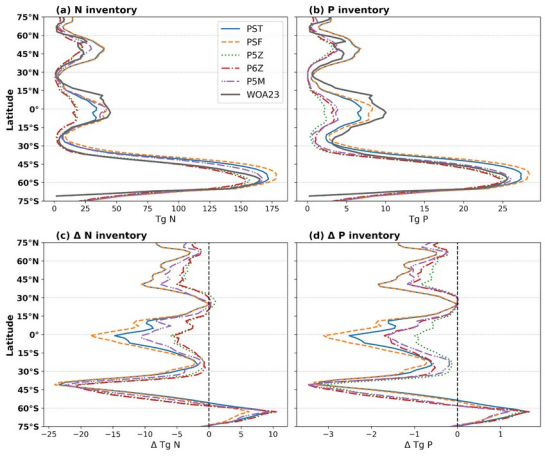
Author response

My co-authors and I would like to thank John Dunne (reviewer 1) and Shengwei Liu (reviewer 2) for taking the time to review our manuscript titled *"Complexity in Biogeochemical Models: Consequences for the Biological Carbon Pump."* The reviewers' comments have significantly improved the scientific quality of the manuscript by strengthening aspects of the methodology, refining the discussion, and identifying typographical and grammatical errors. We respond to the reviewers' comments below, referring to reviewers by number and then following it up with their comment and our response. When page (pg.) or line (ln.) numbers are referenced, they correspond to that of the original submission. We use italics to denote quotations from our manuscript and combine with underlining to highlight additions/changes.

Note, we copy most of our responses from the online EGUsphere author response interactive discussion section.

Number	Reviewers comments	Author response
1	Reviewer 1: ln. 54 - "the present" is constantly changing. Better to use "2023"	Yes, would agree that giving a date or year is more suitable than simply the word 'present' in adding to the long-term readability of the work. We update '*present'* to '*<u>2023'</u>*
2	Reviewer 1: ln. 162 - it would be helpful to add which configurations were used for CMIP5 and CMIP6 given their prominence in previous climate change intercomparisons as point of reference.	For this comment I am not exactly sure how to respond. In our methodology, we state that all the PISCES configurations were forced with identical physical outputs from the IPSL_CM5A_LR climate model and beyond the historical run, we used the high emissions RCP8.5 scenario. So we are using the scenarios that were used within CMIP5.
3	Reviewer 1: ln. 192 - It would be helpful to add a sentence on the impact of adding Mn and Zn modulation of the ecosystem.	In ln. 185 – 190, we give a brief description of the P5M configuration and state that: "*…incorporating manganese (Mn), following Hawco et al. (2022), which included its role in limiting phytoplankton productivity in the Southern Ocean, where observations show Mn as either a primary or co-limiting micronutrient alongside Fe…*". So in brief, Mn and Zn addition impacts phytoplankton growth by imposing an additional nutrient limitation term, alongside the other micro- and macronutrients already represented.
4		Ah, thank you for pointing this out. In our supplementary we show the biome spatial

	Reviewer 1: Figure 1: The acronym definitions for each region should be provided.	map with all the regions labelled. I will amend the caption then to the following: *Figure 1: Model and remote-sensing (RMT) estimates of (a) NPP and (b) C_{exp}, integrated over each RECCAP2 biome (refer to Fig. S1). Black bars indicate ±1 standard deviation across the remote-sensing ensemble.* The biome map was placed in the supplementary material solely to limit the number of figures in the main manuscript. The update now gives clarity and a reference for the reader when we refer to biomes going forward in the manuscript.
5	Reviewer 1: Figure 2: it would be helpful to add the Model and Remote Sensing means and range estimates here. Also, the caption should provide the time ranges used for both the model averages and delta values.	Your suggestion of adding the remote sensing (RMT) means is something we considered and I, as well as my fellow co-authors, can definitely see the value in doing so. However, we opted for a per biome breakdown when comparing NPP and C_{exp} of PISCES and RMT to champion the fact that the modifications to some of the configurations were highly regionally specific: e.g. diazotrophs in the subtropical gyres in P6Z or Mn and Zn impacts in the Southern Ocean for P5M. As for the caption and the time ranges, we do state in the methodology: *For this study, we conducted our analysis of the BCP using averaged model outputs over two time windows, the 'reference' (1986-2005) and 'future' (2091-2100).* We rigidly stick to this nomenclature throughout the manuscript and thus implicitly have the date ranges accounted for in the various tables and plots. Am sure personal preferences prevail, but hope this justification suffices.
6	Reviewer 1: One of the major comments concerned the importance and understanding the differences in model representation of surface nitrate and phosphate.	We address this comment by creating the following figure and adding it to the supplementary section:

