**Response to Reviewers:**

Reviewer comments in black, author responses in blue.

**Reviewer 1:**

General comments: This paper synthesizes our current knowledge on projections of Southern Ocean productivity and biogeochemistry, complemented by some analyses of existing CMIP6 model projections for the high-emission scenario SSP5-8.5. While the topic is generally within the scope of Biogeosciences, I have some concerns regarding the overall scientific advancements and the presentation of the manuscript in its current form. Therefore, I recommend major revisions before this article can be considered for publication for the reasons outlined in more detail in the following. After reading the paper, I was left wondering what the specific aim of this paper is and what its scientific advances are, e.g., relative to two reviews published in the last years on changing Southern Ocean phytoplankton dynamics (Deppler & Davidson, 2017, doi: 10.3389/fmars.2017.00040) and biogeochemistry (Henley et al., 2020, doi: 10.3389/fmars.2020.00581). Unfortunately, as it is, the knowledge gain from this manuscript seems rather small. Something that sets the manuscript by Fisher et al. apart from the aforementioned reviews (and the reason why this paper ultimately should be published) is the inclusion of analyses of CMIP6 data. This is generally valuable and useful, but unfortunately, in its current form, the analyses are not presented (enough) as a central part of the manuscript. The results from these analyses are mixed up with reviews of existing literature, with knowledge gaps resulting from the literature review not being tied closely enough to the performed analyses and the shortcomings of the assessed models. To that aim, in my view, a revised version of the manuscript would benefit from a more "traditional" manuscript structure, with a clear knowledge gap/goal of the paper identified in the introduction, a method section with details on which models are assessed, and dedicated results and discussion sections. I appreciate the extensive work that went into the assessment of existing literature (and I am not arguing to leave any of that out). I think the authors should just work on linking the model analyses more clearly to it, which requires diving deeper into the processes represented (or not represented) in the CMIP6 models in the main text of the paper. In the abstract, the authors set out to "identify key processes where greater observational data coverage can help to improve future model performance". While I agree with the authors that this is a key step forward, it is not entirely clear to me how the presented description and assessment of the CMIP6 models in the current version of the manuscript contribute to that goal. In the revision process, the authors should therefore focus on making it much clearer to the reader what this paper adds to existing literature.

We thank the reviewer for their comments and appreciate the fact they value the analysis of CMIP6 data. In the revised manuscript we have strengthened this aspect by expanding on the CMIP data sets (addition of historic pH, temperature and future projections of Si*). Additionally we have restructured the manuscript to emphasise the methods used and models selected, while streamlining the discussion of literature, particularly in relation to physical processes. Our overall aim in this paper is to show regional variations in the CMIP biological and biogeochemical parameters, which can be used to target observations. In addition, we identify species composition as being an area where uncertainties exist on the order of sign changes in regions of the Southern Ocean, we wish to emphasise this as being a potential source of further uncertainty in the biogeochemical parameter and a topic requiring an expansion in observational and experimental data to support the development of future model generations. In the revised manuscript we propose listing our specific recommendations in the conclusion. We believe the rearrangement of the methods, additional model outputs we include in this response, streamlining of the literature review and explicit listing of recommendations will give a clearer focus to the manuscript and alleviates some of the concerns raised about the focus and purpose of our analyses. Please see our responses to the specific comments below:

Specific comments:

L. 24: I don't understand what such an evidence-based sampling approach would be, given that there is substantial disagreement between the models. Unless I have missed it, you don't mention more about this elsewhere in the manuscript – can you elaborate?

Replaced with "a sampling approach targeting the regions with the greatest rates of climate-driven change in ocean biogeochemistry and community assemblages".

L. 27: A useful reference to add in this section and throughout the paper is Deppeler and Davidson 2017, as these authors have synthesized our understanding of climate change impacts on Southern Ocean phytoplankton in a lot of detail.

Citation to Deppeler and Davidson (2017) added.

L. 28: It is unclear to me why you say that biological carbon uptake is increasingly important (see also a few lines further down where you it plays a minor role in oceanic CO2 uptake). I suggest to simply say "an important process".

Changed to "an important process".

L. 30-32: I think the inserted "with the Southern Ocean […]" breaks the flow of this sentence. Combining this piece with the information in the sentence before and then switching the order of the sentences (first global, then Southern Ocean) might result in a better flow.

Restructured to: "Across the global ocean, uptake of carbon accounts for ~25% of $CO_2$ released by human activities (Friedlingstein et al., 2022). The Southern Ocean is a disproportionately large carbon and heat sink relative to its size (Frölicher et al., 2015), accounting for 30-40% of this global anthropogenic $CO_2$ uptake (e.g. Caldeira and Duffy, 2000, DeVries, 2014), predominantly due to enhanced atmosphere-ocean exchange at increased atmospheric CO2 concentrations (Friedlingstein et al., 2022)."

L. 46: I suggest deleting "especially". I don't disagree per se, but from what you write it is not clear (yet) why this part of carbon cycling should be "especially" vulnerable.

Removed "especially".

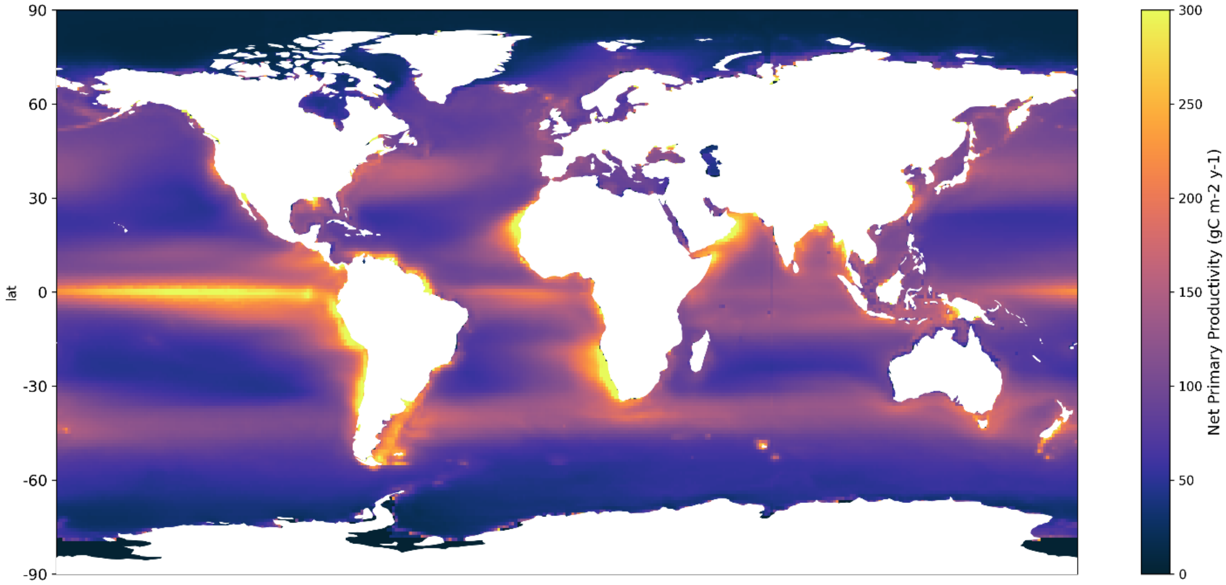
L. 49: Given that this NPP estimate is based on the years 2002-2007, I am wondering here why you're not reporting a more up-to-date estimate. Also, what definition of the Southern Ocean are you using?

