Referee comment on the manuscript "Optimality-Based Non-Redfield Plankton-Ecosystem Model (OPEM v1.0) in the UVic-ESCM 2.9. Part II: Sensitivity Analysis and Model Calibration", by Chia-Te Chien and colleagues.

8 March 2020.

1) General comments

This article presents the sensitivity analysis of an optimality-based plankton-ecosystem model (OPEM), implemented in the University of Victoria Earth-System Climate Model. This model is described in an accompanying paper submitted to GMD (available as a discussion paper; Pahlow et al. 2019), while this article focuses on the evaluation of model performances. Given a set of biogeochemical parameter ranges, 400 simulations were performed using a Latin-Hypercube sampling method. This ensemble of model solutions was then used to select the best model parameterisations using a likelihood-based cost function taking into account both temporal and spatial variations of observed NO_3^-, PO_4^{3-}, O_2, and chlorophyll *a* concentrations at different depth levels. Two biogeochemical models with two different formulations of the temperature dependency for diazotroph's growth were considered (OPEM and OPEM-H) but led to the same choice of best parameter values. However, these best solutions led to low N_2 fixation and denitrification at global scale, as these rates were poorly constrained by the data. Estimates of water-column denitrification were then used to identify the "best" model parameterisation within the ensemble of model solutions. A sensitivity analysis was also conducted for all biogeochemical tracers and all "optimized" parameters. The results revealed that the most important parameters for O_2 concentration were the remineralisation rate, the subsistence N quota of ordinary phytoplankton, and zooplankton maximum specific ingestion rate, underlying the central role of phytoplankton physiology and elemental stoichiometry in global nitrogen cycle in the ocean. From their results and sensitivity analyses, the authors propose new hypotheses on the link between NO_3^- concentrations at global scale and phytoplankton physiology. As a perspective for their work, the authors suggest to explicitly include benthic denitrification and atmospheric deposition in future biogeochemical models to better represent the global nitrogen cycle.

The Introduction section is easy to follow and clearly presents the context and the aim of the study.

In the Materials & Methods section, the authors briefly present the models (the biogeochemical model is fully described in Pahlow et al. 2019). They detail the sensitivity analysis and the model calibration method. The later is based on the definition of a likelihood-based cost function taking into account four different types of observations (monthly and/or annual climatologies of NO_3^-, PO_4^{3-}, O_2, and Chl *a* concentrations at various depth levels, averaged over 17 biogeochemical biomes). This original definition allows for taking into account spatial differences between biomes, as well as temporal differences at various depth levels, which is very good to fully constrain the spatial and temporal dynamics of the tracers.

The Results section then describes the ranges of global averages of major tracer concentrations (for both OPEM and OPEM-H), the sensitivity of biogeochemical tracer estimates (incl. phytoplankton biomass and stoichiometry) to model parameters, and a detailed description of the cost function values of the 400 simulations, especially for different global estimates of biogeochemical tracers.

The Discussion section focuses on 1) the sensitivity of key parameters (including remineralisation rate, phytoplankton subsistence nitrogen quota, and maximum specific ingestion rate of zooplankton), 2) the main model limitations (e.g., lack of benthic denitrification and atmospheric deposition may explain the high simulated volume of oxygen deficient zones compared to the observation), and 3) the insight provided by the use of their original cost function, especially on the usefulness of including variance information in this cost function and on the consideration of several biomes for its calculus.

The Conclusion section is clear. In this section, the main results are listed, the originality of the study is underlined, and the perspectives are indicated.

The text is well written and very clear. Most of the figures are properly described. Typo errors are extremely rare.

Given the very good scientific quality of the manuscript, **I recommend minor revisions before publication.** See below.