		 We also add the following description of the figure in the supplementary: *Differences in biogeochemical parameterisations of phytoplankton growth processes across the five PISCES configurations result in some intramodel variability in N and P inventories. In Fig. (S5a, b), the Monod-quota models resolve slightly higher global nutrient inventories than the Quota-based configurations for the reference period but both modelling frameworks exhibit consistent spatial patterns; especially when compared against observational data from the WOA23. Similar future declines in nutrient inventories occur across configurations (-11.78±1.85% for N and -12.39±1.28% for P).* Furthermore, we add the following within the manuscript: *This top-down interpretation is further supported by the broadly consistent spatial patterns and projected changes in N and P inventories across configurations (Fig. S5).* The additions address the reviewer's comment and fundamentally showcase that N/P variability are not likely a driving factor of C_{exp} variability in our study. By explicitly stating and demonstrating this point, it gives greater credibility to our top-down conclusions that are explored in the manuscript.
7		We agree with the reviewers' comments and thus make the following edit:

	Reviewer 1 & 2: Both reviewers requested a refinement on our discussion when contrasting the variability in NPP and C_{exp} captured in our PISCES-only ensemble with that of the broader CMIP (CMIP5) ensemble.	*These findings align with previous modelling studies (Bindoff et al., 2019), and fall within the variability of the CMIP5 ensemble (Bopp et al., 2013; Fu et al., 2016). The comparable magnitude of variability in NPP and Cexp (Tab. 2) indicates that differences in parameterisations among the selected PISCES configurations can generate a spread in results of similar order to that found across CMIP5 models (Séférian et al., 2020). While this does not imply that the present ensemble captures the full diversity of CMIP model structural differences, it nevertheless highlights the sensitivity of biogeochemical outputs to relatively subtle differences in the representation of phytoplankton growth processes and ecosystem complexity which do contribute to intermodel variability.* The edits remove strong language claims and reframe our results on this topic as suggestive rather than demonstrative. We also slightly amend the wording in the conclusion to reflect this new nuance.
8	Reviewer 1: ln. 502-503 the reviewer noticed that our statement "*the overall decline in carbon export reduces the ocean's capacity to sequester CO2 from the atmosphere.*" seemed reversed in logic. The reviewer then suggested a useful paper to reference and cite to improve this.	We agree with the reviewer's comments and have addressed them by making the following change: *Global future reductions in NPP and C_{exp} likely reflect reduced nutrient supply to the surface associated with enhanced stratification and increased interior residence times (Dunne, 2023); however, the e-ratio shows little to no significant change for the five PISCES configurations, indicating that BCP efficiency is maintained relative to the reference period.* We now stress that shifts in NPP and C_{exp} are consequences and not drivers. We include the following reference as well: Dunne, J.P.: Physical mechanisms driving enhanced carbon sequestration by the biological pump under climate warming. Global Biogeochem. Cy., 37(11), p.e2023GB007859, https://doi.org/10.1029/2023GB007859, 2023.

