Response to Referee Comments for egusphere-2026-044:
Towards Constraining the Drivers of Variability and Trends in Subantarctic Productivity

Christopher D. Traill, Tyler W. Rohr, Elizabeth H. Shadwick, Pearse J. Buchanan, Alessandro Tagliabue, and Andrew R. Bowie

General response to RC1
Referee 1 is wholeheartedly thanked for their positive feedback and highly constructive comments towards the improvement of the manuscript.

We have attended to the suggestions of the referee and revised the manuscript following their suggestions. In particular, we have revised both the introduction and conclusion sections to more clearly and explicitly state both the study aims, findings and specific future recommendation. We also have revised the results and discussion text to improve the clarity around which significant differences emerge in variables between high and low productivity periods. We have also revised the arrangement of CMIP6 results, moving the novel analysis from the discussion to the results section with corresponding text amendments, and moving the CMIP6 NPP trends to the appendix.

Our responses to both major and minor referee comments with supporting figures, analyses and references are given below. Line numbers referenced in author comments refer to lines in the revised manuscript.

Referee Comments 1:
Major Comments:

1. Aims and Conclusions

The introduction to the paper sets up variability in NPP algorithms well and describes the environmental setting in good detail. However, I feel the introduction is lacking in setting out any real aims or hypotheses for the study. The final paragraph describes what the paper does but lacks a "why" aspect from the rationale behind the work. I suggest that setting clear aims perhaps as numbered or bullet points would make this aspect stronger. For example, "We aimed to determine the extent to which different environmental drivers contribute towards interannual variability in NPP within the SAZ".

Having a clear idea of what the paper aims to achieve would then also strengthen the conclusion which I found to be overly vague in areas. For example, on line 522: "Moreover, divergence in both the indicative drivers and NPP trends showed that the mechanistic relationships are not consistently represented between algorithms". The mechanistic relationships of the NPP algorithms are surely a product of the parameters they incorporate, so the fact that different algorithms represent different relationships is an inherent aspect of using different algorithms, i.e. we would know that the algorithms represent these mechanisms differently without requiring any analysis because that is what they are setup to do, the Table in Figure 2 shows this. I think it would be more useful for the conclusion to link to the aims of the study and provide a critical analysis of which algorithms are most suited to representing dynamics of the study region based on the mechanisms/parameters they include.

There is some allusion to this on line 471: While I acknowledge that the study does not seek to validate the best NPP algorithm. If, as the authors say, the AbPM, CAFE and CbPM models are more aligned with the observational trends then perhaps a useful conclusion of the study is that future observations need to incorporate more of the parameters used by these algorithms (e.g. backscatter, absorption, K_d and MLD). Including some element of observational recommendations alongside modelling recommendations in the conclusion would enhance the usefulness of the paper in directing future work. A similar point could be made with respect to the mechanisms identified. Is it just the case that we need to better incorporate multi-resource limitation, grazing and losses in to models or do we also need to improve the evidence base for these mechanisms through further observation?

We have significantly amended the final paragraph in the introduction to explicitly state the study aims and provide a basis for stronger recommendations in the conclusion section, including a stronger focus on the CMIP6 analysis context:

> **Lines 81-86**: "Firstly, we aimed to determine the relationship between interannual productivity variability and broadscale drivers from observational NPP data streams, specifically probing the influence of the subtropical front (STF) as a proxy for the extension of the East Australian Current (EAC) into the region. Secondly, using a biogeochemical model, we aimed to determine the explicit mechanisms by which physical drivers influence phytoplankton growth and loss rates and ultimately NPP changes by interrogating the explicit nutrient limitation, temperature limitation, light limitation and grazing terms. Finally, we aim to determine the variability in drivers of productivity across a suite of CMIP6 models, and discuss the implications of these findings for NPP trends, helping to disentangle uncertainty in first-order productivity drivers and observational/model disagreement."

The conclusion has also been amended to provide more specific assertions in support of the more detailed carbon and absorption based NPP algorithms, including suggestions on future efforts to better constrain in-situ bio-optical parameters.

> **Lines 646-652: "**While this is in part due to the exclusion of biogeochemical and ecological mechanisms in remote sensing algorithms, these estimates are still essential tools in assessing the impact of climate change on productivity systems. In this regard, our analysis supports greater confidence in CbPM, AbPM and CAFE NPP algorithms that integrate physical and optical parameters for estimating of large-scale Southern Ocean productivity change. Future observational efforts should focus on constraining the distinct bio-optical properties (e.g. Figure 2 legend) in the Southern Ocean towards developing consistent remote sensing NPP algorithms. **"**

> **Lines 674 – 680: "**Moreover the explicit mechanisms identified here are contributing to our understanding of earth system model NPP drivers, and are directly linked to the most important mechanisms that require attention across model intercomparisons (Henson et al., 2024), which can be used to inform future biogeochemical model development. Notwithstanding the need for target mechanistic development in biogeochemical models, observational efforts form a critical part in ground-truthing mechanism concepts. Future observational studies that better constrain changes in seasonal grazing pressures and multi-resource limitation in the Southern Ocean are a critical component towards reciprocal observation-model improvements. **"**

2. Statistics and data presentation

Data presentation/interpretation:

There is a lot of data presented within the article, most of which is styled as bar charts (Figure 2,4,5,6,7,11 plus Supplementary). I understand the choice for this, but this analysis relies on the reader essentially eyeballing significant bars in the NPP chart and then looking for significant bars which appear for the same month and model in the environmental drivers. The significant trends become lost in the noise, and there are examples of where the trends shown do not wholly match what is described in the text, for example Line 268: "Mixed layer depths were generally deeper in high NPP periods [...] (Figure 4b)." However, MLD is only significantly deeper in August and Feb and the greatest NPP months are Dec- Feb. So only one of the high NPP months correlates with a significant increase in MLD?

The sentence and paragraph has been amended to provide greater clarity on how MLD and NPP might be decoupled, and sets up the disentanglement story of competing bottom-up and top-down mechanisms in the model:

> **Lines 309 – 313:** "In high NPP years, MLDs were generally deeper in winter through to spring and sporadically in summer and autumn, with significantly positive ΔMLD only in August and February (Figure 4b). However, ΔMLD was very small from November to January and during March (Figure 4b), months where positive total ΔNPP was strongest (Figure 4a). Changes in mixed layer depth during spring may have a lasting memory effect into summer by modifying mixed layer nutrient, light and predator-prey interactions."

Figure 2 also differs from the in-text interpretation on line 389 which says: "The only significant relationship identified was in the CbPM algorithm where shallower MLD was correlated with increased NPP in November", however October also appears to have a significantly shallower MLD and significantly greater NPP for the CbPM algorithm (Figure 2).

This error has been amended:

> **Lines 473-475:** "Rather, the only significant relationship identified was in the CbPM algorithm, where shallower MLD was correlated with increased NPP in October and November (Figure 2). While this relationship was also indicated in the VGPM, eVGPM and CAFE algorithms, it was not significant (Figure 2). "

I think discussing trends between NPP and environmental drivers in generality rather than through statistical correlation creates confusion for the reader. I would suggest to the authors that creating a heatmap style figure which could summarise the significant environmental drivers for months with significant trends in NPP would be an easier way of visualising a large amount of complex data. I think this would be particularly useful for Figures 2 and 6.

