The manuscript *Physiological flexibility of phytoplankton impacts modeled biomass and primary production across the North Pacific Ocean* by Y. Sasai and colleagues assess the importance of optimal nutrient uptake, photoacclimation and variable stoichiometry on the emergence of large scale chlorophyll and primary production patterns in the ocean, using the North Pacific as a testbed. To do that, the authors designed an experiment based on the comparison of two strategic NPZD biogeochemical (BGC) models coupled to a state of the art, eddy-resolving model of ocean physics (OFES2). One of the BGC models implements optimal uptake kinetics (Smith et al 2009), whereas the other incorporates interactive, optimal photoacclimation (Chl:C ratios) and variable stoichiometry (C:N deviating from Redfield ratios) as proposed in the context of the FlexPFT theory of Smith et al 2016). The experiments revealed clear deviations between predicitions from both models, with a clear gain in accuracy when using the more complex FlexPFT model (bulk properties like Chl and fluxes like PP closer to observations, better resolution of the deep chlorophyll maximum (DCM), etc.). The authors conclude recommending the adoption of similar approaches by the BGC community.

The topic, approach and results are of great appeal and the manuscript is already an important contribution. The authors designed a clean test of their hypothesis about the emergence of surface and vertical gradients in phytoplankton growth and biomass, and the result clearly support their ideas. However, the manuscript is not easy to follow as it stands, especially the combined *Results and Discussion* section. The narrative in this section is quite descriptive, and it fails to provide a clear picture of the ability of each model to reproduce observed patterns. As commented below, these issues seem to arise in part from the lack of motivation and rationale for model assessment in *Methods*. There are other aspects missing like the limitations of the current model (*e.g.* what about nitrogen fixation?), alternative explanations to the emergence of DCM (Cullen 2015; there may be more recent reviews), or about the performance of similar BGC models in simulating Chl and PP in the Pacific. Together, these issues led me recommend a major revision of the manuscript. I provide some major concerns, and a long list of minor issues and suggestions below. I hope the authors find both lists useful.

Major comments

1. The manuscript is in general well written and structured, but there are two sections that in my opinion deserve another thought, namely *Methods and Materials* and the *Results and Discussion* (see next three points). Although the description of the models is in general easy to follow (see however some minor suggestions below), the fact that the text moves from a complex model to a simpler one is not an optimal choice. I recommend the authors to present first their general approach with the components that are common to both models, and then detail first the simpler model featuring just optimal uptake followed by the more complex model featuring also photoacclimation and variable stoichiometry. I am aware that this might read as a minor issue, but I think it is quite important to ensure that readers realize that, despite their names, FlexPFT is something more complex than InFlexPFT. It is not clear whether one model is a nested version of the other or not (in the sense of a simpler formulation or the result of setting from variables as a constant). For instance, there is certain temptation to just check Table 1 and conclude that InFlexPFT results in reduced Chl and PP when compared with FlexPFT just because μ_{max} is lower in the former. There is also

some confusion about whether the model implements only photoacclimation or photoacclimation and variable stoichiometry, and about whether one or both of them are simple NPZDs or not.

2. Merging *Results* and *Discussion* has certain risks. In my opinion, that section of the manuscript needs to dig a bit more into the results and provide more quantitative tests that enable readers to assess the relative merits of each model and to frame the results in the context of similar work. The text reads well, but it lacks any quantitative comparison except toward the end, when the authors comment on the huge variability of available NPP estimates and provide large scale estimates for the overall production of the North Pacific. The manuscript would benefit from a more systematic assessment featuring regional averages (say, at the biome scale?) and some kind of statistical metric or test.

3. There may be other things to say about the choice of the data for the comparison, and about how model output was preprocessed (for instance, how did you process Chl profiles, was there any attempt to mimic the way the ocean color satellites operate?). Since the simulations were forced using JRA-do reanalysis data, one would expect that the target for the models would be to reproduce or match available data. It is not clear what was the aim and objectives of the study, and perhaps that explanation is the only thing missing. The objectives and the rationale for choosing some data and patterns over potential alternatives needs to be justified. The models seems to be doing more than decent job, but the authors need to clarify to what extent some of the apparent biases observed both in surface and subsurface fields reflect are due to biases in simulated physical and chemical conditions or to differences on the phytoplankton model.

4. Finally, a key aspect that the authors need to make clear earlier in the paper is the feasibility that proposed and discussed mechanisms may be actually working in the field. There is room to discuss alternative mechanisms currently ignored by the two models assessed. For instance, interactions between grazers and phytoplankters, potential biases in export and recycling, the metabolic diversity of phytoplankton (*e.g.* nitrogen fixers, picophytoplankton), etc.

5. As a bonus question, although it does not seem central to the study at hand, the formulation of zooplankton grazing was quite intriguing for two reasons the deserve further comment;

    1. the numerical response seems to be nonstandard and deserves further comment, as well as the closure term
    2. Eq (A2) in L429 includes a quadratic mortality term for phytoplankton. That effectively means that phytoplankton dynamics follow logistic growth, which seems redundant with the formulation of phytoplankton growth as a function of available nutrients, and underwater light and temperature conditions, and with the common assumption of a population controlled though grazing by zooplankton. Perhaps I am missing something?

Minor comments

Abstract

L001 - active voice? [Light and nutrient conditions ...]

L002 - define photoacclimation?

L002 - at the end your model features both photoacclimation and variable stoichiometry, perhaps it is worth highlihgting it

L003 - break the sentence at the comme (it is already a bit twisted), and perhaps join with the next one?