We have performed our own additional analysis of NPP now to update Okin et al., (2011). This update is based on an average between 2015 and 2023, from a multi-model ensemble following SSP2-4.5 (as the most realistic climate scenario for recent years, with $CO_2$ emissions tracking current levels). Note, we chose SSP2-4.5 over historical models because the historical period ends in 2015 so would itself be quite out of date. This analysis showed a Southern Ocean carbon production of 12.06 Pg C $yr^{-1}$, similar to the 11.4 Pg C $yr^{-1}$ proposed in Okin et al. (2011). However, global declines in productivity increased the relative contribution of the Southern Ocean from the ~25% previously cited to 35.96% in our analysis.

For this analysis, we defined the Southern Ocean as south of 30°, based on Gregg et al. (2003). This is the same boundary used by Okin et al. (2011) in their analysis of NPP.

The text has been revised to:

"Between 2015 and 2023, Southern Ocean phytoplankton represented ~35.96% of marine net primary productivity globally, equivalent to 12.06 Pg C $yr^{-1}$ (Supplementary Figure 1)".




Supplementary Figure S1: Global Net Primary Productivity (2015-2023) based on a CMIP6 multi-model mean for SSP2-45. Models included in this analysis are given in Table 1 as variable "intpp (245)".

L. 50: I suggest adding Laufkötter et al. (2015) to the discussion, as the authors describe projected changes in primary productivity in CMIP5 models.

We now include a new projection of global NPP in CMIP6 (above). We have added this citation later in the text in response to a further comment below.

L. 53: Please rephrase "sea ice duration" and "increasing ocean surface area". In addition, maybe rather say "light availability" than "light supply"?

Rephrased to: "a reduced duration and extent of sea ice coverage, allowing for a greater supply of light to surface waters".

L. 54: I find this statement a little misleading. Given that the increase in upwelling is predominantly found in open ocean waters, the link to sedimentary iron supply is not obvious. From what I know, sedimentary iron supply does not play a huge role in deep, open ocean regions, as the upwelling does not necessarily bring waters from the seafloor to the upper ocean. Can you clarify this?

This statement mainly refers to coastal examples of upwelling of sedimentary iron input (as observed on the West Antarctic Peninsula, e.g. Annett et al. (2015)). However, wind driven upwelling has the potential to deliver iron of hydrothermal or sedimentary origin (via lateral transport beneath the mixed layer) to surface waters in deep ocean open waters Moreau et al. (2023).

We have clarified this statement to indicate that upwelling of iron is not necessarily sedimentary, but can also be hydrothermal.

“Strengthened upwelling is also likely to increase the flux of existing iron supplies to the coastal (Annett et al., 2015) and open ocean (Moreau et al., 2023) from sedimentary or hydrothermal sources”

L. 62: Please check here and throughout the text: Species names are usually given in italic.

Species names have now been italicised throughout

L. 75: globally?

Added “to the global ocean”.

L. 76: I suggest adding Nissen et al. (2021) here, as the authors describe the impact of changed Southern Ocean nutrient utilization on nutrient fluxes to lower latitudes.

Added Nissen et al. (2021).

L. 78: What do you mean by “regional nutrient supplies” here? Please clarify.

Replaced with “Southern Ocean” (the region intended).

L. 82/83: Please note that ocean-only (hindcast) models are used in the Global Carbon Budget, not Earth System Models. I suggest adding a reference to the latest IPCC report here, as Earth System Models are used there.

We have added a reference to: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I. and Huang, M., 2021. Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change.

L. 84: Please specify in what exactly the spread of models has increased. In all variables?

Added: “the spread of model projections across vertical and horizontal physics as well as the number of phytoplankton functional types”.

L. 88: How do you know if a model projection is deficient? Maybe “differences” is more appropriate here?

Changed to “variance”.

L. 90: In addition to the factors you list here, I think the representation of phytoplankton loss processes, and zooplankton and particle dynamics also impact differences in model projections of biogeochemistry (as discussed in Henson et al. 2022).

Added: “Deficiencies in model projections of phytoplankton and ocean biogeochemistry have been linked to the use of fixed C:N:P elemental stoichiometry (Kwiatkowski et al., 2018), an inability to reflect physiological adaptations, e.g. the ability of diatoms to maintain growth under iron limitation (Person et al., 2018), **and complexities in modelling export fluxes, particularly in constraining phytoplankton losses through zooplankton grazing (Henson et al., 2022). “**

L. 96: I am not sure that Laufkötter et al. (2018) is the most suitable reference here. Please double-check.

Removed, Holm-Hansen et al (2004) remains as the reference for this point instead.

L. 99: I am left somewhat confused after this first section. I expected a knowledge gap being highlighted that will be filled in this paper. Is the purpose of your paper to synthesize current knowledge or to provide new insights based on new assessment of CMIP6 output? I think making this very clear early in the paper will help the reader. I think a better management of the reader's expectation will improve the paper.

This is intended primarily as a synthesis paper, with additional insights on future changes in phytoplankton dynamics derived from CMIP6 models. As CMIP6 arguably makes incremental progress on CMIP5 in the areas we focus on (Seferian et al., 2020), so the scientific knowledge gained from updated these predictions is perhaps small minimal. However, our paper addresses this by aiming to find processes or regions where large uncertainties exist, our overall aim is not to critique the models per-se but to draw attention to where additional biological and biogeochemical observations can be most valuable in improving the next generation of CMIP. Additionally, in summarising many of the biological and biogeochemical parameters, we are providing the most up-to-date model output in order to to underpin our understanding of future change.

L. 105: Can you clarify what you mean by "microscale" here?

This sentence is no longer part of the revised physical oceanography section.

L .108: Based on the title of the section, I was not expecting the discussion of any biology here. I suggest rethinking the title or reorganize the content of this section.

Section title changed to "Physical climate drives biological changes in Southern Ocean water masses".

L. 110: What CMIP6 models are shown here? How many models are you considering in this ensemble? I suggest to not expect the reader to go to the supplement for this information and to instead include this in the main paper.

Supplementary table 1 outlines the models selected for each parameter, this has been moved to the methods section of the main text (now Table 1).

L. 112: Since the importance of the limitation factors over time is not shown in the Figure (only the annual mean is given), from what do you infer the "before"? This is unclear from what you write and show.

Text revised to:

"CMIP6 models project the greatest increase in productivity to occur across the coastal zone of the Southern Ocean (65-90°S) (Figure 2c), where irradiance limitation is reduced (Figure 2d) due to a shallowing of the mixed layer (Figure 2b), and due to increased light delivery in the Weddell Sea (Figure 2e), despite experiencing stronger iron limitation (Figure 2f)".