We have provided for the benefit of the referee our analysis as a heatmap using a normalised scale for the differences in variable magnitude (delta-variable) for NPP-algorithms (Figures AC1) and PISCES-QUOTA-FE output (AC2). Code for these figures are provided in the included in publicly available Github repository: https://doi.org/10.5281/zenodo.18148572.

We agree with RC1 that it is challenging to succinctly condense many variable into a cohesive picture. It also important to contextualise statistically significant results relative to each phytoplankton group, the average phytoplankton composition, and to the magnitude and units of the differences that arise.

Owing to the need to retain useful information on the absolute magnitude and units of differences between highest and lowest productivity years, the bar plot format was chosen as a balance between dimensional reduction and meaningful, real-world values, while allowing direct comparison across algorithms rather than comparison between environmental variables. We found that this information is lost in the normalised results of the heat map visualisation without adding variable specific scales, which would also necessitate eyeballing to compare differences across phytoplankton groups and determine in which groups the most significant differences occurred.

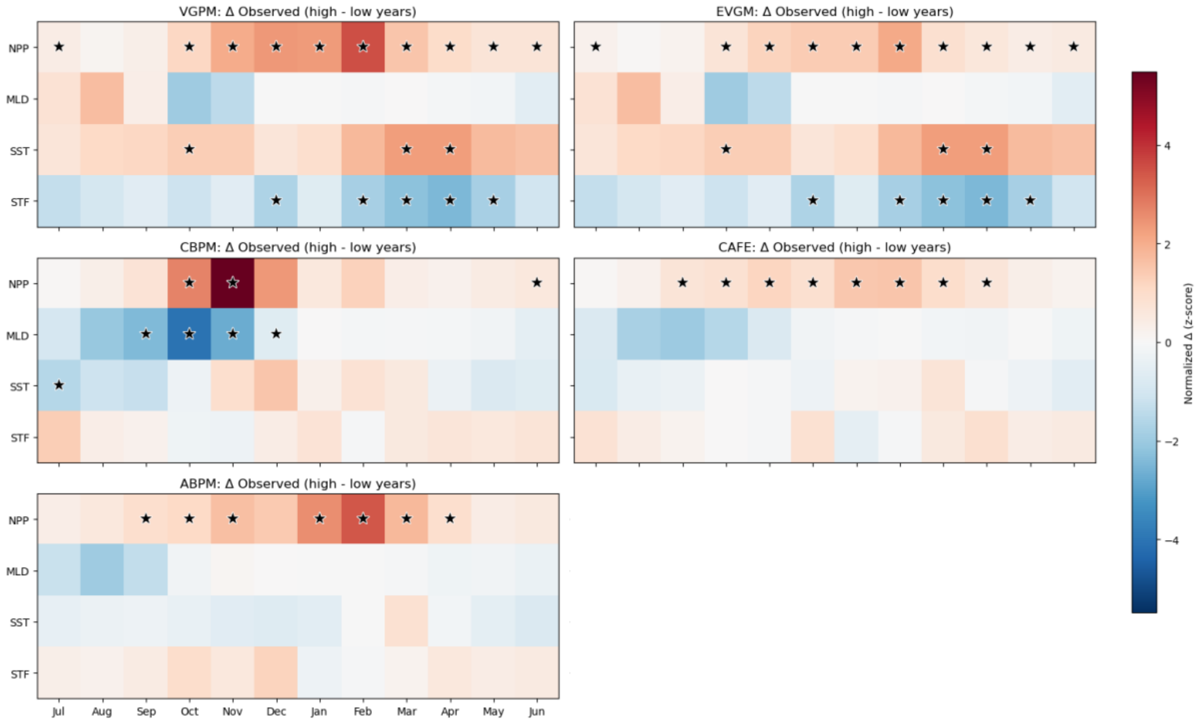


Figure AC1. Heatmap approach to presenting NPP-algorithms changes (Δ-values) in phytoplankton class specific and total mean climatological terms determined from the mean climatological difference between high and low integrated total NPP seasons (see Figure 2 in manuscript). Panels are separated by satellite NPP algorithms, with colours denoting normalised Δ-values across variables of NPP, MLD, SST and relative STF position. Stars denote that mean monthly climatological difference in the selected high and low NPP season groups is significantly exceeds the full time series climatological variability at the $p < 0.05$ level (10,000 iteration random year group bootstrapping analysis).

