Responses to Referee #2 :

The paper by Cioffi et al. uses model projections from the CMIP6 ensemble to investigate future changes in net primary production (NPP) in eastern boundary upwelling systems (EBUS), and their driving mechanisms. This is a very important topic that has been mostly analyzed within individual EBUS until now, with studies across EBUS focusing on upwelling itself rather than primary production. The paper was interesting and well written and nicely demonstrates that the future of NPP within EBUS is highly uncertain.

However, I believe that the topic of future NPP changes deserves very careful evaluation due to the large uncertainties in biogeochemical (BGC) modeling (as pointed out by the authors). This is particularly problematic in climate models in EBUS that are known to be biased, notably due to a lack of model resolution (e.g., Varela et al., 2022, https://doi.org/10.3390/jmse10121970 or papers cited in the conclusion). As such, several aspects need to be improved before publishing the paper.

We thank the reviewer for the useful comments and have made significant revisions to the manuscript to improve its completeness and clarity. We have developed a more robust model evaluation by analysing the performance of individual CMIP6 models and observational products in capturing mean EBUS NPP values. We have introduced new metrics, including the evaluation of models' ability to simulate historical NPP trends and key drivers such as nitrates and wind stress. Furthermore, we conducted a preliminary analysis to reduce projected EBUS NPP uncertainty and test potential relationships between historical and projected trends. We have expanded the Discussion section to provide a more comprehensive examination of the role of key drivers such as geostrophic transport. Finally, we have addressed several specific points raised by the reviewer, paying particular attention to improving the description of sources of NPP uncertainty at the EBUS scale.

Below are the authors' point-by-point responses to the comments, with reviewer's comments in bold and author's responses in normal font.

First, the "model evaluation" section (2.4) is wholly insufficient. The only evaluation presented compares mean NPP from all CMIP6 models to mean NPP from five data products. This is problematic on several levels: (1) each CMIP6 model should be evaluated separately instead of evaluating their mean (this will be important for the comment in the next paragraph); (2) the data NPP products should be used separately and not averaged; (3) other metrics beyond mean bias should be considered. Regarding (2), NPP satellite algorithms are known to display contrasting patterns and trends, and averaging products of likely varying quality is difficult to justify. I suggest the authors read the comment by Toby Westberry on the paper by Ryan-Keogh et al. (2023, https://doi.org/10.5194/essd-15-4829-2023) at https://essd.copernicus.org/articles/15/4829/2023/essd-15-4829-2023-discussion.html. Also note that the authors of that paper themselves analyzed the NPP products separately in their subsequent paper (Ryan-Keogh et al. 2025, https://doi.org/10.1038/s43247-025-02051-4 - I have no affiliation with either paper) with a discussion of this in the section "Assessing the merits of the different remote sensing algorithms". Regarding (3), I would suggest looking at other metrics beyond the mean, in particular the 1998-2014 trend – before assessing the model trends in the future, it would be useful to evaluate their trend in the reference period. Other metrics could be used (eg standard deviation) if the authors think they are pertinent to evaluating how likely models are to be "right" in the future.

We agree that a more robust model evaluation would be valuable in grounding our analysis. We have therefore expanded the Model Evaluation section and moved it to the Results section. We have developed the following main points:

- We propose keeping Figure 1 as it is in the current manuscript, as we believe it effectively demonstrates the models' general ability to simulate the historical mean state of NPP and provides a straightforward entry point for readers.
- We have assessed the ability of individual models to simulate the historical mean state of NPP. We propose adding in the Supplementary Information a new figure (Figure A of this Document) displaying a scatter plot of the mean NPP for each individual CMIP6 model alongside each of the five observational products (Ryan-Keogh et al., 2023) across the four EBUSs. This explicitly shows model-specific biases and spread within the observational data products.
- We investigated historical NPP trends (1998–2014) by comparing the CMIP6 models with the five observational products. We propose including a box-and-whisker plot in the main manuscript, comparing the historical normalised trends of the observational products versus the CMIP6 models at the sub-EBUS scale (Figure B of this document). This illustrates that the CMIP6 models are generally capable of capturing the trends, as in most regions, the simulated trends fall within the range of observational values. In some sub-systems, the interquartile range is even greater for the observational data.
- We extended the model evaluation to two key drivers of NPP:
    - Equatorward wind stress $\boldsymbol{\tau}$: we added an evaluation of both mean states and normalised trends using box-and-whisker plots (Figure C of this document), over the historical period (1998–2014). We compared the CMIP6 models with the ERA-Interim reanalysis (data available at https://esgf.nccs.nasa.gov/thredds/catalog/bypass/CREATE-IP/reanalysis/ECMWF/IFS-Cy31r2/ERA-Interim/mon/atmos/catalog.html) and found that they generally reproduced the mean state well, while the trends are more variable, potentially due to the important role of interannual variability over the considered period.
    - Nitrates: we added an evaluation of climatological states of both surface and subsurface (200 m) NO_3 using box-and-whisker plots (Figure D of this document), comparing CMIP6 models with World Ocean Atlas 2018 climatological data (Garcia et al., 2019) for the period 1900–2017 (WOA) and 1900–2014 (CMIP6 models). Our results show that the models effectively capture the nutrient fields at the EBUS scale