L. 116/117: It is unclear what changes to the forcing affect iron supply. Can you clarify?

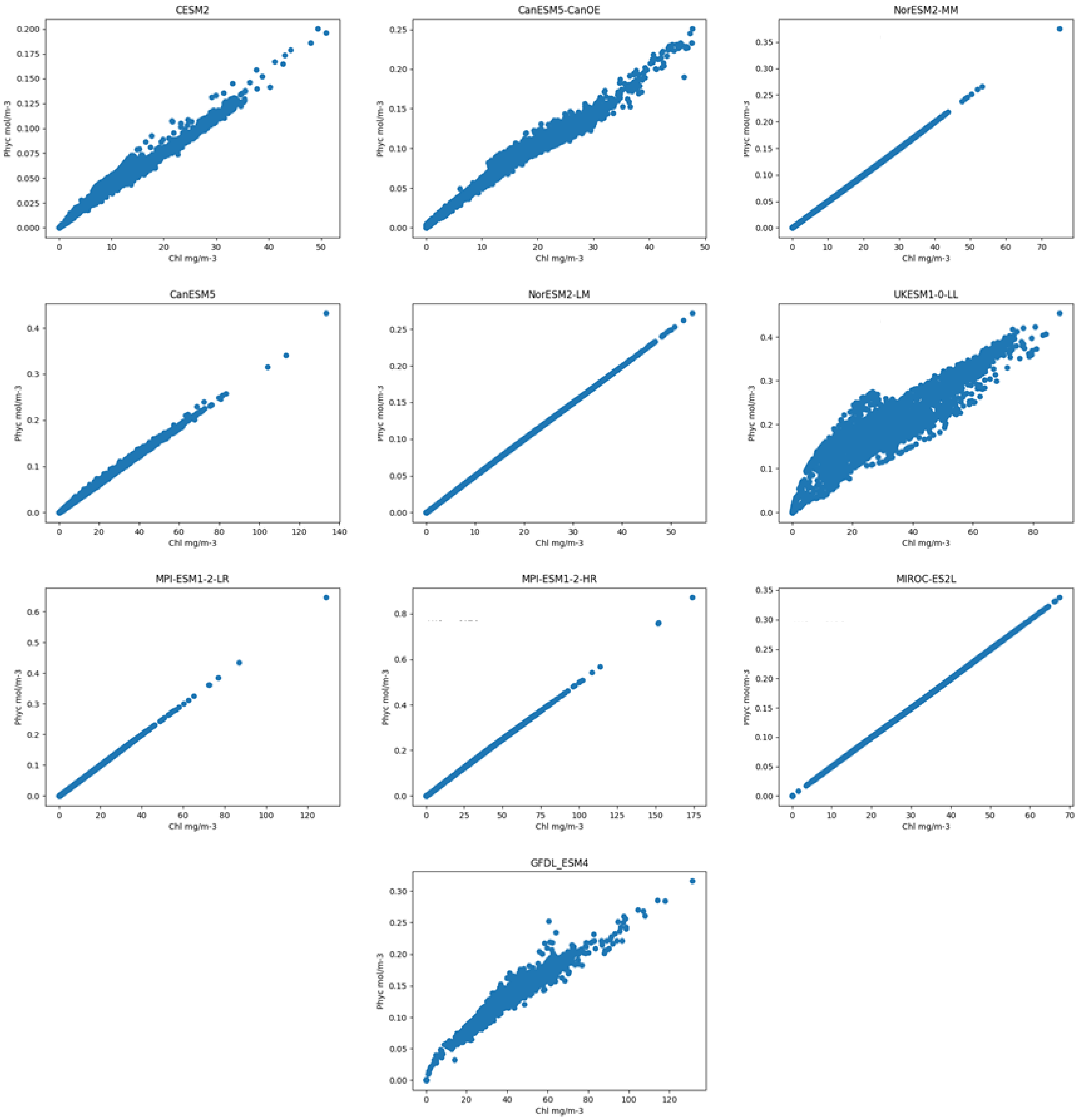
Clarified to: "Iron supply to the surface is subject to changes in the properties and movement of water masses, which lead to variable circulation strengths, depth boundaries, heat content and carbon sequestration resulting from climate-driven perturbations to the ice-ocean-atmosphere system (Meredith et al., 2019, Bindoff et al., 2019)."

L. 123: Please check the Figure reference – are you referring to Fig. 2a?

2d changed to 2a.

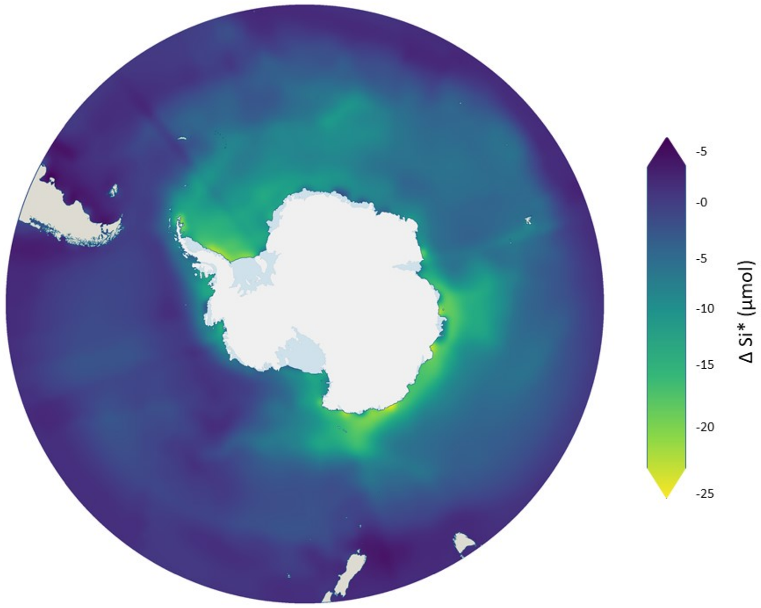
L. 123-148: In my view, this part does currently not fit very well into the narrative of the paper. Don't get me wrong, it is a nice summary of the physical changes. However, there is currently too few ties to what the paper is about according to its title: the impact of changes in phytoplankton productivity and community structure on Southern Ocean biogeochemistry. I suggest revising this whole section to fit the aim of the paper.

The physical oceanography section will be shortened in the revised manuscript to reinforce only those processes which are directly related to the biogeochemical parameters for which we show model outputs.



The original text between line 203 and 206 has been revised to:

"Si* is projected to decline across the Southern Ocean, with the greatest declines being in Ross and Weddell seas, as well as the Indian sector (Table 2). These changes in Si* correlate with increases in chlorophyll concentration across the same regions (Figure 3), suggesting that increase uptake of silicate is driving increased chlorophyll production. However, increases in chlorophyll appear independent from projected changes in primary productivity (Figure 2C). For example, the west Antarctic Peninsula and Amundsen sea regions show the greatest increase in primary productivity, but are among the lowest regions of change for both Si* and chlorophyll. The divergence between chlorophyll and primary productivity indicates variability in chlorophyll:carbon, with silicate consuming diatoms typically expressing more chlorophyll per unit of carbon (Sathyendranath et al., 2009). Therefore, the large chlorophyll increase, large Si* decline signal projected in the Weddell Sea is likely driven by an increase in diatoms, whereas the productivity increase with only small changes in both chlorophyll and Si* seen on the west Antarctic Peninsula likely results from an expansion of non-diatom phytoplankton with a lower chlorophyll:carbon."




**Supplementary Figure S6: Anomaly in Si* ($[Si(OH)_4]–[NO_3^-]$) (μmol) between 2100 (SSP5-8.5) and a historical average (1985-2015).** Representative of a multi-model ensemble of CMIP6 models, models included are detailed in Table 1.

L. 212: This is another good location to refer to Nissen et al. (2021), in which the authors have performed sensitivity experiments of the importance of dissolution/remineralization depth of nutrient resupply to the surface and nutrient distributions across the Southern Ocean.

Added Nissen et al. (2021).

L. 216: Unclear to the reader why "especially in low sea ice years".

We know that sea ice itself can store nutrients, e.g. *(Jones et al. (2023)).* As sea ice cover modulates light availability, and therefore phytoplankton nutrient uptake, it is an important control on nutrient dynamics. The point here is that low sea ice years are increasingly common but have been sparse over the period in which we have conducted much oceanographic sampling in the Southern Ocean.

Changed to: "particularly in increasingly common low sea ice years".

L. 220: "surface" pH. Add "under the high-emission scenario SSP5-8.5".

Added "surface" and "under the high emission scenario (SSP5-8.5)"

L. 225: Start a new sentence before "for example…".

Sentence split

L. 227: The wording is unclear to me here for "Petrou et al. (2019)'s modelled OA conditions".

Changed to "the OA conditions modelled by Petrou et al. (2019)"

L. 247: While I agree with you that there is lots of uncertainty surrounding the implementation of CO2 dependencies into marine ecosystem models, I disagree with them that their implementation is not feasible. Acknowledging that there are still gaps in our understanding, there have, in fact, been several attempts to implement these dependencies into both ocean only and climate models (e.g., Dutkiewicz et al., 2015; Krumhardt et al., 2019; Seifert et al., 2022 and references therein). I suggest rephrasing this part to reflect the state of the art more accurately.