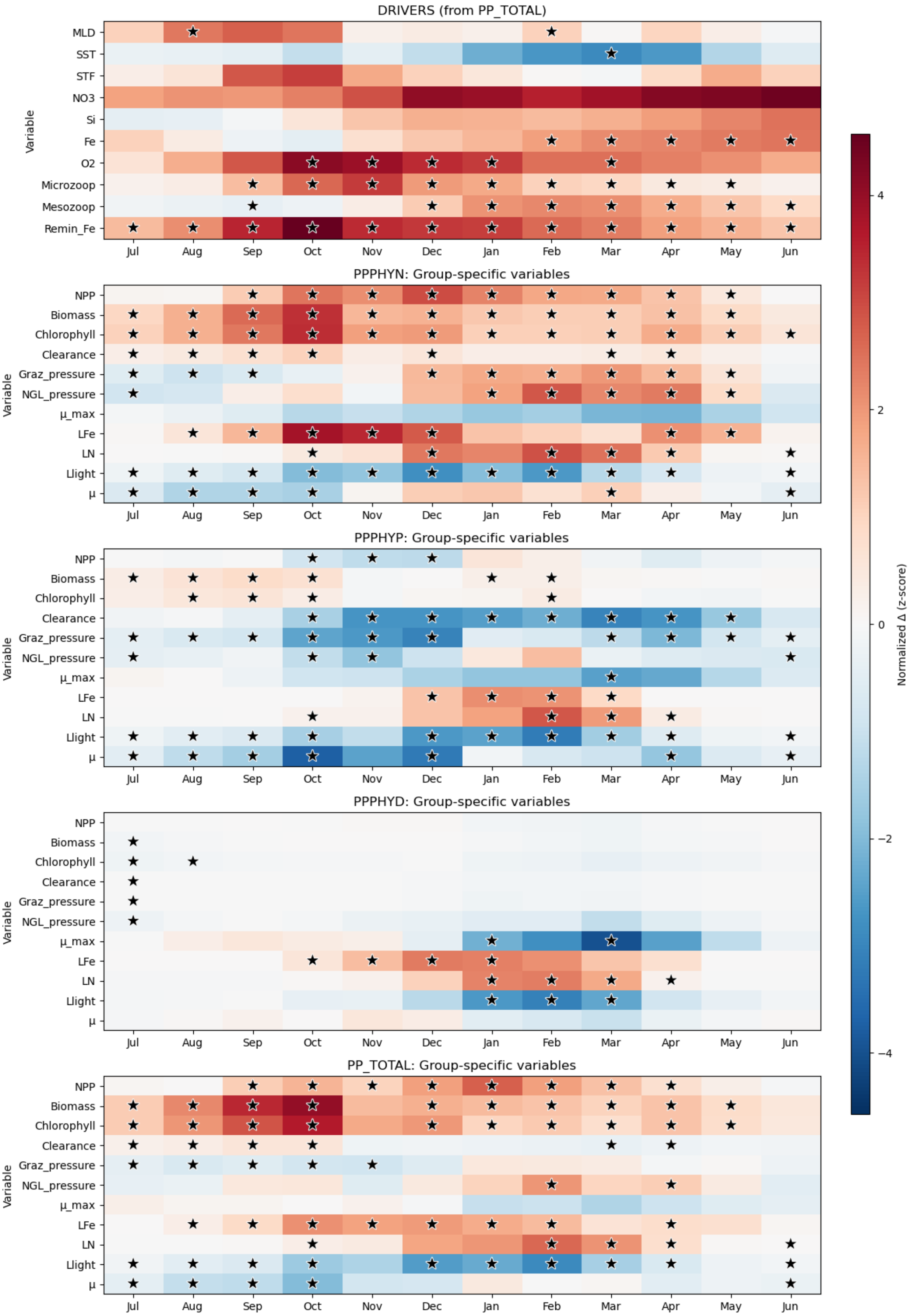


Figure AC2. Heatmap approach to presenting the PISCES-QUOTA-FE simulated changes (Δ) in phytoplankton class specific and total mean climatological terms determined from the mean climatological difference between high and low integrated total NPP seasons (see Figure 6 in manuscript). Panels are separated by biomass weighted depth-averaged phytoplankton groups and total phytoplankton composition. Stars denote that for each variable (y-axis), mean monthly climatological difference in the selected high and low NPP season groups significantly exceeds the full time series climatological variability at the $p < 0.05$ level (10,000 iteration random year group bootstrapping analysis).

Statistical approach:

I understand the choice to use the anomalously high NPP months as means of identifying the environmental drivers. Since the study is using the top/bottom 5 years in the observational time series, where variability could be explained by external forcings (e.g. ENSO/SAM), it would be useful for the authors to expand on which years were selected as a result of this analysis and where their NPP values sit relative to the climatological mean. I am left wondering whether the high/low years selected are consecutive to one another or interspersed across the time series for example?

In the publicly available code, outputs are explicitly specified to print the year subsets (which are centred on January and run July to the following June) in all NPP algorithm time series and the PISCES output timeseries. This ensures that this information is clearly conveyed for any dataset processed by other users.

In the PISCES output section, the years used are:
Highest 5 highest integrated NPP (all phytoplankton groups) July-years: [1990 2000 2003 2015 2019]
Lowest 5 integrated NPP (all phytoplankton groups) July-years: [1975 1979 1980 1988 1989]

In the observational NPP algorithm section, the years used are:
Selected July years max CbPM-NPP: [1999 2004 2005 2009 2012]
Selected July years min CbPM-NPP: [2008 2010 2014 2015 2016]

Selected July years max eVGPM-NPP: [2009 2013 2017 2021 2022]
Selected July years min eVGPM-NPP: [1998 2001 2003 2005 2010]

Selected July years max CAFE-NPP: [1999 2000 2002 2009 2011]
Selected July years min CAFE-NPP: [2014 2015 2017 2018 2019]

Selected July years max VGPM-NPP: [2009 2013 2017 2021 2022]
Selected July years min VGPM-NPP: [1998 2001 2003 2005 2010]

Selected July years max AbPM-NPP: [2000 2002 2008 2009 2011]
Selected July years min AbPM-NPP: [2015 2017 2018 2019 2022]

In regard to the influence on climate mode influence on anomalously high NPP (e.g. ENSO/SAM), it is possible that some physical conditions may not be independent from each other. However, our analysis and interpretation operate under the assumption that interannual differences are attributed to physical and biological processes operating at the sub-seasonal scale such that they are not autocorrelated between years. Our focus is on the explicit processes operating within the seasonal timescale, that generate differences in NPP between years. What drives the physics (i.e. underlying climate modality in the JRA55 reanalysis) is not currently assessed in this study, and so we have not explicitly stated the year selection in the main text.

3. CMIP6 data

I find the introduction of CMIP6 data in the discussion to be misplaced. I don't think new datasets should be added to the paper in the discussion. I would prefer this to either be incorporated in the results or the supplementary material. I am actually not convinced that the paper needs to include

CMIP6 analysis, the main purpose of the inclusion seems to be to show an agreement of an increasing trend in NPP with PISCES-Quota-Fe, but as the paper already points out this has already been shown for the SAZ in other CMIP6 analyses (Fisher et al., 2025, Ryan-Keogh et al., 2025), so referencing the trend shown in these papers would be sufficient in my view.

Indeed, because the data features only in the discussion the results are never discussed in much detail, e.g. Figure 10 is only referenced once in the paper and Figures 11 a) and d) not at all, so they make only a minor contribution to the overall picture. I would encourage the authors to either expand on the CMIP6 aspect and include it as an additional dataset within the results or move to the supplementary if it is to remain as a minor section.

Figure 10 results (CMIP6 NPP trends) have been moved to the appendix and are cited now as Figure A8. Figure 11 (CMIP6 bootstrap analysis) has been moved from the discussion to a new results section (3.5). We have added further text in the results section 3.5 and discussion section 4.3 to more comprehensively incorporate CMIP6 results and the application of our statistical methods into the manuscript:

> ***Lines 439 - 455:***
>
> *"***3.5 – Environmental drivers of subantarctic NPP in CMIP6 models**
> In the highest NPP years, increased SSTs and more southern STF position were generally consistent across the whole year across ACCESS-ESM1.5, CanESM5-CanOE, CanESM5-CMOC, CESM2 and CMCC-ESM2 models, suggesting that thermal drivers dominated highest NPP years for this subset of CMIP6 models (Figure 8c). In PISCES-v2 containing models IPSL-CM6A and CNRM-ESM2-1, SSTs were only significantly warmer from July to October (Figure 8c), which opposes our findings of cooler SSTs in PISCES-QUOTA-FE assessment (Figure 4c). Despite generally warmer SSTs, the significant differences between highest and lowest NPP year groups in CMIP6 models were variable, with ΔNPP ranging from 2 to 40 mmol C m^2 d^{-1} (Figure 8a). The highest differences in NPP between highest and lowest NPP year groups were in CESM2 and lowest differences in ACCESS-ESM1.5, CanESM5-CMOC, GFDL-ESM4 and CMCC-ESM2 models (Figure 8a).
>
> Between highest and lowest NPP year groups, ΔMLD was large and varied substantially across models (Figure 8b). Through winter and spring, ACCESS-ESM1.5 ΔMLD was significantly shallower than any other model, with ΔMLD values ranging from 100 to 300 m shallower. during high NPP years (Figure 8b), Deeper MLDs were significant for only CanESM5-CMOC in October, and IPSL-CM6A and CNRM-ESM2-1 in November (Figure 8b). IPSL-CM6A and CNRM-ESM2-1 use PISCES-v2 biogeochemical modules, and deeper MLDs in November are consistent with our assessment of PISCES-QUOTA-FE."
>
> **Lines 621-625**: "This is particularly the case with variability between models in the direction of ΔMLD during high NPP years (Figure 8b) yet consistently increased ΔSSTs suggesting thermally driven NPP trends (Figure 8c). Moreover, broad variability in the difference between highest and lowest NPP years in CMIP6 models (Figure 8a) highlights that NPP response mechanism to warmer temperatures vary broadly."

Minor comments:

Abstract

Line 14: I would introduce the concept and scale of NPP interannual variability (could you add a quantification?) and then lead in to mentioning light/iron and the unknowns about mechanisms which cause it.

