region. For instance, we considered the non-diatoms in the model to represent cyanobacteria, rather than the

nanoflagellates and dinoflagellates that would play important roles near the coast. 3) the noted (and substantial) errors in the Campeche Bank are disproportionately important to the temporal signal of <1000 m chl when averaged in the model.

A mean time series for all the coastal regions in the GoM does not seem to me appropriate, because there are important differences among shelf regions. I would suggest an independent comparison for the shelves off Louisiana-Texas, Mississippi-Alabama, west Florida, and Yucatan.

We fully acknowledge that there are issues (as we noted above) with including only a single plot for all coastal regions. However, our focus is not on the coastal regions, and this is already a very long manuscript, so we do not believe that separate panels for each of the shelf regions is warranted.

A figure showing chlorophyll vertical profiles should be included in the paper main body or the Supplement, ideally displaying data for each season.

We agree with referee #3 that vertical profiles of Chl are useful to increase confidence in the model solution. When writing the manuscript we did not have access to CTD data from our 2018 cruise. Since then we have obtained this data. In our revised manuscript we plan on including an additional validation figure devoted to show point-to-point model-data comparisons of Chl profiles based on the samples collected during both May 2017 and 2018 lagrangian process study cruises. This figure will also include a histogram of DCM depths from SEAMAP fluorescence profiles and simulated DCM depth at corresponding times and locations.

The in-situ data used for model validation was mostly based on measurement collected in two cruises during May of 2017 and 2018, with all the cruises stations located in the open GoM. These data do not allow evaluating whether the model is reproducing cross-shore patterns or seasonal variability. There is abundant data in the northern GoM that the authors could use to improve the model validation, like the Gulf of Mexico coastal ocean observing system (GCOOS).

We appreciate the suggestion and investigated the GCOOS website but found no mesozooplankton grazing rate measurements were available. In the manuscript we use the incredibly extensive SEAMAP data set for validating cross-shelf patterns. We will acknowledge the need to have similar measurements on the shelf to the ones we collected in May 2017 and 2018.

The original Kishi et al model included a small, large and predatory zooplankton to represent ciliates, copepods (mesozooplankton), and euphausiids (macrozooplankton). Shropshire et al. redefined large and predatory zooplankton as two size-classes of mesozooplankton. This conceptual change is not indicated in Section 2.1.2. Do the zooplankton parameters in the model, like maximum growth rate and maximum grazing rate need to be revisited after this redefinition? I dislike that the model validation results in Section 3 can be dependent of the size-class arbitrary choose to define large and predatory zooplankton.

Originally PZ are defined in NEMURO as a group that includes gelatinous plankton and euphausiids. From the original Kishi manuscript “the biomass of the highest predator ZP is in a

sense unrealistic in that it represents the total biomass of a number of species. We included ZP in NEMURO to get a more accurate representation of the biomass of ZL, which plays an important role in lower trophic ecosystems in the Northern Pacific, as well as to represent a suitable prey functional group for the higher trophic level linkages.” We have treated LZ and PZ in much the same way, but in order to compare them to field data (and in a coming manuscript to the prey field for larval Bluefin tuna) we have chosen to consider Kishi’s LZ (mainly copepod) trophic group to be representative of mostly 0.2 – 1.0 mm organisms and Kishi’s PZ (gelatinous predators and krill) to be representative of mostly >1.0 mm organisms.