We have clarified the phrasing here. We did not intend to suggest that $CO_2$ dependencies cannot be implemented in models, rather that there are still fundamental unknowns in the biological adaptation of phytoplankton to acidification.

New wording: "Interspecies divergence in responses to OA, and differential responses at different $pCO_2$ thresholds presents a significant challenge to modelling OA-induced ecophysiologial behaviour. Improving our ability to model these biological responses at the basin scale will require expanded observation of Southern Ocean $CO_2$ , through programs such as the Southern Ocean $CO_2$ Atlas (SOCAT) (Bakker et al., 2014), alongside experimental work to determine the response of different phytoplankton types to different acidification conditions in combination with multiple stressors."

L. 265/Table 1: It is unclear to me why you only report SI* from WOA and not from the models. Do the models agree with this spatial distribution suggested by WOA data? It would then be great to also report Si* from the projections. Further, I suggest providing the historical average for pH and temperature to put the reported changes into perspective. Is the reported standard deviation across models, across time, or across space? Please clarify. Please also consider adding Freeman et al. (2018) to the respective paragraph in the text for completeness, who have looked at the projected changes in Si* in the Southern Ocean in CESM.

We have now added the projected change in Si* (to 2100, SSP5-8.5, per our other analyses) from CMIP6 in addition to the historical data from WOA. Further, we have added historical data for temperature (WOA) and pH (historical run of CMIP6 (1985-2105) as pH is not included in WOA).

The standard deviation is across space, this has been added to the figure caption.

Added: "Freeman et al. (2018) show that increased biological uptake of silicate, through increased diatom growth, leads to a poleward shift in the silicate front and a potential decoupling from the Antarctic polar front."

**Table 2: Biogeochemical parameters of the Southern Ocean Observing System regions. SOOS regional working groups (as defined at: www.soos.aq/activities/rwg) indicated on Figure 3; section C is an overlap section of sections B and D. Data shown are: Si* ($[Si(OH)_4]–[NO_3^-]$) values and temperature determined from objectively analysed annual means of World Ocean Atlas 2018 data. pH was determined from a historical run of a multi-model ensemble of CMIP6 models (1985-2015). Delta values are anomalies of multi-model means of pH, temperature and Si* based on comparisons between the mean annual historical value (1985-2015) and projected values for 2100 under SSP5-8.5 for a CMIP6 ensemble (detailed in Table 1). Values in brackets are standard deviations, representing spatial variation across the region.**

| Section | SOOS Region | Si* (µmol) | ΔSi*(µmol) | pH | Δ pH | Temperature (°C) | Δ Temperature (°C) |
|---|---|---|---|---|---|---|---|
| A | West Antarctic Peninsula & Scotia Arc | **17.24 (17.82)** | **-4.32 (3.48)** | **8.06 (0.12)** | **-0.45 (0.01)** | **0.93 (1.73)** | **1.99 (0.40)** |
| B | Weddell Sea & Dronning Maud Land (WSDML) | **37.37 (9.70)** | **-10.02 (2.98)** | **8.08 (0.10)** | **-0.43 (0.01)** | **-1.12 (0.52)** | **1.61 (0.49)** |
| C | SOIS/WSDML | **23.16 (6.67)** | **-7.12 (1.24)** | **8.09 (0.02)** | **-0.44 (0.01)** | **-0.40 (0.85)** | **2.09 (0.37)** |
| D | Southern Ocean Indian Sector (SOIS) | **4.71 (3.72)** | **-8.45 (3.71)** | **8.05 (0.25)** | **-0.44 (0.01)** | **0.41 (1.42)** | **1.99 (0.52)** |
| E | Ross Sea | **19.82 (18.49)** | **-6.68 (4.77)** | **8.08 (0.15)** | **-0.42 (0.02)** | **-0.65 (1.64)** | **1.16 (0.47)** |
| F | Amundsen and Bellingshausen Seas | **17.59 (14.02)** | **-2.49 (3.82)** | **8.09 (0.13)** | **-0.45 (0.01)** | **-0.77 (0.78)** | **1.97 (0.32)** |

L. 274: It might just be me, but in my mind, "POC" is most commonly used in the literature for sinking particulate organic carbon and not for organic matter in phytoplankton. I would appreciate if you could clarify the terminology here to be more consistent with existing literature. Since you're referring to primary productivity here, why not use (N)PP?

We use POC because the model definition of the parameter examined (intpp) is: "Vertically integrated total primary (organic carbon) production by phytoplankton". However, this is the same as NPP. To include both, we have reworded this to "Across CMIP6 models, net primary productivity (NPP) is expressed by the intpp parameter, defined as the depth-integrated production of particulate organic carbon (POC)."

L. 275: Maybe "quantities" rather than "parameters"? (see also title)

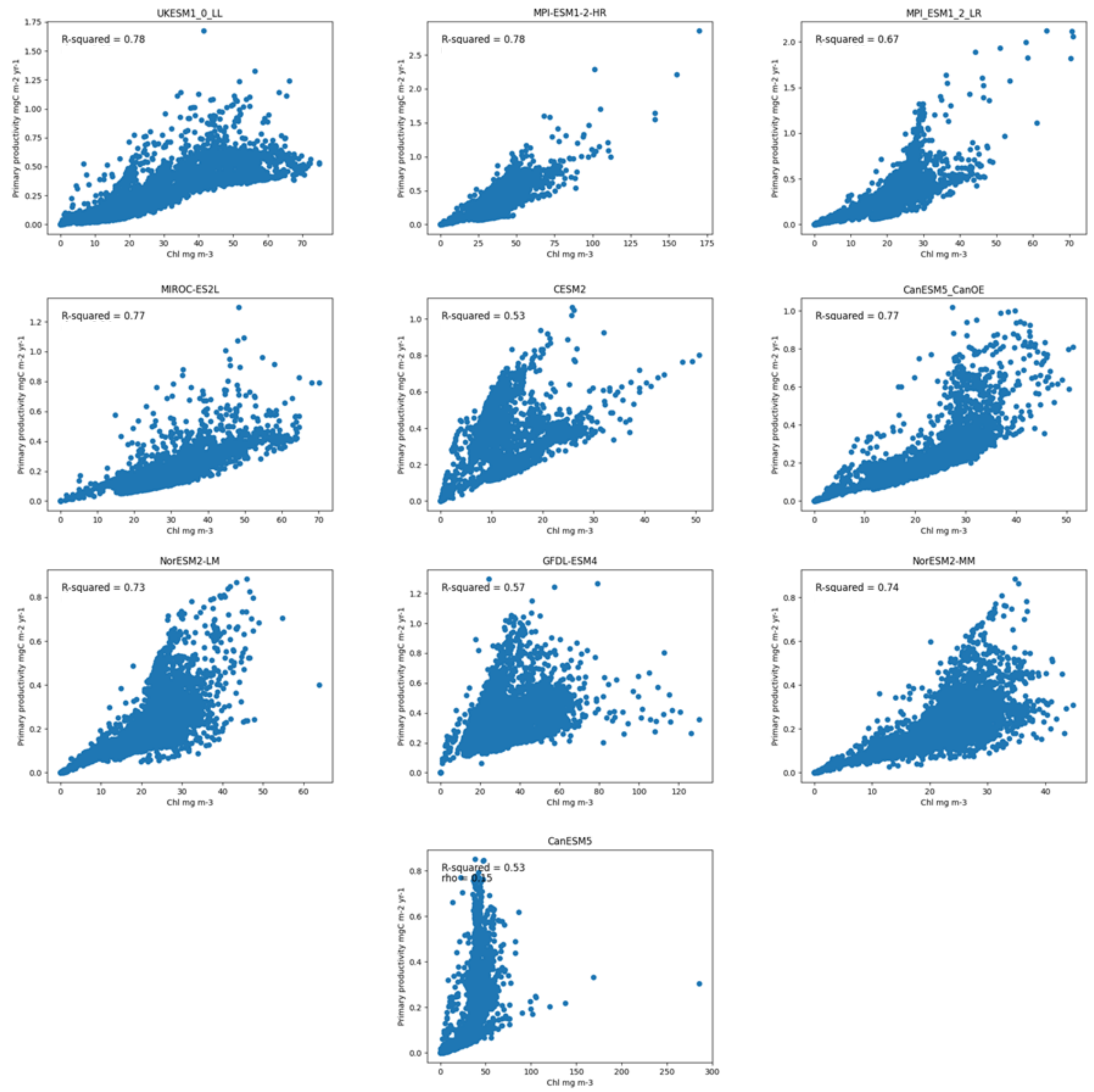
Deleted parameters on line 275, subtitle changed from parameters to variables.