L004 - as commented above, I recommend to go from simple to complex [say optimum nutrient uptake PFT to full FlexPFT], and provide a one sentence description with general details about the two models

L007 - mention OFES2 by name [and acronym]?

L010 - [...] subsurface Chl maxima *in the subtropical gyre* [to provide context]. Otherwise please detail where exactly that happens (especially the overestimation of Chl). As commented above, a figure detailing the magnitude of deviations with satellite data would be very useful.

As a general comment about the abstract; data used for validation is not mentioned at all. Readers might just assume you are testing your model against "oceanic observations" [L003], which may be too vague

---

Introduction

L035 - not sure if there is something else besides pursuing efficiency and simplicity 😉 ...

L077 - if InFexlPFT is a typical NPZD, then call it NPZD, or state here too that it is an NPZD implementing optimal uptake kinetics as per Smith et al 2009 [L135ff]?

L080 - perhaps deter giving the full name and details of OFES2 to MatMet?

Methods and Materials

L085 - I understand you extensively modified the simple NPZD to implement either OU kinetics or FlexPFT. I mean, perhaps it is worth mention it and state that the default configuration consisted in a simple NPZD? [or maybe just provide those details later when talking about the BGC component of the model]

L089 - I would put all details about the configuration of the experiments together [L101]. The sentence in L93 is especially intriguing and disconnected from the rest of the explanation [L104].

L113 - Eq (1): perhaps Q(I,T) instead of just Q?

L119 - Eq (2): I really did not like the symbol $\mu_{Flex}(I,T)$, it seems potentially confusing ... what about just $\mu_{\mathrm{max}} , S(I,T) , F(T)$ or $\mu_{Flex}^{*}(I,T)$ [where I would suggest the former] ... otherwise it may lead users to think you need some functional or to iteratively solve the equation?

L124 - Please detail how do you determine maximum affinity [if it is optimized on a daily basis, etc]

L131 - please detail how the optimal value of θ is updated [I mean, that it is not a constant]

L137 - ideally, it would be nice to see how one can go from Eq (7) to Eq (1) [if that is possible], Otherwise it may be worth stating whether the models are truly nested or they just feature different terms for nutrient the dependent growth [though they propagate to the other terms in FlexPFT]

L153ff - perhaps explicitly include formulas for Chl and PP [$Chl = P , \theta / Q \quad \text{in FlexPFT}$, etc]

L155 - I also miss some details here about what kind of outputs were compared to observations. In principle, since JRA-do is a renaalysis dataset, you might expect a direct match between simulated fields and observations at sea.

L165 - perhaps follow the order physics, chemistry, biology? Again, missing some rationale for the choices and the way the model was evaluated

---

Results and Discussion

-> *general comment* as an author myself I can understand the preference for pooling the two sections, as a reader I am not such a fan.

L174ff - it seems that the physical component of the model was evaluated elsewhere; if that is the case it would be better to explicitly state so, but it would be ok to go beyond the ability of the model to reproduce major circulation features to mention at least its skill in reproducing temperature and nutrient fields.

L185 - there is certain paradox here since the initial focus of the manuscript on phytoplankton biomass and productivity mutates here on a large section devoted to two sections devoted to chlorophyll (which, needless to say in the context of a photoacclimation paper, is not biomass)

L185 and 200 - perhaps the titles of these sections should somehow detail that they refer to model to data intercomparisons (w.r.t. the section starting on L321)

L190 - please provide some quantitative summary of deviations between models and obs

L200ff - the narrative here becomes a bit difficult to follow to me ... perhaps an alternative structure [grouping results per biome for instance], or just a diagram or table summarizing the main findings [obs, simulated patters, most important regulating factor, etc] might make the section easier to follow

L277 - why? Is it possible to partition the amount of variation due to each to I, N and T?

L327 - horizontally and vertically? Could you develop a bit more what you mean?

L346ff - I think this paragraph belongs to the previous section

---

Conclusions

L375 - perhaps *acclimation* instead of *adaptive response*?

L409 - perahps hte IA abbreviation can be omitted here to detail instantaneous acclimation

---

Appendix A

L429 - please note comment above about Eq (A2)

L435 - please detail the type of numerical response (*i.e.* no need to force interested readers to go to Sasai et al 2016). My excursion to that paper suggest it is a Gompertz function with a threshold ... did not seem entirely standard (I mean, a commonly used formulation).

Figures

Figure 1 - nice maps! Some suggestions doubts to comment in teh text;

- add transect lines to panels c and k?
- what happened in the Gulf of Alaska and at Bering Sea?
- the distinct shape of the gyre suggest there may be underlying biases in ocean physics propagating to chl [what about simulated MLD?]

Figures 2 and 3 - again nice figures and amazing results

- physics, chemistry, biology? [order of columns]
- why not directly comparing data for 2002/2003? [it would be nice to check whether the model reproduces small scale heterogeneity]

Figures 6 and 7

- I like the figures but still feel they fail to clearly convey whether N and I are more important than T ... How would an equivalent figure with T in the abscissa look like? How can you partition which variable contributes more variability?

Figure 8

I think the results for FlexPFT would compare well with satellite based NPP products. Indeed, it would be great if, beyond biases in InFlexPFT the authors can show that actually the simpler model fails to capture large scale gradients [or at least, that is the impression I got].

Figure 9

Is it possible to complement these profiles with a time series plot? [perhaps the monthly climatology at each site]