Review items

1) Does the paper address relevant scientific modelling questions within the scope of GMD? **YES**
Does the paper present a model, advances in modelling science, or a modelling protocol that is suitable for addressing relevant scientific questions within the scope of EGU? **YES**
2) Does the paper present novel concepts, ideas, tools, or data? **YES**
3) Does the paper represent a sufficiently substantial advance in modelling science? **YES**
4) Are the methods and assumptions valid and clearly outlined? **YES**
5) Are the results sufficient to support the interpretations and conclusions? **YES**
6) Is the description sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? *In the case of model description papers, it should in theory be possible for an independent scientist to construct a model that, while not necessarily numerically identical, will produce scientifically equivalent results. Model development papers should be similarly reproducible. For MIP and benchmarking papers, it should be possible for the protocol to be precisely reproduced for an independent model. Descriptions of numerical advances should be precisely reproducible*. **YES**
7) Do the authors give proper credit to related work and clearly indicate their own new/original contribution? **YES**
8) Does the title clearly reflect the contents of the paper? The model name and number should be included in papers that deal with only one model. **YES**
9) Does the abstract provide a concise and complete summary? **YES**
10) Is the overall presentation well structured and clear? **YES**
11) Is the language fluent and precise? **YES**
12) Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? **YES**
13) Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? **NO**
14) Are the number and quality of references appropriate? **YES**
15) Is the amount and quality of supplementary material appropriate? *For model description papers, authors are strongly encouraged to submit supplementary material containing the model code and a user manual. For development, technical, and benchmarking papers, the submission of code to perform calculations described in the text is strongly encouraged.* **YES**

2) Specific comments

Remarks on the Methods:

-Equations 1 and 2: why don't you also test the sensitivity of the model to the parameters of these two equations? Please justify it in the text.

- Table 1: More clearly indicate in the legend of Table 1 that the identified "best" values for trade-off simulations were the same for the two model configurations OPEM and OPEM-H. It is indicated in the text later, but it has to be clearly mentioned here for the reader.

- Lines 102-107: I am not entirely convinced by the arguments given here to justify why the parallel setup is better than systematic calibration approaches. Indeed, one can imagine a systematic calibration where X values are systematically tested for the 13 parameters. In that case, I do not see how this would lead to individual model simulations that would depend on other/previous combinations of parameters, neither how it would prevent re-evaluation with different metrics. However, the first item would be true for a parameterisation based microgenetic algorithm for instance. The authors may need to rephrase this sentence to make it accurate (e.g., by replacing the term "*systematic*" by an other one).

Remarks on the Results:

- Table 2: Could you add a column with observation values, at least for the depth levels where the data are available (and also add rows of simulated values at these different depth levels), so that the reader can also estimate if the observed concentrations/fluxes fall into the range of simulated values? Otherwise, I have the feeling that this table could be removed. As it is, is in unclear, even from lines 170-175, which main result(s) the reader should keep in mind from this Table.

- Positive comment: Figure 1 nicely shows that the model outputs are highly sensitive to nu_det and Q_0^Nphy (and to g_max and phy_phy at the second order)!

- Section 3.3.3 is quite long, but it presents a very detailed and interesting description and associated comments of the results presented in Figure 1. Keep it as it is.

- Figure 3: Justify in the methods the reason(s) of your choice of performing a regional splitting into latitudinal bands. This is missing in the article. Indeed, you mention later line 259: "*sensitivities of dissolved N:P ratio to parameters in [...] three geographical settings (low, high latitudes and global*)". It has to be mentioned (and justified) earlier.

- Line 250: "*where diazotrophs are abundant in high latitudes*": yet, this is not visible from your results. If this comes from Pahlow et al. 2019 please indicate it.

- Figure 2: Add a legend for black (OPEM) and grey (OPEM-H) as you did in Figure 1. Same comment for Figure 3.

- Line 285-288: It seems to me that these sentences should rather be included in the Methods section, not in the Results section.

- Figure 3: It may be nice to highlight the values that differ between OPEM and OPEM-H, for instance with rectangles (around the bars) and or stars (below or above the bars), so that it would clearly strike the eyes that the differences between the two configurations are obtained for the 60°S-70°S latitudinal band for C:N, C:P, and N:P. Also indicate in the legend the choices you made for the "*different latitude bands*".

- Figure 5 is not easy to read as it is because purple and black symbols look very similar. Smaller symbols may be used to help. Drawing horizontal and vertical lines to better underline the location of the WOA 2013 values (green square) may also be a good idea (although the figure is very well described lines 306-313).

- Lines 313-314: "*Overall, we stress that the minimum-cost and trade-off solutions appear at the margin of the full spread of the ensembles, which could be interpreted as indicating a model deficiency.*": I do not understand what you mean here. For me, it seems that they are in a patch of simulations with symbols in black, indicated log_10 of cost values lower than 8, which seems OK. What are you referring to by the term "*model deficiency*"?