L. 275: Do all considered CMIP models parametrize Chl:C variability in the same/a similar way?

We previously showed the correlation between Chl: phytoplankton carbon (in response to comments on L191). In this context there is no variability in Chl:phyC, all models employ a fixed ratio. This differs from reality where Chl:C is variable, particularly on a seasonal cycle in the Southern Ocean (Thomalla et al., 2017).

The reviewer comment here refers to an introduction of the chlorophyll and primary productivity parameters, therefore, we plotted Chl:primary productivity (below). Beyond differences in scale, we see a general positive correlation but with much greater spread and less linearity than for phytoplankton carbon content. This suggests that models are parametrising Chl:C variability in a similar way because of the similarity in output trends, but the lack of direct relationship between primary productivity and chlorophyll results in variance between the models. In our previous discussion on chlorophyll vs primary productivity in the context of Si* we attributed the difference between primary productivity and chlorophyll to phytoplankton functional groups, as diatoms have a greater chl:C than smaller plankton classes.

Note: For the purpose of this correlation, we plot a random 1% subsample of grid values south of 65 degrees latitude; values shown are a time average from 2100 under SSP5-8.5.




L. 280: This addition of remote sensing might confuse the reader in this paragraph about the models.

The mention of remote sensing has been removed.

L. 287: I think you’re introducing the different phytoplankton types way too late. You’ve already been mentioning diatoms and other types throughout the text. Please build your paper in a logical way, so that things are properly introduced when first mentioned.

The section introducing phytoplankton types have been moved to the introduction section of the manuscript.

L. 291: Nissen et al. (2018) is not a good reference for this point. I suggest citing Luo et al. (2012) instead.

Reference changed to Luo et al. (2012).

L. 293: The citation of Nissen et al. (2018) should be moved here.

Added Nissen et al. (2018)

L. 293: I don't understand what you mean by "90 distinct models".

This sentence has been removed in the revised text.

L. 295: If you're referring to all CMIP models that include 2 phytoplankton function types, isn't the second type typically a mixed phytoplankton, i.e., combining different size classes and not only picophytoplankton? Please double-check.

There are 2 different variables here, intpppico defined as: Vertically integrated primary (organic carbon) production by the picophytoplankton component alone. And intppmisc "Vertically integrated total primary (organic carbon) production by other phytoplankton components alone". The section referred to reads *"…11 [models] specifically include diatoms under future warming conditions and only three of these additionally consider picophytoplankton"*. That is to say that 11 models contain the intppdiat parameter, and of these 11, 3 have a specific picophytoplankton component (GFDL-ESM4, CESM2-WACCM, CESM2). These 3 models then also contain parameters for diazotrophs and miscellaneous phytoplankton (the mixed component the reviewer refers to). The other 8 models that contain diatoms only have 2 functional types (diatoms and miscellaneous).

We have brought supplementary table 1 from the original manuscript in to the main text; this outlines which models were chosen for which parameters.

L. 296ff: If these two models are so different in their future projections, I am wondering which one of the two simulates a more realistic community composition for the present-day. Have you looked into that? Additionally, it is unclear, based on what you write alone, why you decided to show these two models in particular.

We have not compared the model output to a real world composition. This is difficult because most of the phytoplankton data which exist for the Southern Ocean are spatially and temporally constrained too much to be properly representative.

These two models were chosen because they are the only two models to include irradiance limitation of picophytoplankton; this has been explicitly added to the methods section.

L. 301: closing parenthesis missing

Closing parenthesis added.

L. 301ff: I am not sure I follow your argument here. I assume that the two models you're showing here are already different in their present-day community composition. If one is more realistic than the other, there might not be a fundamental problem regarding the parametrization of phytoplankton functional types across models, but instead only in some of the models. Additionally, models account for phytoplankton-zooplankton already. While I agree with you that more work needs to be done on that end, a more nuanced statement might be necessary here to better reflect where the shortcomings of current models are.

The divergence in productivity projections is not directly linked to parametrization of functional types, but here we simply demonstrate that that divergence is not removed by further parametrization. I.e. adding more functional groups to increasingly complex models is not necessarily a solution to directionality uncertainties.

We have altered this statement to add nuance "expanded range of biological interactions **involved in** phytoplankton-zooplankton predation and bacterially-driven mixotrophic effects, **which can substantially alter trophic energy transfer and export fluxes (Ward and Follows, 2016)**"

L. 314: delete "ocean"

Deleted ocean.

L. 329: This sentence about seasonal progression appears disconnected from the topic sentence of this paragraph. Please revise.

This paragraph has been removed entirely.

L. 348: Reading through the paper, this section title confused me, as you have referred to CMIP6 projections throughout the paper already. I suggest revising the section structure. Additionally, please check for redundancies throughout the manuscript. I suggest including Laufkötter et al. (2015) in this section.

Section title changed to "Latitudinal productivity projections in CMIP6" to be more specific. We have removed some examples of repetition from earlier in the manuscript.

We added a reference to the findings in Laufkötter et al. (2015) at the start of this section. Previous analysis by Laufkötter et al. (2015) using a different set of models (a mix of marine ecosystem models employed in CMIP5 and the Marine Ecosystem Model Intercomparison Project) found substantial disagreement between models in projecting which phytoplankton groups drove NPP changes in the Southern Ocean.

L. 384: How much do you think this statement would change if CMIP models included more diversity for the high-latitude phytoplankton community? I am not convinced that this is a robust statement, given all the uncertainties surrounding CMIP-type projections in high-latitude waters.

Yes, this is true that if the models contained more ecological diversity then the trends in community composition and group level carbon production could be different. However, this statement is reflecting the current state of the CMIP ensemble, not an overall prediction of what might happen.

L. 396ff: Please list some of the processes you're referring to here.

Added (e.g. nutrient upwelling, viral losses, composition of the grazer community)

L. 404: Isn't the lack of eddies to a first degree a result of the rather coarse model resolution?

Mesoscale eddies are parametrised in CMIP6 but are not typically resolved south of 50°S, even in high resolution iterations (1/4°). This is because the Rossby radius, which sets the length scale at which rotational effects are as significant as buoyancy, has a value of ~250km at the equator but reduces to ~10km in the high latitudes (Hewitt et al., 2020). The main limiting factor in applying mesoscale eddies appears to be computational cost, while it is true that even the most high resolution models would struggle to replicate sub-mesoscale processes. Even if we will struggle to include the eddies themselves in high latitude projections, this comment (calling for more observation of eddies) is mostly focused on addressing a knowledge gap, understanding the temporal and spatial extent of eddy effects on phytoplankton. Although the eddy may be too small to resolve, the extent of the impact of the eddy may be relevant at existing model resolutions.

“eddies” changed to “biogeochemical impacts of eddies” for clarity.