Amended clarify interannual NPP variability concept:

Line 14: "Seasonal iron and light limitation control annual net primary production (NPP) in this region, but the explicit mechanisms that drive differences in NPP between years (NPP interannual variability) remain elusive due to sparse observations"

Line 15 that follows indicates that variability quantification is difficult given circumstantial to NPP product selection, which can vary as much as 27 to 36% of the maximum NPP across algorithms in a given year.

> Line 15: "This uncertainty is reflected in inconsistent interannual variability and trend estimates of remotely-sensed NPP algorithms."

Line 19: analyse -> analysed
Line 19: Include which biogeochemical model is being used (PISCES-Quota-Fe).

Corrected line 19: analyse -> analysed

Added model details:

> **Line 19:** "…we analysed the explicit bottom-up and top-down controls of depth integrated NPP in a biogeochemical ocean model historical run (PISCES-QUOTA-FE; 1958-2022) in the subantarctic zone south of Tasmania, Australia."

Line 20: The Indian Sector is much broader than the area this study is working in, I would be more specific about it being the region around SOTS.

Amended to more regional specificity:

> **Line 19:** "…we analysed the explicit bottom-up and top-down controls of depth integrated NPP in a biogeochemical ocean model historical run (PISCES-QUOTA-FE; 1958-2022) in the subantarctic zone south of Tasmania, Australia."

Line 27: "We applied our analysis to CMIP6 models, and while all historical simulations converged...". Please edit this to clarify a selection/ensemble of models were analysed, the current wording could imply that the trend was across all CMIP6 models.

Clarified that a subset of CMIP6 models were analysed.

> **Line 28:** "We applied our analysis to a subset of CMIP6 models, and while all historical simulations converged with respect to positive trends in NPP,…"

Line 33: Can you be more specific about which observational datasets are needed? The conclusion also mentions future work needed in model development through mechanistic inclusion, that should also be added to the abstract as a takeaway from the paper.

Amended final abstract sentence:

> **Lines 33-35:** "Therefore, the need for cautionary approaches to NPP trend interpretation is highlighted. Additionally, further observational data, including zooplankton grazing and phytoplankton biomass distributions over depth are needed to ground truth mechanistic understanding of NPP drivers and develop models through mechanistic inclusion. "

Introduction

Line 38: Could you add reference to the size of the CO2 sink? 1.39 (± 0.3) mol C m^{-2} yr^{-1} according to Shadwick et al., (2023).

The reference to Shadwick et al (2023 is provided in **Line 47:** "Over the annual cycle, this region is a net CO_2 sink (Lenton et al., 2013; Shadwick et al., 2023), driven predominantly by biological processes (primary production) (Shadwick et al., 2015; Yang et al., 2024b)."

We have not provided a quantitative estimate of the net CO_2 sink given the variability in annual CO_2 sink estimates depending on methodology and discussed in the cited references, as this is outside the scope of this study.

Line 43: "How primary productivity responds to these changes in uncertain". Yes, but previous studies have looked at the impact of warming, freshening, stratification and acidification on primary productivity in the SAZ, this should be referenced. I would recommend Deppeler and Davidson (2017) as they discuss the potential for increased stratification to reduce nutrient delivery and productivity and the impact of temperature change on grazing/growth rates. Constable et al., (2014) could be useful also.

The referee is thanked for the suggestion and we have added a citation for Deppeler & Davidson (2017) to the cited references.

> **Line 54:** 'How primary productivity responds to these changes is uncertain but is critical to assessing how regional biogeochemical cycles and ecosystems change (Deppeler & Davidson, 2017; Henley et al., 2020; Anugerahanti & Tagliabue, 2024; Henson et al., 2024; Hutchins & Tagliabue, 2024).

Line 48: "show varied relationships to plausible physical drivers". Can you give examples of what these physical drivers are?

Added examples from Ryan-Keogh et al (2025b).

> **Lines 57-58:** "Trends in NPP from different satellite estimates show varied relationships to plausible physical drivers including sea surface temperature (SST) and mixed layer depth (MLD) (Ryan-Keogh et al., 2025), "

Line 64: "decoupling of iron distributions from the prevailing physical supply pathways due to biological processes". Which processes?

Detail added on specific biological processes:

> **Lines 64-65:** "However, the decoupling of iron distributions from the prevailing physical supply pathways due to biological processes, including uptake and remineralisation "

Line 83: STF and EAC need to be defined on their first use.

Amended.

> **Line 85:** '…influence of the subtropical front (STF) as a proxy for the extension of the East Australian Current (EAC) into the region.'

General comment: It would be useful to mention how representative the SOTS site is of the wider SAZ (or Indian Sector), how important are the findings of this study for understanding the wider region?

Lines added with relevant references to the SOTS regional representativeness

> **Lines 132-134**: "The SOTS observatory, nominally located at 142°E, 47°S, is situated in the SAZ, is considered broadly representative of SAZ waters south of Australia between 90 – 145 °E (Trull et al., 2001; Shadwick et al., 2015; Yang et al., 2024a)."