9	Reviewer 2: A major comment concerned the justification of the time-windows we used, especially when comparing our reference model period (averaged over 1986-2005) with remote-sensing products (averaged over 1998-2005).	We agree that a justification is needed. Therefore, in the methodology, we add the following: *Although the averaging periods differ between the model outputs and remote-sensing products, the use of multi-year means reduces the influence of interannual variability and allows for a consistent comparison of large-scale patterns.* Limitations in our ability to harmonise time-windows are explained in detail in the online response under AC2 response on the EGUsphere interactive discussion (https://doi.org/10.5194/egusphere-2025-6505-AC2)
10	Reviewer 2: The reviewer asked if there was a way to justify or perhaps remove NPP and C_{exp} remote-sensing algorithms on the outset to better refine our comparison of the model outputs vs remote-sensing products.	In short, we do not exclude any algorithms a priori. Instead, we consider the spread across algorithms as a representation of structural uncertainty, and interpret the results in terms of ensemble behaviour rather than reliance on any single product.
11	Reviewer 2: the reviewer requested a refinement and addition to our brief explanation on the statistical analysis of Supp. (S8).	For greater explanation of the results present in Supp. (S8), I would agree with the reviewer. We did not want Supp. (S8) to be a major focus but we like the reviewer's contribution. Within the discussion, we therefore add the following, as suggested by the reviewer: *Increased complexity may improve global NPP magnitude without improving, and possibly worsening, spatial pattern fidelity, while for C_{exp} there is little difference in skill across configurations* And indeed, the added statement is more precise than the current broad statement we have regarding increasing complexity and model skill.
12	\Reviewer 2: With reference to comment #11, the reviewer asked why these statistical differences existed between the different PISCES versions we used in our study.	Answering why these improvements/worsening in NPP occur is beyond the scope of the paper. But it is important to understand that Supp. (S8) is the ensemble mean of all the NPP and C_{exp} remote-sensing products used in the study.

		Referring back to Fig. (1) showing the range in remote-sensing for the different regions, the magnitude is captured well but we cannot infer spatial patterns. To do so would require us to individually assess the remote-sensing products against the different model configurations. This in itself would be an interesting study, and granted, would be a very valuable question to answer.
13	Reviewer 2: "How would other key BGC metrics like alkalinity (involved in e.g. nitrogen cycling and soft-tissue pump) be impacted in your model configurations? Will they be more helpful for the model-observation comparison?"	Alkalinity was not explored in the work, but acknowledge that it could have been an additional facet to the study.
14	Typographical errors noted by Reviewers 1 & 2 are corrected	*Refer to mark-up*
15	Reviewer 1 & 2: Both reviewers asked or questioned why we chose to compare our PISCES-ensemble with that of CMIP5 and not CMIP6, noting of course that the version of PISCES we used (Aumont *et al,* 2015) was used in CMIP6.	We chose to compare our results with CMIP5 because the experimental framework used in this study is consistent with CMIP5 protocols, in particular the use of IPSL-CM5A-LR forcing and the RCP8.5 scenario. While the PISCES v2 biogeochemical model is also used within CMIP6 models, our simulations are not directly based on a CMIP6 experimental configuration (e.g. SSP-based forcing and associated coupled model setups). For this reason, CMIP5 provides the most consistent and directly comparable ensemble for contextualising our results. We note that CMIP6 results and developments are discussed in the Introduction to provide broader context on recent model developments and findings related to NPP and C_{exp}.
16	Personal: referencing method for supplementary figures is updated.	In the original submission we used Supp. (Sx) to designate a certain figure or table to reference in the supplementary. However, our supplementary includes both tables and figures. Thus, we use Fig. (Sx) and Tab. (Sx) when referencing our supplementary material for greater clarity.

References

AC1 response, *EGUsphere interactive discussion*, https://doi.org/10.5194/egusphere-2025-6505-AC1, 2026.

AC2 response, *EGUsphere interactive discussion,* https://doi.org/10.5194/egusphere-2025-6505-AC2*,* 2026.