

We would like to thank the reviewer for both their overall positive evaluation and the important points they raised. We agree that addressing these will further improve the quality of the paper.

Please find below each reviewer comment in blue and italic and our response in black, standard font.

In their major comments the reviewer clearly highlights two key points:

1. The uncertainty in the primary production measurement methods, leading to high uncertainties in (*in situ*) observations of primary production, e.g. *“First and foremost, in my opinion, the single largest challenge in improving estimates of primary production and its changes is the difficulty of reliably measuring primary production. Any attempt to use data assimilation or machine learning tools to understand productivity patterns requires reliable observations, or at least observations with a reliable error model to extract those insights from.”*

Although we did mention uncertainties in *in situ* observations of primary production at several points in the manuscript, where relevant to the main focus, we would like to follow the reviewer’s suggestion and highlight these points better. To do this we will dedicate an additional paragraph to the potential uncertainties in *in situ* data early on in the manuscript (in the Background section around the line 199), to make clear that the accuracy of *in situ* data can vary with consistency in methods and the specific method used (i.e., measuring GPP, NPP or other forms) and that this could also lead to errors in the comparison with satellite-based or ecosystem-model-based estimates of primary production. We will also add comments on how the *in situ* primary production data could become more useful, e.g. by developing reliable error models. We will also address all the minor comments of the reviewer towards this point -- wherever this remains appropriate given the intended modifications to the manuscript. This may include: a) making a comment on this issue right in the Abstract, around the line 39, b) adding comments on this early into the Introduction, i.e. into the paragraph on the lines 90-109, c) adding a comment on this into the satellite-based model section 3.1 (paragraph 322-327), d) adding another comment on this into the paragraph 632-667 of the “A way forward” section.

2. The reviewer suggests to better highlight the potential synergies between the different types of models (especially across satellite-based and ecosystem models), in particular how they can be combined, or learn from each other, for their mutual benefit. E.g. *“My second observation relates to the potential synergies between satellite NPP models and ecosystem models that estimate NPP. The paper outlines many differences and commonalities between these two types of models. I think it would benefit, however, from a more concise distillation of the basic differences and pathways for productive interaction that leverage their relative strengths.”*

We fully agree that exploring model synergies — especially by increasing model consistency through model inter-calibration exercises (e.g., estimating parameter variability in less complex models using simulations from more complex models) — should be discussed more prominently in our manuscript. This includes using ecosystem model simulations to estimate parameter variability in some satellite-based models. We also agree that this point would follow more naturally after better highlighting the differences between satellite-based and ecosystem models in Section 3.3.

To be more concrete (and also following reviewer minor comments) we suggest to make this topic more central by: a) add sentences on this issue right into the Abstract around the lines 50-53. b) Add comments on this in the Background section into the paragraph on the lines 228-245. c) Update the section 3.3 as appropriate and highlight the satellite-ecosystem model synergies at the end of this section. d) Discuss this point more in detail in the “A way forward” section, by substantially expanding the paragraph 626-631 and also briefly mentioning it in the following paragraph, starting from the line 632. e) Mention it again in the Conclusions section, in the paragraph on the lines 696-705.

Furthermore, please see our suggestions regarding the reviewer minor comments:

*Minor comments:*

1. *Line 39-40; 45-47: Consider this statement relative to my first primary comment. It is important to acknowledge that the in-situ data is also highly uncertain.*

Thank you, we will do as appropriate (please see our response to reviewer's major point 1).

2. *Line 48-52: The concrete and achievable synergy between ecosystem and satellite-based models (lines 626-631) highlighted in my second primary comment seems central to these points. I would consider a clearer and more prominent articulation of this synergistic path forward.*

Thank you, we will do (please see our response to reviewer's major point 2).

3. *Line 71-73: Is the purpose here only to review, or do you intend to suggest a path forward?*

Thank you, we will address this (it is both).

4. *Line 90: Perhaps reframing this paragraph to address my first general comment is: "A central challenge in understanding primary production and predicting primary production is that it is difficult to measure", then consolidating these issues here? You would then just need to articulate how these challenges could be overcome.*

Thank you, we will do as appropriate (please see our response to reviewer's major point 1).

5. *Line 97-98: Most ecosystem models predict and report NPP, not GPP. Respiration is often included in the growth function (e.g., Geider), basal respiration is either omitted or accounted for in reported NPP.*

Thank you, we will address this.

6. *Line 132-133: I would argue that not including T-dependent parameters for a global algorithm, or even a large regional one across seasons, is not defensible given the data. Can't we at least eliminate this?*

We agree with the reviewer that such approaches are problematic, but we would like to try to refrain from commenting on specific approaches that are being compared. If possible, we prefer to remain sufficiently general in both, reviewing the state-of-the-art PP models, and suggesting the way forward.