- Line 315: "*Figures 6 and 7 show zonally averaged NO3 – and O2 in simulations with low and high NO3 – and the trade-off simulations":* Would it be possible to delineate these simulations in Figure 5? Indeed, it is unclear if the concentrations presented Figures 6 and 7 come from one simulation only, or from several (how many?) simulations. When describing these two figures, also underline the fact that the outputs from OPEM and OPEM-H are very similar here. If this is indeed the trade-off simulation (as indicated in the legend), then the results should be the same and there is no need to show twice the same figures.

- Line 332: "*because of intense denitrification in the ODZ*" => the last (and first) time that you used the abbreviation ODZ was line 193. As it has not been used since, I recommend giving the full name here again and not just the abbreviation (as you do it later line 402).

- Line 334: "*widespread ODZs, occupying much of the deep water in the northern and equatorial Pacific as well as the Indian Ocean (Figure 6)"* => Please indicate these areas clearly on the Figure 6, using arrows for instance.

- Figure 8: clearly mention in the legend that the two trade-off simulations for OPEM and OPEM-H are in fact the same, and use only one symbol for this trade-off simulation for figure clarity.

Remarks on the Discussion:

- The section 4.1.1. (especially the lines 348-362) provides new and very interesting hypotheses on the link between NO3 inventory at global scale and phytoplankton physiology. I appreciate this section.

- Line 404: "*ODZ volumes in the trade-off simulations are more than twice that in the WOA 2013 (Figure 10)*" => I do not see where it is visible on the Figure 10. I guess it could be inferred from Figure 10C from an expert eye, but I would rather give the precise value in the legend of Figure 10, with the corresponding vertical lines on Figure 10C, if you decide to keep the text as it is. Besides, this is the fist mention of Figure 10, that will be mentioned again line 439. I recommend clearly describing this figure here and later in the discussion, to fully explain and exploit it.

- Line 436-437: "*A peculiarity of our cost function is that it complements the data-model misfit, i.e. the residuals of spatial mean log- transformed values, with an additional term that resolves differences in spatial variances*" => Yes, indeed! I have particularly appreciated this.

- Line 439: "*The cost function's variance term introduces a strong penalty to approximately 30 % of all ensemble model solutions (Figure 10)."* => As mentioned above, Figure 10 lacks a clear description. I do not see what in Figure 10 supports this, but I am sure the authors could give more explanation for helping the reader through this.

Additional remarks:

- I am wondering why keeping the quarter of the 400 simulations with the highest (worst) cost values in all the analyses, and not keeping only the 200 to 300 best ones?

3) Technical corrections:

<u>Minor comments and typos:</u>
- Lines 44-46: "*Our new ecosystem model [...] offers new features and it improves the representation **of some biogeochemical tracers** on the global scale (see accompanying study, Pahlow et al. (2019)*" => Which biogeochemical tracers? Give examples in brackets.
- Lines 48-49: "*This model approach yields mass flux estimates with spatial and temporal variations in the elemental C:N:P stoichiometry of both inorganic nutrients and organic matter*." => Add at the end of this sentence: "*as observed in situ*" and give some references to justify (e.g. Martiny, A.C., Vrugt, J.A., Primeau, F.W., Lomas, M.W., 2013. Regional variation in the particulate organic carbon to nitrogen ratio in the surface ocean. *Global Biogeochem. Cycles* 27, 1-9.)
- Line 79: "*Our setup comprises ensembles of 400 simulations for each of two model configurations. The two model configurations differ in how temperature affects diazotrophy."* => This could be replaced by "*Our setup comprises ensembles of 400 simulations for each of the two model configurations that differ in how temperature affects diazotrophy*."
- Line 102: "*the parallel setup with different parameter combinations has a some advantages*" => Remove "*a*".

- Line 103: Replace "*Individual*" by "*individual*".
- Legend of Figure 4: Replace "*minmum-cost*" by "*minimum-cost*".
- Line 337: a space is missing after the term "*quota*".
- Line 360: "*our simulations: A more intense..."* replace "*A*" by "*a*".
- Line 378: You may want to change "*do contribute some variations to most of the tracers*" by *"do contribute to some variations of most of the tracers"*
- Line 393: Figure 5 instead of Fig. 5 (for homogeneity).
- Line 421: " *The mean global estimates ±1 standard deviation in OPEM and OPEM-H are*..."=> You may want to replace "*±1"* by "±*"*.
- Line 496: "*and*" instead of "*adn*"