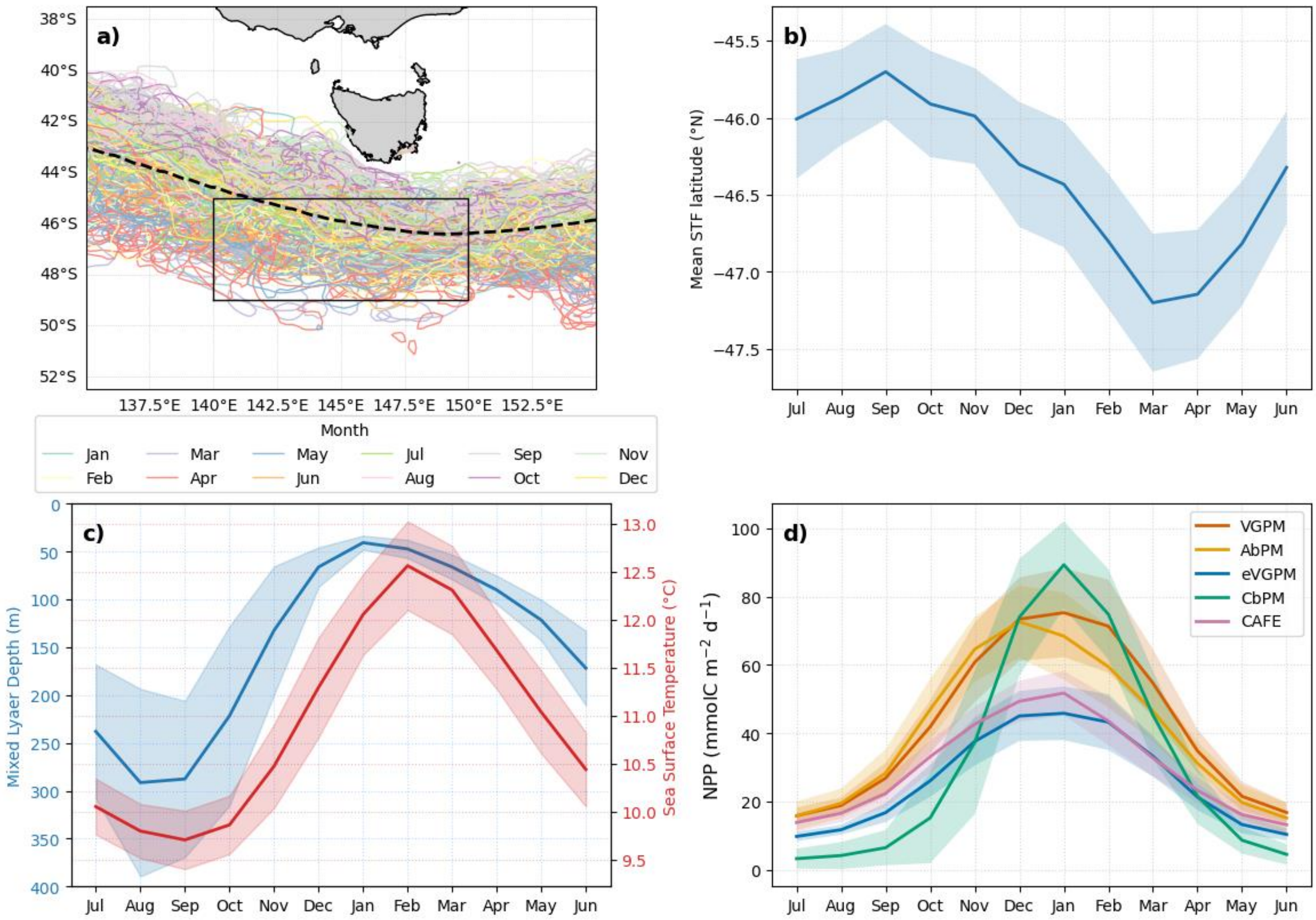
Methods

Figure 1: For panel C can the y axis be refitted to show the 0 depth, this will put the data above 100m within the scale.

Figure 1c y-axis scale has been amended to show the 0 m value (see revised figure 1)

Figure 1 (and throughout): I would recommend changing the colour scheme used between VGPM and CbPM to something colourblind friendly.

NPP-algorithm Figures 1, 2, and 8 have been modified to an Okabe-Ito colourblind friendly palette. The referee is thanked for highlighting this issue relating to fair accessibility for all readers.



Revised Figure 1. Including changes to panel c y-axis and colorblind friendly palette (Okabe-Ito categorical) in panel d.

Line 108: Probably “Section” rather than “Chapter”.

Deleted referral to ‘Chapter’ (previous draft relic).

Section 2.2: Where does the PAR, backscatter and absorption data used in CAFE, CbPM, AbPM come from (as listed in Figure 2?), this isn’t mentioned here. The data comes from an existing dataset, but it would be useful to know where the initial parameters came from, I’m guessing BGC-Argo?

An exhaustive list of every OC-CCI-based NPP algorithm dependency source is beyond the scope of this study, as these details are already given in the references provided:

> **Line 154:** “Estimates of NPP from five remote sensing algorithms were obtained from the OC-CCI-based dataset processed by (Ryan-Keogh et al., 2023a)…”
>
> **Lines 159-160:** “Detailed discussion on the algorithm input dependencies and assessment is given in Figure 2, Ryan-Keogh et al. (2023a) and Ryan-Keogh et al. (2025). ”

For the benefit of the referee, PAR is from Globcolor, while backscatter and absorption data are from OC-CCI v6.0, with exception of the spectral slope of backscatter which is a derived variable from specific remote sensing reflectance wavelengths.

Line 127: Ryan-Keogh et al., (2023a).

Citation format amended from ‘(Ryan-Keogh et al., 2023a)’ to ‘Ryan-Keogh et al., (2023a)’.

Figure 2a: If we compare the January productivity for the high anomalies against the mean algorithm data (Figure 1d) we see quite different trends. How much of the intra-algorithm variability in the mean NPP data (Figure 1d) is described by high anomalies approach shown in Figure 2a?
Line 168: Just because the data is high/low (10% percentile) it doesn’t mean the data is necessarily anomalous. Like my previous comment on this, it would be good to see where the selected years lie on the distribution.

The referee highlights an important observation. The discrepancy between the magnitude of intra-algorithm variability (Figure 1d) and that of the difference in mean monthly climatological values for NPP between the high- and low-NPP years in Figure 2a is because the data in Figure 2a is the difference between the averages of the 5 year top and bottom years respectively. This results in difference magnitudes being smaller than the positive and negative range of 1 standard deviation from the mean temporal variable of the spatial mean. This may also results from spatial mean differences in subsetted years driving difference sin difference magnitudes.

We retain the use of the term ‘anomalous’, since our statistical approach is use to determine if differences between between the high- and low-NPP years are significant relative to the full time series variability (i.e. Figure 1d) at the $p < 0.05$ level.

Line 179,180: I think some parts of the Equation 1 are missing from the text here.

Amended.

> **Lines 179-183: “**Where $\Delta var(\mathrm{m})$ is the difference in mean monthly climatological values for variable *var* in month *m* between the high- and low-NPP years, $\overline{var}_{high}(m)$ is the mean monthly climatological value of *var* at month m calculated over the five highest integrated NPP years, and $\overline{var}_{low}(m)$ is the mean monthly climatological value of *var* at month m calculated over the five lowest-NPP years.”

Table 1: In PISCES-Quota-Fe, what does the limitation term represent. In CMIP it is the growth under environmental nutrient concentrations / maximum growth under non limiting concentrations. Is it the same here?

In PISCES-QUOTA, limitation of growth rate by nutrients are similarly, modelled, by specifically follow the formalism of the chain-model of Pahlow and Oschlies (2009). Limitation terms are determined from normalised quotients relative ot minimum and maximum nutrient quotas, and for Fe, optimum iron quotas that allow for luxury uptake beyond a quota maximum for growth only. See Kwiatkowski et al. (2018) for details of limitation term derivation in PISCES-QUOTA.

It should be noted that individual CMIP6 biogeochemical modules are likely to deviate from the simplified (or apparent) quotient calculation of nutrient limitation terms as indicated by the referee, especially over large spatial and temporal averages. It is often not appropriate to apply limitation based on modelled growth under nutrient conditions alone, since in multiple phytoplankton group models since nutrient requirements differ across different phytoplankton and must be weighted across groups proportionately if examining total productivity. Additionally, some models also quota based nutrient limitation which are linked to other nutrients that can co-factors for uptake, or have quotas that not only influenced by grid cell nutrient concentrations but other parameterised biological processes (e.g. luxury iron uptake beyond cellular requirements). The explicit nutrient limitation terms, as presented in our study, are seldom saved in output, but are the correct indicator needed to really understand how growth in the model is limited.

Results

Line 243: Mean temperatures also look cooler in PISCES compared to observations.

See figure 3d: SST is right on the Obs line in the Taylor diagram.

Figure 3c: Should the errors from the upper panel be carried forward to the % contribution?

