We thank the reviewer for their thorough and careful review. Our responses are below under the reviewers comments (italic text).

General Comments

(1) To rate the study, it would be important to know how large the impact of the resolution is compared to other uncertainties as e.g., the surface forcing, the parametrization of the mixed layer depth, the solar absorption profiles etc.. Also, preceding studies addressing similar issues should be mentioned.

Earlier studies have been added to the introductory text, detailing the sensitivity of biological models to various factors such as horizontal resolution, surface forcing and model parameters. The consensus is that physics dominates and hence mixed layer depth behavior is key when considering the vertical only, as in this study. Hill et al (2012), Burchard and Beckers (2004) and Hanert et al. (2006) all show how adaptive meshes can be used in conjunction with vertical turbulence models to improve a models accuracy at simulating mixed layer behaviour. As such, adaptive meshes should also increase the numerical accuracy of biological ecosystem models as the consensus is that the physical representation is key to increasing model skill. This study focuses on demonstrating adaptive meshes ability at improving numerical accuracy of ocean biology models and limits itself to numerical accuracy, not model skill and is the first such study to show this. Earlier adaptive mesh studies are described, but in the adaptivity section. They have now been brought forward into the introduction section.

(2) Generally biogeochemical models are tuned based on a certain resolution to fit some observations as good as possible. It can thus not be expected that a higher resolution will generally make the model look more realistic.

We think the reviewer is conflating numerical accuracy with model skill (how well the model fits data). We don't expect higher resolution to increase model skill – we use increased resolution to show numerical convergence in the model, then show adaptivity gives as good an answer as fixed mesh with fewer elements. We cannot therefore conclude adaptive meshes results in a better match to data, i.e. model skill, and at no point in the manuscript do we claim this. However, the model will represent the physics better with higher vertical resolution. The best possible model will be one that resolves all turbulent length scales without parameterisation. That is infeasible for large-scale models, but it has been shown by a number of authors that strategically changing resolution improves the physical representation of the system (see above). If you have improved the physical representation, then the biological model should behave more realistically within the limits of that model (i.e. depending on the model parameters). Oschlies and Garçon (1999) showed that large primary productivity can be a result of large numerical (i.e. non-realistic) diffusion and furthermore counteracting this effect of unrealistically large implicit diffusion by changes in the biological model could easily lead to misconceptions in the interpretation of ecosystem dynamics. Instead increasing resolution or implementing improved numerical schemes could reduce the artificial diffusion. Our results are entirely consistent with this narrative, in that you must have numerical confidence in the model before tuning parameters. We have clarified this point further in the discussion section of the paper.

Also, a good fit to the observations cannot be expected in this study since advective processes are neglected.

This is only true at those stations where advection dominates, which is not the case at Station Papa. We should also reiterate we are not claiming adaptive mesh technologies will necessarily increase model skill. We have developed this point further in the discussion.

The argumentation in the paper should be changed sometimes and probably idealized studies, illustrating sensitivities and effects, would give a better picture. Of course the stations could be still mentioned to illustrate that the idealized studies are based on realistic conditions.

We disagree that more idealised studies would be of use here. Moreover, one could argue we have used idealised studies with only pseudo-1D representations of the ocean sites being used. Idealised (i.e. laboratory-type) studies in combination with adaptive mesh technologies have been carried out in a number of previous papers and, whilst not specifically aimed at biogeochemical models, tracers have been part of the studies. A combined sensitivity study of biological tuning parameters, models, adaptivity metric weights, surface forcing and turbulence parameterisation would be a very useful study, but is beyond the scope of this present paper.

For which resolution and region was the model originally designed?

The ecosystem model was originally designed for the global model OCCAM (Popova et al 2006). We minimally tuned parameters to fit the data at all three stations presented here for use in Fluidity. This has been clarified in the text.

(3) Typically, 3D ocean circulation models feature a much higher resolution at the surface than at depth, while the present paper assumes equal spacing at all depth levels. Typical resolutions in 3-D ocean models should at least be mentioned or (better) be added.

We agree that this is an omission of this present study. We have therefore compared one of the key examples (tracking of detritus at Station Papa) with a similar model using a-priori variation in vertical resolution based on the popular NEMO model. We have added this result to the relevant section and clearly show the advantage that adaptive meshes can give in the representation of transient behavior.