L. 408: Please rephrase “diversity in fate of phytoplankton species”. It sounds odd to me.

Changed to “the extent to which different phytoplankton species undergo losses via different pathways”

L. 409: Replace “required for improving” by “which needs to be filled to improve”

Phrasing replaced.

L. 414: It very unusual to have a method section as the last section of a manuscript. As you could see from my detailed comments, there was a few locations where I was looking for information on the methods, suggesting that the current placement of this section might not be the best (especially since this section is not referenced anywhere in the text, implying that the reader does not even know it exists). I suggest moving the method section to a more “traditional” location in the revision process. Further, please include information on how many and which models were included in the main text.

The methods section has been moved to after the introduction, prior to any results in the main text. Additional methods information including models chosen has been added to this methods section from the supplementary information in the original manuscript.

L. 419: “quantity” instead of “parameter”

Changed to quantity

L. 421: I don’t understand what these “compatibility issues” are. Can you clarify? As it is, I don’t find this very convincing, as data structures can be homogenized in postprocessing, and I am left wondering why some models were (seemingly randomly) excluded from the analysis.

Most of our analysis is showing relative change over the 21$^{st}$ century by calculating the anomaly between SSP5-8.5 and the historical period. For some models they might contain either the SSP5-8.5 value or the historical value but not both; in these cases the models are deemed not compatible with the analysis and removed. It is not an issue with postprocessing, just availability of the parameters within the different models. We have removed the term “compatibility issues” in the methods.

Fig. 1: I really appreciate that you included a sketch! Just three minor comments: Please add “phytoplankton” before “size class” (otherwise it seems like it’s referring to bloom size classes, which is not what you mean I think). Further, the arrows related to upwelling should not (all) come from the seafloor. In my view, the way it is currently drawn is not an accurate representation of what is happening in the open ocean. Note also that the cited paper for the bottom warming only refers to warming of bottom waters on Antarctic continental shelves (Purich & England 2021).

This schematic figure will be updated in the revised manuscript.

Fig. 1 and Supplement: Note that ideally, there should be no references in the supplementary material that are not also cited in the main text. At the moment, Purich & England (2021) and McNeil et al. (2008) are only cited in the supplement. I suggest adding the references of the numbers given in Fig. 1 to the Figure caption in the main text.

References added to figure caption in the main text.

Fig. 4: Why is iron limitation only shown for the GFDL model?

The limfe parameter is only present in the GFDL model; this has now been stated explicitly in the revised methods section.

Fig. 5: I find it a bit difficult to identify the individual models for the different shown quantities. Maybe by showing the model spread as a shading around each multi-model mean this figure would be easier to read.

This cosmetic change to the figure will be applied in the revised manuscript.

Reviewer 2:

**General comments:**

This study provides a synthesis of previous works in regard to the Southern Ocean under climate change from the aspects of physics, biogeochemistry (including nutrients and ocean acidification), and phytoplankton. On top of the synthesis of previous works, this study is trying to investigate the underlying mechanisms that drive the change in phytoplankton productivity and analyse the potential phytoplankton structure change at different latitudes under the SSP5-8.5 scenario using CMIP6 models.

The paper is trying to address the question of how future physical and biogeochemical changes will affect phytoplankton, which is a significant question and fits the scope of the journal. And I can see the authors went through tremendous effort to conduct a comprehensive literature review. However, the current formatting of the paper appears to lean more towards a review paper rather than a traditional research article. For the part of the CMIP6 ensemble analysis, despite the general research idea being interesting and valuable, the results are very little and not well presented (buried in the literature review), which makes the overall storyline of the model results unclear. And also, as listed in the method part in the supplementary (Table S1), the analysis of different variables is mostly based on the mean value of different model ensembles. Also mentioned in this paper and from previous studies, model projections, especially of phytoplankton, could be very different. I understand CMIP6 model output towards phytoplankton, especially towards detailed phytoplankton limitations and community structure, is very rare. However, comparing the mean values of different variables when the members of the model ensembles are significantly different is not very convincing.

Overall, I think the current version of the paper requires major revisions and improvements before it can be considered for publication.

We thank the reviewer for their comments which we have used to improve the clarity, purpose and value of our paper. We have addressed the specific comments made below:

**Specific comments:**

L24: unclear what an "evidence-based sampling approach" is.

Replaced with “a sampling approach targeting the regions with the greatest rates of climate-driven change in ocean biogeochemistry and community assemblages”.

L74: I don’t think temperature-driven zooplankton metabolism changes will modulate the zooplankton grazing pressure. Could you explain more or/and add a reference?

We have added an additional reference referring to increased zooplankton metabolism at higher temperatures and a further reference for this increasing the grazing pressure, altering phytoplankton biomass.

"Additionally, water temperature, alongside changes to zooplankton abundance and diversity, has been shown to increase zooplankton metabolism (López-Urrutia et al., 2006, Mayzaud and Pakhomov, 2014), which can in turn be expected to modulate the grazing pressure and phytoplankton biomass (Lewandowska et al., 2014)."

L90: sentence "As the main source … long term storage" is long and complex. Could you cut it short and make it more clear?

Revised to: "As phytoplankton are the main source of organic carbon in the Southern Ocean, uncertainty in projections of phytoplankton composition compounds existing model uncertainty in the biological carbon flux to the ocean's interior and seafloor (Henson et al., 2022), where a fraction is available for long term storage."

L97: Is "a targeted approach" the same as "an evidence-based sampling approach" in the abstract? If so, try to keep the same name and explain the approach more in detail.

Resolved by removing evidence-based, see response to L24.

L123: Figure 2a instead of 2d

Corrected

Figure 2: "*Units in panels E and F" change to "*Units in panels **D** and F". Also, have you mentioned anywhere how exactly you calculated the limitation of irradiance and iron on phytoplankton? What you show in Figure 2 is for surface phytoplankton or averaged over the water column, and how?

Corrected panel labelling.

Irradiance and iron limitation are parameters in CMIP6 (limirr, limfe). The CMIP definition of these parameters are:

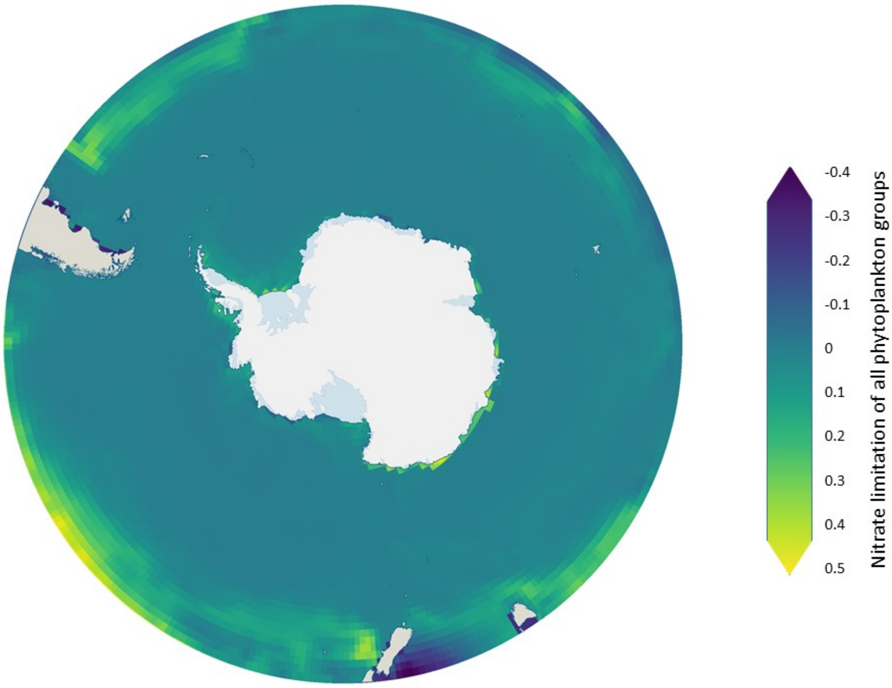
"Iron growth limitation" means the ratio of the growth rate of a species population in the environment (where there is a finite availability of iron) to the theoretical growth rate if there were no such limit on iron availability.