Error shading has been added to Figure 3c % contribution panel:

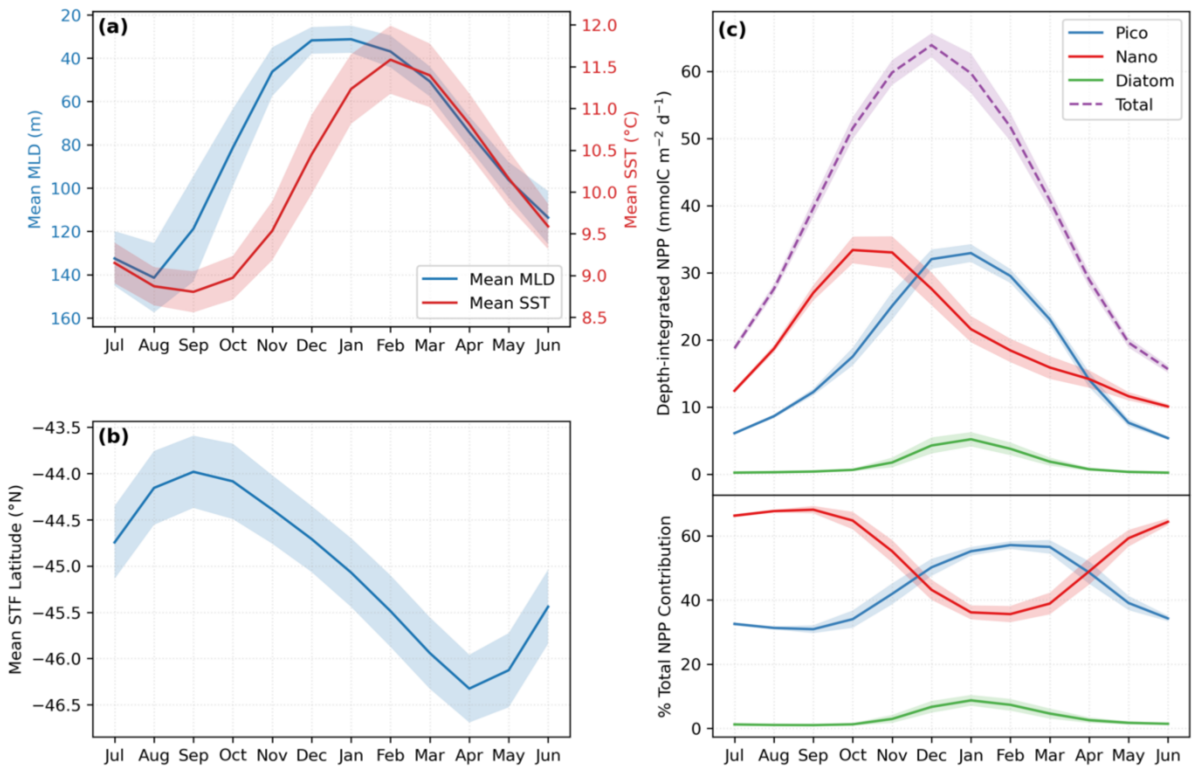


Figure 3d: How is relative bias defined? Also, it would be useful to specify that colour denotes the directionality of the bias.

$$\% \, Bias = \frac{\text{mean of simulated variable} \; - \; \text{mean of observed variable}}{\text{mean of observed variable}} \times 100$$

Amended Figure 3d caption:

> **Line 285:** Relative bias (%) is given as marker size and colour denoting bias direction.

I think there should be more text describing the results of Figure 3, for example the contribution of different phytoplankton types is omitted. It would also be useful to add more detail around the trends shown in Figure 3d as this is a complex figure to understand.

Text in the results section has been expanded to highlight the changes in phytoplankton contribution and aid the interpretation of Figure 3d Taylor diagram:

> **Line 275-285**: "The simulated climatology of NPP, MLD, SST and STF mean latitude in PISCES-Quota-Fe (Figure 3) was generally consistent with observations (Figure 1) despite some biases. Seasonal mixed layer depth amplitude was notably lower than observational derived MLD, and the STF mean latitude was further north in simulations compared to observations (Figure 3d). Simulated NPP over the seasonal cycle was larger than CbPM, eVGPM and CAFE algorithms and lower than VGPM and AbPM algorithms, while simulated NPP variability (normalized standard deviation) was lower than CbPM, VGPM and ABPM algorithms and higher compared to eVPGM and CAFE algorithms (Figure 3d). The good agreement between simulated NPP and remote sensing algorithms (correlation values > 0.9; Figure 3d) provides a strong basis with which to test the simulated mechanistic drivers of anomalously high and low NPP"

Line 256: "Cooler temperatures at depth between high and low NPP seasons were observed during summer but were generally insignificant". Given that there is no stippling (significance) in the temperature-depth profile below 100m (Figure 4e) it would be more accurate to say "insignificant" rather than generally insignificant, if there are no cases in which it is a significant trend.

Amended sentence accuracy.

> **Lines 298**: "Cooler temperatures between high and low NPP seasons were observed during summer, but were insignificant except in the surface 75 m in March (Figure 4e). "

Line 248: "Lower temperatures at depth were linked to a northern shift in the STF". How are you defining depth here, because Figure 4e shows warming or no change in temperature at 250- 300m?

Sentence rephrased:

> **Line 299:** "STF latitude did not vary significantly between high and low NPP periods (Figure 4d), despite cooler temperatures co-occurring with a northern shift in the STF in spring and late autumn. While this was more consistent with CbPM, CAFE and AbPM NPP algorithms than VGPM/eVGPM (Figure 4d), variability in simulated NPP is not statistically linked to increases in subtropical water mass presence in this region. "

Line 288: "Total Biomass was largest in spring, peaking in October when the mixed layer shoals (Figure 3a)". If the Biomass is integrated to 300m then why would the shoaling of the mixed layer impact the depth integrated biomass so much, it would just change the depth of the biomass

within the water column? Is the significant increase in Nanoplankton grazing loss rate following the spring biomass accumulation (Figure 6h) not accounting for the reduction in biomass?

To avoid confusion and mis-implied causation, the reference to MLD shoaling is removed in the line.

> **Line 338:** "Total ΔBiomass was largest in spring, while ΔNPP was largest in January."

To clarify, Figure 6b shows ΔBiomass, that is, the difference biomass between high and low NPP years – not absolute biomass in a climatological sense.

Line 320: I think this paragraph is upside down. The bottom-up controls (e.g. entrainment) are introduced at the end of the paragraph but are mentioned as drivers of increased NPP at the start of the paragraph. I encourage the authors to bring the remineralisation/ entrainment aspect to the start of the paragraph before going on to discuss the impact on productivity.

The paragraph has been re-ordered to improve the logical flow as suggested by the referee:

> **Lines 373 – 384:** "We found that increased NPP during the spring arose from the unique ways in which bottom-up and top-down processes responded to deeper mixing. Initially, strong bottom-up controls linked to relief of iron limitation appeared to drive increased NPP. Fe limitation was reduced through July-November, with consistently and significantly increased class specific and total phytoplankton community averaged L_{Fe} terms in high NPP years (Figure 6d). This effect was confined to depths from the surface to just below the mixed layer (Figure 7d), following increased mixed layer DFe inventories (Figure 5c). Increased DFe mixed layer inventories during spring were not from remineralised iron sources (Figure 5e), suggesting that springtime relief of iron limitation was driven by entrainment. Significantly higher depth resolved Δμ within the mixed layer reflected these better growth conditions for the average phytoplankton (Figure 7c). This corresponded to increases in surface NPP (Figure 7a), despite a

> non-significant increase in surface biomass during the July-November period (Figure 7b, Appendix Figures A3, A4, A5). "