(4) The whole section about method explanations is difficult to read/understand. To make it easier, the explanations should stick to the version actually used in the study and not other available options in the applied software package.

The explanation only applies to the methods used. We have described the GLS model in some detail as we have added an additional term to that described in Hill et al 2012. We have shortened the explanation to refer only to the k-epsilon version of the GLS.

The explanations are sometimes relatively vague, e.g. it does not get clear how the weights are chosen to optimize the adaptive grid,

We have given a thorough overview of how the mesh is adapted, along with a number of references that provide more detail. We have expanded the text on how the weights of this study were chosen.

which initial values were used for the biogeochemical model

All initial values have been added to the manuscript.

and if the model shows some initial drift.

The model shows no initial drift as shown in figures 2, 3 and 5.

The criterium how the MLD is defined is apparently missing (is it the same criterium for model and observations?)

This has been added to the manuscript. We use density differences to surface values (as per observations) when comparing to observations.

and also I did not find how "surface values" are defined.

The surface values are at those at zero depth, which are the same as the values through the mixed-layer depth. We have clarified this in the manuscript.

When introducing the biogeochemical model briefly in the text all prognostic variables should be mentioned and particular also how chlorophyll is treated (since many models assume a fixed Chlorophyll_a:P ratio). At the same time, some other explanatory parts could be shortened considerably

All prognostic variables are already listed, but this fact has been made clearer. Chlorophyll_a is treated as a separate variable.

(e.g. not all formula need to be given but citation is sufficient (or if they are given they should be explained and the notations should be used consistently throughout the manuscript); the first passage when explaining the biogeochemical model is non-necessary etc.).

We have considerably shortened the GLS explanation along with several other paragraphs.

(5) The authors state that the use of an adaptive grid could save considerable computing time since the number of necessary vertical levels can be strongly reduced while the results remain basically identical. How difficult is it to implement a varying number of vertical levels in 3D and how much would this again increase the computational cost?

This has already been explained as part of the response to the first reviewer.. We have added information on projected computational savings.

Specific Comments

ABSTRACT: In 1D most of the mentioned processes cannot be addressed. The abstract should stick more to what actually was done.

We disagree with this point as we are motivating the use of adaptive grids and clearly state this paper is a first step towards three dimensional simulations. We have however, reduced the level of discussion on three-dimensional studies.

INTRODUCTION Also the introduction could stick a bit more to the investigated issue. 3D ocean models using adaptive grids in the vertical are already used, e.g., at IOW Warnemuende. It would be good to mention and discuss preceding studies. What is new about the present study?

These papers are already cited in the main text, but have been brought up to the introduction. We have clarified the novelty of this study.

HYDRODYNAMICS MODEL I would recommend to change the title of this section (Methods?). Some parts are common knowledge and could be shortened, while sometimes important information is lacking (see major point 4). Particularly, the mesh optimization did in principle but not in detail get clear to me.

We have changed the section title and added more text on the adaptive metric formulation as part of review 1.

MODEL VERIFICATION I would rather call this "model evaluation" – however, as stated above I *don't think it makes much sense to measure the model-data misfit in the present setup and the section might be renamed anyway. Also, it is a bit lengthy to go through the same results for all three stations. Instead only remarkable situations could be highlighted.*

We disagree with the reviewer here. In order to properly show the ability of the adaptive mesh it is important to convince the reader of the ability of the model to a) have reasonable model skill which we do only by a qualitative comparison (i.e. we do not assess model skill quantitatively) and b) show numerical convergence. We feel the figures are an important part of this.

ADAPTIVITY The "Metric formulation" a well as the interpolation methods should be moved to the Method section (which is now called "hydrodynamical model").

Some text is moved to methods, but some of this text describes a key result – the effect of interpolation on the simulated system, for example, so has remained in the results section.

FIGURES: The number of figures is much too high while many figures show basically the same thing. Also, it would generally help a lot if the name of the respective variable would be written into the sub-panels (Fig. 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15). The symbols representing the observations are sometimes very small.

We feel it is vitally important that the reader is convinced of the adaptive mesh capabilities and the figures are a key part of this, despite their similarity. We have added the respective variable on the panels to aid interpretation. The symbols representing observations are deliberately small in order allow comparison to the model results but without swamping the numerical results. The reader is not meant to be able to decipher individual symbol values.