7. *Line 156: I would call the 17-83 range the range, not the uncertainty. The lower and upper bounds can clearly be rejected with existing constraints. There seems to be an unwillingness to do so directly, but maybe the effort you are calling for could help us do so in an objective and constructive fashion.*

Thank you, we will address the first comment. As we mentioned previously, whilst we agree with the reviewer that the estimates are unlikely accurate, we would avoid outright rejecting some of these estimates in this paper.

8. *Line 207-213: See general comment 1. The difficulty of observing NPP deserves a more central, consolidated place in your narrative and needs to be more directly addressed in your call to action.*

Thank you, we will do as appropriate (please see our response to reviewer's major point 1)..

9. *Line 232-234; 238-241: Is there enough data to do this, or would an ecosystem-/satellite-based model synergy be the primary driver of this process? Would intermediate-/high-complexity ecosystem models be an intended target of such activity, or would it be a tool helping to understand and introduce this added structure into other models?*

We believe that both are necessary: calibration against the same observational data sets and, where possible, inter-calibration among models. Please note that when we refer to the availability of large observational data sets for model calibration, we mean data covering a broad range of biogeochemical variables (chlorophyll-a, oxygen, pH, nutrients, etc.). These variables are generally better constrained than the PP measurements mentioned by the reviewer. Therefore, stating that sufficient data may exist for a highly constrained model calibration is, in our view, fully consistent with the acknowledged need to improve the accuracy of *in situ* PP measurements.

10. *Line 322-327: See general comment 1.*

Thank you, we will address this as appropriate (please see our response to reviewer's major point 1).

11. *Line 347-349: Not always ignored; Sometimes treated implicitly.*

Thank you, we said here "often ignored" which we believe was accurate enough?

12. *Line 404-405: Do you think it is defensible to not have a temperature dependence?*

Please see our response to the reviewer comment 6.

13. *Line 434-440: This seems key to the synergy between these two model types, and comes up again on lines 626-631. It would be nice to have a more cohesive presentation of these points (see general comments).*

*Emergent Patterns: The satellite-based models that do not have explicit nutrient and temperature dependencies, inherently contain those dependencies in the model parameter values (line 434-435); BGC models have these dependences and allow you to probe emergent outcomes.*

Thank you, we will do (please see our response to reviewer's major point 2).

14. *Line 460: See general comments.*

Thank you, we will address this (please see our response to reviewer's major point 2).

15. *Fig. 6A: See Stock (2019) for an example of how to derive the equivalent photosynthesis-irradiance parameter from an ecosystem model.*

Thank you, we will include the reference.

16. *Line 535-537: This flexibility allows models to account for the diversity of plankton and processes that are not explicitly represented in current models (BGC models capture quite a bit, but I could see a data-driven approach capture more. The only problem is that we don't have reliable measures of NPP to find emergent patterns).*

Whenever we discuss calibration of parameters, we consider also other measurements than PP, please see our response to the reviewer's comment 9. But we agree that the uncertainty of PP measurements needs to be better discussed in the paper and we suggest to do this as proposed in our response to reviewer's major point 1.

17. *Line 585: "Together, these considerations suggest that investigating parameter assignment and parameter variability may be an important route to understand and potentially reduce many of the apparent*

*differences between marine PP models, and hence in the estimated magnitudes of production.” – I feel like the recommendations that follow could be more concrete, and worry a bit that this statement comes nearly 600 lines into the paper.*

We propose to make such key points better visible earlier on, e.g. in the Abstract/Introduction/Background.

18. *Line 628-631: This seems like a concrete path forward yet the treatment is very limited (see general comments).*

Thank you, as already suggested, we will address this.

19. *Line 634-636: An considerable uncertainty in in-situ NPP observations?*

Thank you, we will address this as appropriate (please see our response to reviewer’s major point 1).

20. *Line 655: New paragraph?*

Yes, thank you.

21. *Line 656-675, Fig. 10: Yes, data volume has increased rapidly, but has it yielded reliable, consistent NPP estimates across global ocean biomes? If this hasn’t happened yet, is the time really right? Is there a way to address the fact that it hasn’t in a rigorous way using new techniques to provide a self-consistent data source for much of the analysis you’ve proposed. I feel like this section should have come earlier in the text as one of the central challenges to executing what you are proposing, rather than just before the conclusion.*

The uncertainty of in situ observations will be discussed much more thoroughly (see our response to reviewer’s major point 1), and we also refer the reviewer to our responses to their comments 6 and 12. We emphasize that we are referring to *highly constrained* estimation, so we are being very cautious in our statements. We will make this point clearer in the dedicated paragraphs, such as on the lines 602–625.

22. *Line 690: I’m not sure what you mean by “assignment is quite rare”.*

We will explain this, thanks.

23. *Line 693-695: Would the natural next sentence be: A combination of data-driven approaches and natural synergies with intermediate- and high-complexity ecosystem models could provide the means to fruitfully uncover these variations.”*

Thank you, we will address this reviewer’s point across the manuscript, as mentioned in our response to reviewer’s major point 2.