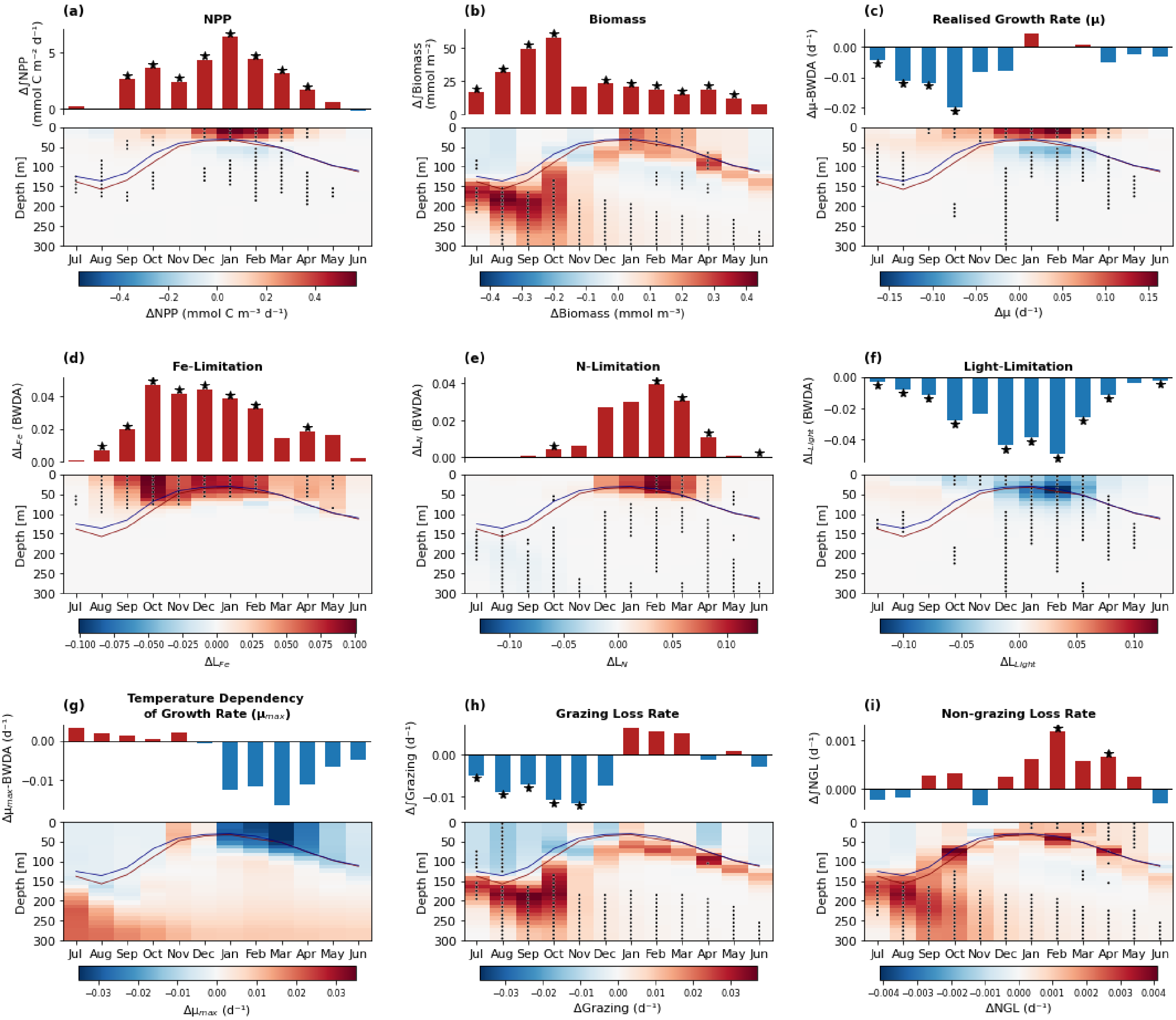
Line 343 (and Figure 7): If using the mean MLD is covering up mixing events would it make more sense to show the individual MLD's or some indication of variability in Figure 7?

Mean MLDs for the 5 -year top and bottom NPP periods are shown in Figure 7 panels a-i. Text is added in the caption to indicate this.

> **Lines 422 (Figure 7 caption):** "Red and blue solid lines denote mean MLDs in respective high and low NPP year groups."

Line 348: "shifting biomass to greater depths where growth conditions were unfavourable due to insufficient light". Figure 7b shows significant increases in biomass below the mixed layer, while Figure 7f shows significant depth integrated increases in light limitation during the spring these mostly seem confined to the upper ocean, not below the mixed layer? The evidence for N- limitation below the mixed layer in Figure 7e seems as strong, if not stronger than light limitation. How did you determine that light is the primary limiting factor on growth here?

See Manuscript Figure 7 below: During spring when biomass is disproportionately shifted to depth, the biomass weighted depth averaged light limitation is negative (blue), indicating smaller light limitation terms and stronger light limitation during higher NPP years. For the same period, the biomass weighted depth averaged N-limitation and Fe-limitation terms are positive, and due to the limitation of growth rate as function of N, Fe (Si), and light limitation terms (amongst other parameters), light is the dominant factor affect the growth rate. The benefit of biomass weighted depth averages is they show you the dominant effect experienced by the average phytoplankton, but omit at where that changes occurs, the signal of which can lose statistical significance in the depth resolved sense due to reduction in sample size.



Manuscript Figure 7.

Discussion:

Line 384/385: Is it useful to consider how nutrient limitation could impact plankton fluorescence through quenching and therefore measurements of chlorophyll/absorption also? E.g. Schallenberg et al., 2022.

This is indeed an important area of research – a comparison between differences in chlorophyll estimates using different methods, including accounting for fluorescence quenching, is discussed in Traill et al. (2025) in collaboration with C. Schallenberg. This discussion of this effect in the context of satellite NPP algorithms is beyond the scope of this study. See Ryan-Keogh et al. (2023b) for interaction between iron limitation and fluorescence yields from remote sensing.

Line 392: I have some issue with "enhanced light availability for mixed layer biomass", since biomass itself is not necessarily photosynthetic. I would suggest productivity is a better word.

Changed biomass to productivity.

> **Line 455: "**This does provide some indication that earlier shoaling of the MLD enhanced light availability for mixed layer productivity, consistent with observations of light as a seasonal productivity driver in this region."

Line 419: Could you expand on why reduced iron limitation benefited nanophytoplankton over Picos and diatoms. Is this because nanophytoplankton have a greater iron requirement in the model?

The text has been amended to highlight that nanophytoplankton benefit with regards to driing ht strongest change in NPP, due to their domination of the total biomass composition. When iron limitation is relieved and growth conditions improve, nanophytoplankton drive the greats NPP response with regards to improved nutrient availability.

> **Lines 481-484:** "Firstly, the responses of phytoplankton to bottom-up controls (class specific nutrient limitation) varied between groups, where reduced iron limitation in high NPP seasons allowed the strongest NPP contribution from nanophytoplankton as the dominant phytoplankton group over picophytoplankton and diatoms (cf. Appendix Figures A3, A4, A5). "

Prescriptive sensitivity of phytoplankton groups to iron concentration changes in the model is only partly made explicit by potential half saturation concentrations of iron for each group (pico are most sensitive with the smallest K-values, followed by nanophytoplankton and diatoms; Kwiatkowski et al. 2018).

Line 425: Change "the rate at which clearance rate" to "the clearance rate".

Added text to explain how the derivative of the clearance rate is a metric used to assess functional response stability in modelled ecosystems. See references by (Gentleman et al., 2003; Gentleman and Neuheimer, 2008).