"Growth limitation due to solar irradiance" means the ratio of the growth rate of a species population in the environment (where the amount of sunlight reaching a location may be limited) to the theoretical growth rate if there were no such limit on solar irradiance.

Figure 2C contains the intpp variable, this is a 2D variable in CMIP which is already provided as a depth integrated value. The iron and light limitation parameters are similarly 2 dimensional, considering all phytoplankton in the water column. Note that this differs from chlorophyll in Figure 3 which is a 3D variable that we have integrated between 0 and 500m.

L191–192: "Models do not show any increases in nitrate limitation over the remainder of the century." I wonder how you got this conclusion from figure 3 (or any other figure not listed).

We have added the nitrate limitation to the supplementary information as Supplementary figure 7 and added the "limn" parameter to Table 1. Note that only one model (GFDL—ESM4) contains the nitrate limitation parameter.

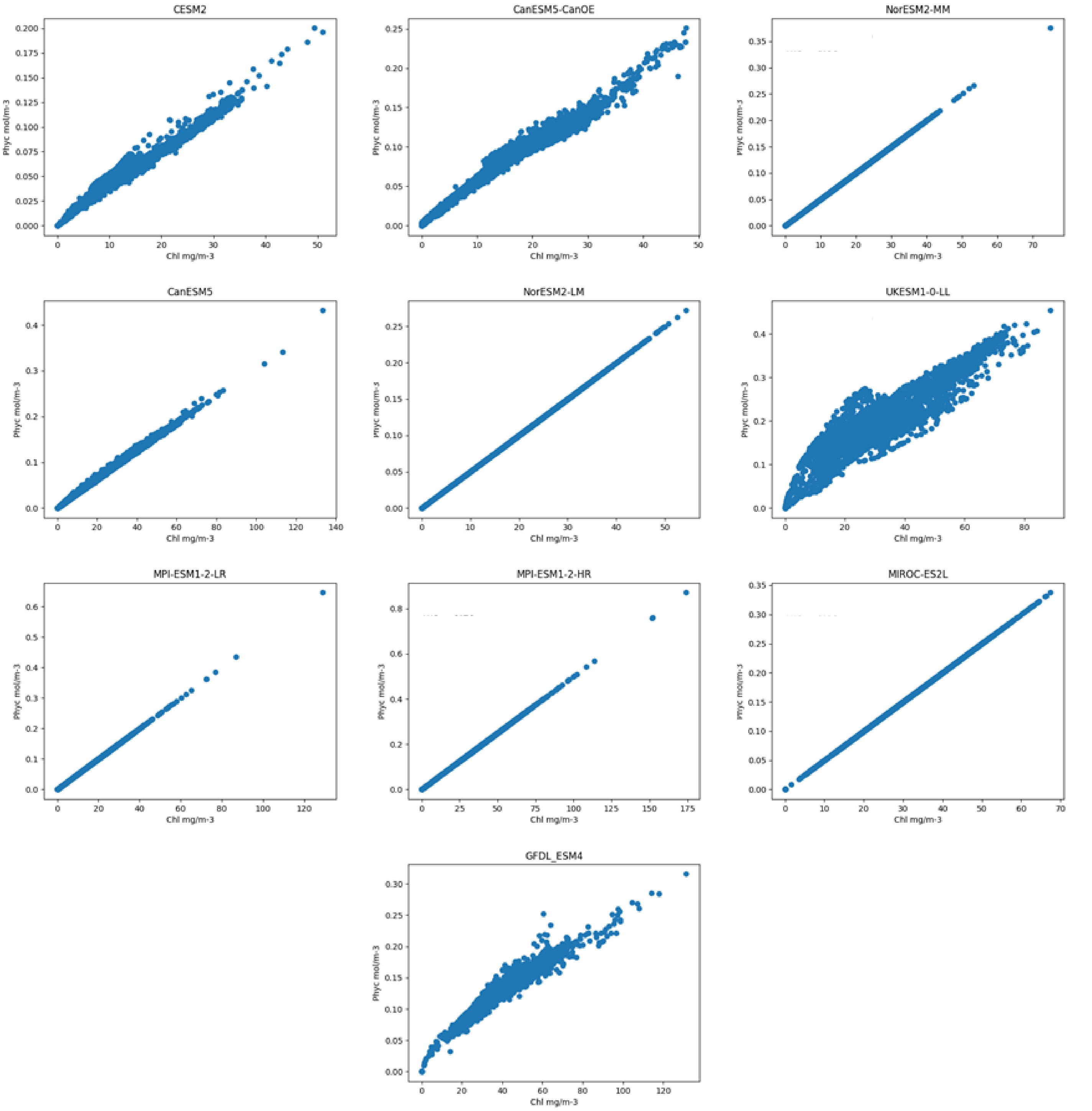



**Supplementary Figure S7**: Anomaly in nitrate limitation of all groups of phytoplankton across the Southern Ocean for SSP5-8.5 compared to a historical mean (1985-2015). Data shown for GFDL-ESM4, being the only CMIP model to include nitrate limitation. Limitation of all groups is derived as the sum of "limndiat","limnpico" and "limnmisc". Units are the anomaly value between a ratio of growth under environmental nitrate concentrations and theoretical growth under unlimited nitrate.

Figure 3: CMIP6 models output both chlorophyll and phytoplankton biomass in carbon data. If you want to see how integrated phytoplankton change, I think phyC would be a more appropriate variable, as you don't have to disentangle the varying chl:C ratio, which is also not included in every model.

We thank the reviewer for their suggestion. We correlated chl against phyc and found that chl:C variability is very low for the models we consider, in the majority of models there is a direct correlation. As a paper focusing on augmenting observations, we believe that chlorophyll is the most useful and widely understood parameter for the observational community.

Note: For the purpose of this correlation, we plot a random 1% subsample of grid values south of 65 degrees latitude; values shown are a time average from 2100 under SSP5-8.5.




L213: What is "rain rate"? Similar to export efficiency?

Rain rate is the same as export efficiency (the amount of carbon travelling through the water column), but usually used in the context of carbon arriving at the seafloor. We have changed this to “the rate of carbon export to the seafloor”.

L274: Why do you discuss phytoplankton productivity (intpp) in carbon and biomass in chlorophyll, when you can directly compare carbon to carbon (phyc)?

In CMIP6 phyc covers all phytoplankton types and cannot be split by groups, so offers no insight in to community composition changes. We chose to focus on chl and intpp because both of these parameters can be varied by phytoplankton group. Additionally, as a paper aiming to discuss targets for observations we wished to include chlorophyll as one of the most widely measured biological variables. In response to reviewer 1 we showed that for the models we consider chl is almost always directly correlated with phyc. Considering phyc instead of chl would not change any of the trends in our data.

L275: I don't think phytoplankton productivity and biomass are two independent variables.

Changed from independent to "variable". In our response to reviewer 1 we included chl:phyc (direct correlation) and chl:intpp which has more variability. We believe this to be a factor of differential chl/c ratios in different phytoplankton groups.

L296: It is figure 4 instead of figure 3.

Corrected

Figure 4: The research area is not explicit in the caption.

Figure caption updated to: "Evaluation of GFDL-ESM4 and CESM2-WACCM models using an anomaly between 2100 (SSP5-8.5) and a historical average (1985-2015) for the Antarctic zone (65-90 °S)".

L299–301: I like this part of the result comparing the different signals from GFDL and CESM.