> **Line 492:** "This was likely driven by the rate at which clearance rate, the volumetric rate at which phytoplankton are grazed per unit zooplankton, changes with phytoplankton concentrations (i.e. the speed at which clearance rate changes with respect to changing phytoplankton concentrations). "

Line 443: "Unfortunately, the multispecies predator prey system of PISCES-Quota-Fe challenges the quantification of inherent changes in stability from the rate of clearance rate changes with respect to phytoplankton concentration, due to multi-prey responses sensitive to community composition." This is a complicated sentence to understand for anyone not familiar with PISCES, I suggest rewording. The main point sees to be that different zooplankton groups graze on different phytoplankton groups which makes it difficult to compare the stability of the loss term of phytoplankton because it is not a single value?

The text has been clarified:

> **Lines 508-512**: "Unfortunately, the multispecies predator-prey system of PISCES-Quota-Fe challenges post-hoc quantification of inherent changes in stability from the rate of clearance rate changes with respect to phytoplankton concentration. This is due to multiple grazing rates

> across multiple phytoplankton groups and concentrations, driving differential phytoplankton loss rates that are lost across monthly and group averages. "

Line 493: "Without a specific growth limitation term archived for historical CMIP6 models". What growth limitation term is this in reference to? Multiple models contain iron limitation in historical simulations (IPSL-CM6A-LR, CNRM-ESM2-1, EC-Earth3-CC, UKESM1-0-LL), may need to be more specific about what is missing?

See reply to Methods: Table 1 comment. Derived limitation terms are not equivalent to group specific 'experienced' limitation terms explicitly saved in biogeochemical modules.

Text has been added to clarify what is needed from models:

> **Lines 565:** "Without explicit growth limitation terms from biogeochemical module componentry saved for historical CMIP6 models, there remains large uncertainty surrounding the full suite of environmental drivers. "

Conclusions:
See major comment on the structuring of the conclusion. Line 539: "environment" should be "environmental".

See reply to major comment 1.

Amended grammar.

> **Lines 613**: "…changes in nutrients, light or grazing, rather than environmental drivers such as temperature and mixing. "

Supporting References:

Anugerahanti, P., & Tagliabue, A. (2024). Response of Southern Ocean Resource Stress in a Changing Climate. *Geophysical Research Letters*, *51*(10). https://doi.org/10.1029/2023gl107870

Deppeler, S. L., & Davidson, A. T. (2017). Southern Ocean Phytoplankton in a Changing Climate. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00040

Henley, S. F., Cavan, E. L., Fawcett, S. E., Kerr, R., Monteiro, T., Sherrell, R. M., Bowie, A. R., Boyd, P. W., Barnes, D. K. A., Schloss, I. R., Marshall, T., Flynn, R., & Smith, S. (2020). Changing Biogeochemistry of the Southern Ocean and Its Ecosystem Implications. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.00581

Henson, S., Baker, C. A., Halloran, P., McQuatters-Gollop, A., Painter, S., Planchat, A., & Tagliabue, A. (2024). Knowledge Gaps in Quantifying the Climate Change Response of Biological Storage of Carbon in the Ocean. *Earth's Future*, *12*(6). https://doi.org/10.1029/2023ef004375

Hutchins, D. A., & Tagliabue, A. (2024). Feedbacks between phytoplankton and nutrient cycles in a warming ocean. *Nature Geoscience*, *17*(6), 495-502. https://doi.org/10.1038/s41561-024-01454-w

Kwiatkowski, L., Aumont, O., Bopp, L., & Ciais, P. (2018). The Impact of Variable Phytoplankton Stoichiometry on Projections of Primary Production, Food Quality, and Carbon Uptake in the Global Ocean. *Global Biogeochemical Cycles*, *32*(4), 516-528. https://doi.org/10.1002/2017gb005799

Lenton, A., Tilbrook, B., Law, R. M., Bakker, D., Doney, S. C., Gruber, N., Ishii, M., Hoppema, M., Lovenduski, N. S., Matear, R. J., McNeil, B. I., Metzl, N., Mikaloff Fletcher, S. E., Monteiro, P. M. S., Rödenbeck, C., Sweeney, C., & Takahashi, T. (2013). Sea–air CO2 fluxes in the Southern Ocean for the period 1990–2009. *Biogeosciences*, *10*(6), 4037-4054. https://doi.org/10.5194/bg-10-4037-2013

Pahlow, M., & Oschlies, A. (2009). Chain model of phytoplankton P, N and light colimitation. *Marine Ecology Progress Series*, *376*, 69-83. https://doi.org/10.3354/meps07748

Ryan-Keogh, T. J., Tagliabue, A., & Thomalla, S. J. (2025). Global decline in net primary production underestimated by climate models. *Commun Earth Environ*, *6*(1), 75. https://doi.org/10.1038/s43247-025-02051-4

Ryan-Keogh, T. J., Thomalla, S. J., Chang, N., & Moalusi, T. (2023a). A new global oceanic multi-model net primary productivity data product. *Earth System Science Data*, *15*(11), 4829-4848. https://doi.org/10.5194/essd-15-4829-2023

Ryan-Keogh, T. J., Thomalla, S. J., Monteiro, P. M. S., & Tagliabue, A. (2023b). Multidecadal trend of increasing iron stress in Southern Ocean phytoplankton. *Science*, *379*(6634), 834-840. https://doi.org/10.1126/science.abl5237

Shadwick, E. H., Trull, T. W., Tilbrook, B., Sutton, A. J., Schulz, E., & Sabine, C. L. (2015). Seasonality of biological and physical controls on surface ocean CO2from hourly observations at the Southern Ocean Time Series site south of Australia. *Global Biogeochemical Cycles*, *29*(2), 223-238. https://doi.org/10.1002/2014gb004906

Shadwick, E. H., Wynn-Edwards, C. A., Matear, R. J., Jansen, P., Schulz, E., & Sutton, A. J. (2023). Observed amplification of the seasonal CO2 cycle at the Southern Ocean Time Series. *Frontiers in Marine Science*, *10*. https://doi.org/10.3389/fmars.2023.1281854

Traill, C. D., Rohr, T., Shadwick, E., Schallenberg, C., Ellwood, M., & Bowie, A. (2025). Coupling Between the Subantarctic Seasonal Iron Cycle and Productivity at the Southern Ocean Time Series (SOTS). *AGU Advances*, *6*(3). https://doi.org/10.1029/2024av001599

Trull, T. W., Bray, S. G., Manganini, S. J., Honjo, S., & François, R. (2001). Moored sediment trap measurements of carbon export in the Subantarctic and Polar Frontal zones of the Southern Ocean, south of Australia. *Journal of Geophysical Research: Oceans*, *106*(C12), 31489-31509. https://doi.org/10.1029/2000JC000308

Yang, X., Wynn-Edwards, C. A., Strutton, P. G., & Shadwick, E. H. (2024a). Carbon Export in the Subantarctic Zone Revealed by Multi-Year Observations From Biogeochemical-Argo Floats and Sediment Traps. *Global Biogeochemical Cycles*, *38*(7). https://doi.org/10.1029/2024gb008135

Yang, X., Wynn-Edwards, C. A., Strutton, P. G., & Shadwick, E. H. (2024b). Drivers of Air-Sea CO2 Flux in the Subantarctic Zone Revealed by Time Series Observations. *Global Biogeochemical Cycles*, *38*(1). https://doi.org/10.1029/2023gb007766