Figure 5: The mean of total productivity (intpp) is calculated as the mean of 15 models (Table S1), while diatom productivity (intppdiat) is calculated as the mean of 6 models. intpp and intppdiat are calculated from very different model ensembles (intpp model ensemble does not even completely cover the intppdiat model ensemble). Therefore, directly compare intpp and intppdiat, and the calculation of non-diatom productivity as intpp-intppdiat, I think, is incorrect. I suggest here to use the models that have outputs of both intpp and intppdiat and then do the comparison and calculation.

Apologies for the confusion here. The reviewer is correct that we can only compare multi-model means where we use the same models. We have only included the 6 models containing the intppdiat parameter to derive the intpp value in the analysis shown in Figure 5. Intpp is listed as having 15 models in Table S1 (now Table 1) because all 15 models are used to create Figure 2C (using intpp to show total productivity). We have updated the figure caption for Figure 5 to make the model selection clearer.

Figure 5 caption: "Six CMIP6 models were used in this analysis, details of the specific models assessed are given in Table 1, only models containing the diatom productivity parameter are included."

We have also put the models used in the time series analysis in bold in Table 1 and added an explanatory *:

*All listed models were used for deriving total primary productivity by all phytoplankton types (Figure 2C). Only those models which also contain intppdiat (shown in bold) are included in the phytoplankton group timeseries (Figure 5).

L363–364: This part is shown in the second panel of Figure 5. Please label Figure 5 and indicate the figure once you talk about it.

Panel labels have been added to Figure 5.

L363–364: Productivity trends are based on the mean of a multi-model ensemble, while the iron variability is only from GFDL. If you want to use it this way, at least in the appendix, show that the GFDL productivity trend is not standing out from other model results.

L367–369: same here, please refer to the figure.

Figure reference added.

L421: Could you explain more why you exclude "some models"? What kind of difference in data structure are you referring to?

This is actually referring to the exclusion criteria the reviewer suggested for Figure 5. We only compare models where all the available parameters are available, therefore models can be excluded because (in the context of figure 5) comparable variables are not available (e.g. here are 15 models containing ‘intpp’ but only 6 with ‘intppdiat’ so 9 models are excluded when creating the multi-model mean for ‘intpp’ in this conext). Or, more commonly, because a parameter will have a value for either SSP5-8.5 or the historical run but not both, so we have to exclude some models to ensure the anomaly is representing the same models at the start and the end of the century. We have removed the term “compatibility issues” and will clarify our model matching process in the methods.

**Additional references**


Annett, A. L., Skiba, M., Henley, S. F., Venables, H. J., Meredith, M. P., Statham, P. J. & Ganeshram, R. S. 2015. Comparative roles of upwelling and glacial iron sources in Ryder Bay, coastal western Antarctic Peninsula. *Marine Chemistry,* 176**,** 21-33.

Bakker, D. C., Pfeil, B., Smith, K., Hankin, S., Olsen, A., Alin, S. R., Cosca, C., Harasawa, S., Kozyr, A. & Nojiri, Y. 2014. An update to the Surface Ocean CO 2 Atlas (SOCAT version 2). *Earth System Science Data,* 6**,** 69-90.

Freeman, N. M., Lovenduski, N. S., Munro, D. R., Krumhardt, K. M., Lindsay, K., Long, M. C. & Maclennan, M. 2018. The Variable and Changing Southern Ocean Silicate Front: Insights From the CESM Large Ensemble. *Global Biogeochemical Cycles,* 32**,** 752-768.

Gregg, W. W., Conkright, M. E., Ginoux, P., O'reilly, J. E. & Casey, N. W. 2003. Ocean primary production and climate: Global decadal changes. *Geophysical Research Letters,* 30.

Hewitt, H. T., Roberts, M., Mathiot, P., Biastoch, A., Blockley, E., Chassignet, E. P., Fox-Kemper, B., Hyder, P., Marshall, D. P., Popova, E., Treguier, A.-M., Zanna, L., Yool, A., Yu, Y., Beadling, R., Bell, M., Kuhlbrodt, T., Arsouze, T., Bellucci, A., Castruccio, F., Gan, B., Putrasahan, D., Roberts, C. D., Van Roekel, L. & Zhang, Q. 2020. Resolving and Parameterising the Ocean Mesoscale in Earth System Models. *Current Climate Change Reports,* 6**,** 137-152.

Jones, E. M., Henley, S. F., Van Leeuwe, M. A., Stefels, J., Meredith, M. P., Fenton, M. & Venables, H. J. 2023. Carbon and nutrient cycling in Antarctic landfast sea ice from winter to summer. *Limnology and Oceanography,* 68**,** 208-231.

Laufkötter, C., Vogt, M., Gruber, N., Aita-Noguchi, M., Aumont, O., Bopp, L., Buitenhuis, E., Doney, S. C., Dunne, J. & Hashioka, T. 2015. Drivers and uncertainties of future global marine primary production in marine ecosystem models. *Biogeosciences,* 12**,** 6955-6984.

Lewandowska, A. M., Hillebrand, H., Lengfellner, K. & Sommer, U. 2014. Temperature effects on phytoplankton diversity — The zooplankton link. *Journal of Sea Research,* 85**,** 359-364.

López-Urrutia, Á., San Martin, E., Harris, R. P. & Irigoien, X. 2006. Scaling the metabolic balance of the oceans. *Proceedings of the National Academy of Sciences,* 103**,** 8739-8744.

Mayzaud, P. & Pakhomov, E. A. 2014. The role of zooplankton communities in carbon recycling in the Ocean: the case of the Southern Ocean. *Journal of Plankton Research,* 36**,** 1543-1556.

Moreau, S., Hattermann, T., De Steur, L., Kauko, H. M., Ahonen, H., Ardelan, M., Assmy, P., Chierici, M., Descamps, S., Dinter, T., Falkenhaug, T., Fransson, A., Grønningsæter, E., Hallfredsson, E. H., Huhn, O., Lebrun, A., Lowther, A., Lübcker, N., Monteiro, P., Peeken, I., Roychoudhury, A., Różańska, M., Ryan-Keogh, T., Sanchez, N., Singh, A., Simonsen, J. H., Steiger, N., Thomalla, S. J., Van Tonder, A., Wiktor, J. M. & Steen, H. 2023. Wind-driven upwelling of iron sustains dense blooms and food webs in the eastern Weddell Gyre. *Nature Communications,* 14**,** 1303.

Okin, G. S., Baker, A. R., Tegen, I., Mahowald, N. M., Dentener, F. J., Duce, R. A., Galloway, J. N., Hunter, K., Kanakidou, M., Kubilay, N., Prospero, J. M., Sarin, M., Surapipith, V., Uematsu, M. & Zhu, T. 2011. Impacts of atmospheric nutrient deposition on marine productivity: Roles of nitrogen, phosphorus, and iron. *Global Biogeochemical Cycles,* 25.

Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M.-H., Maass, H. & Platt, T. 2009. Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. *Marine Ecology Progress Series,* 383**,** 73-84.

Seferian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J. G., Li, H., Long, M. C., Luo, J. Y., Nakano, H., Romanou, A., Schwinger, J., Stock, C., Santana-Falcon, Y., Takano, Y., Tjiputra, J., Tsujino, H., Watanabe, M., Wu, T., Wu, F. & Yamamoto, A. 2020. Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Curr Clim Change Rep,* 6**,** 1-25.

Thomalla, S. J., Ogunkoya, A. G., Vichi, M. & Swart, S. 2017. Using Optical Sensors on Gliders to Estimate Phytoplankton Carbon Concentrations and Chlorophyll-to-Carbon Ratios in the Southern Ocean. *Frontiers in Marine Science,* 4.

Ward, B. A. & Follows, M. J. 2016. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proceedings of the National Academy of Sciences,* 113**,** 2958-2963.