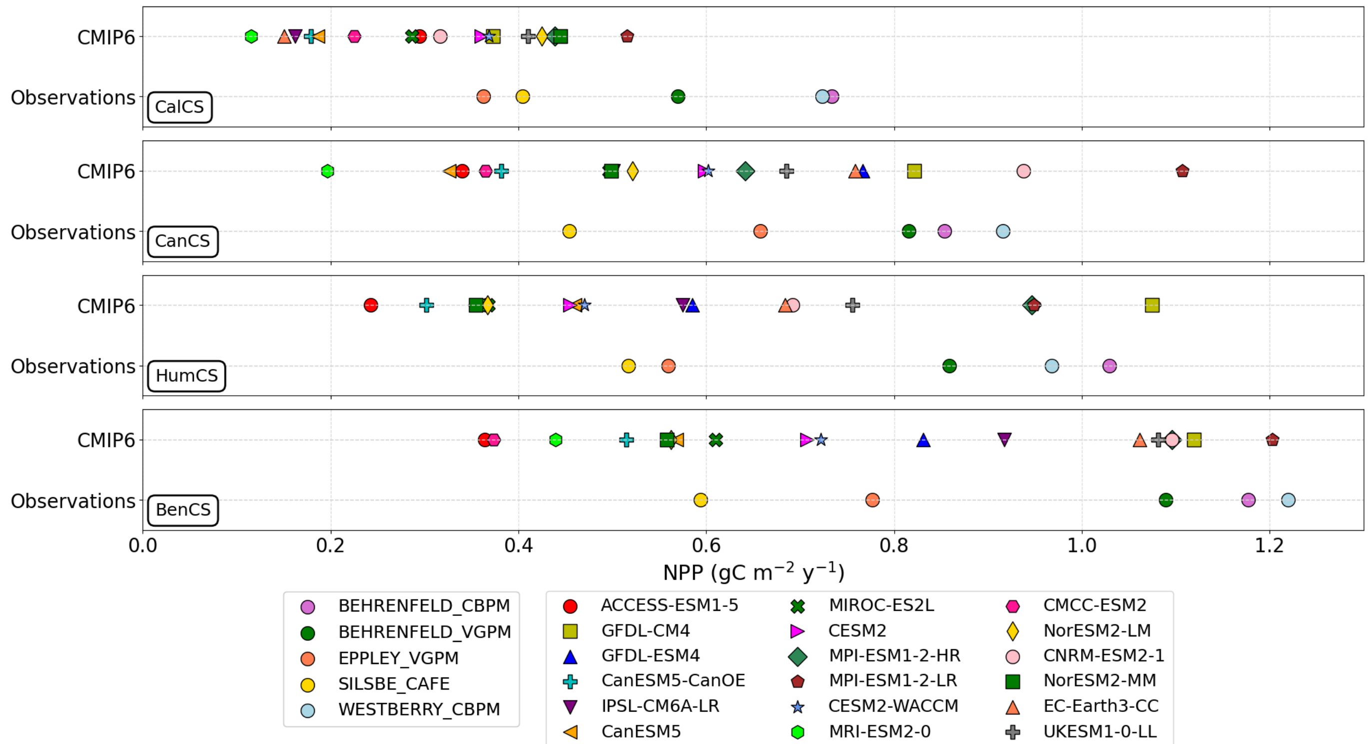

Fig. A) Evaluation of 18 CMIP6 models versus different data products of observations. Each colored marker represents the temporal (1998-2014) and spatial (EBUS-scale) average of NPP.

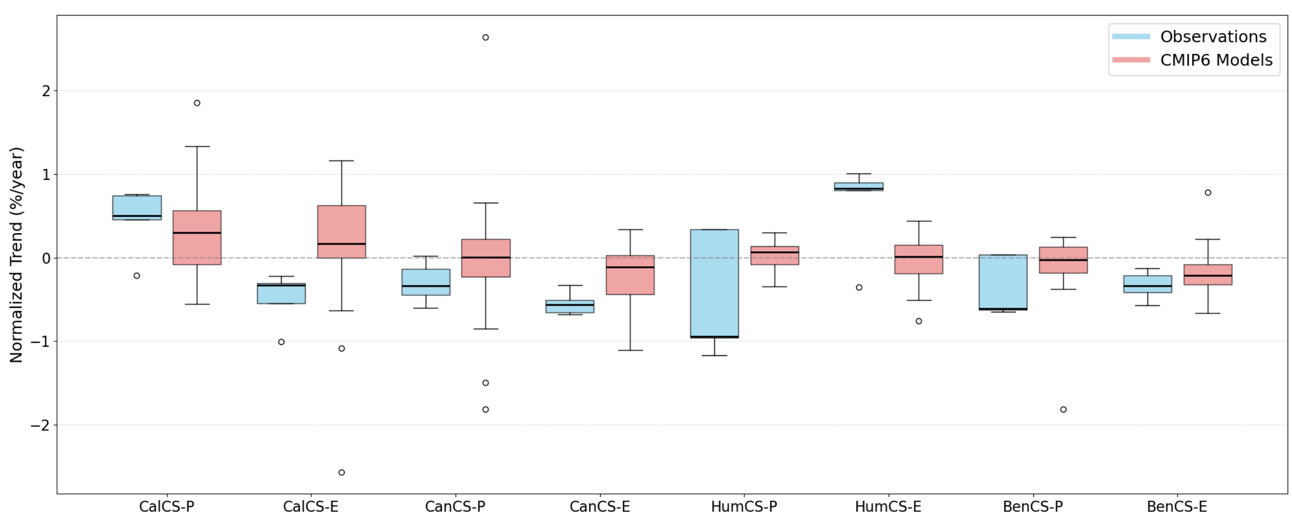

Fig. B) Box-and-whisker plots showing the distribution of observational data products and CMIP6 model estimates of historical (1998-2014) NPP trends across sub-EBUSs. For each region, boxes represent the interquartile range (IQR), whiskers extend to 1.5×IQR, and the median is shown as a thick black line, in blue for observations (five data products obtained combining ESA OC-CCIv4.1 ocean color observations with five different algorithms) and salmon for CMIP6 models.

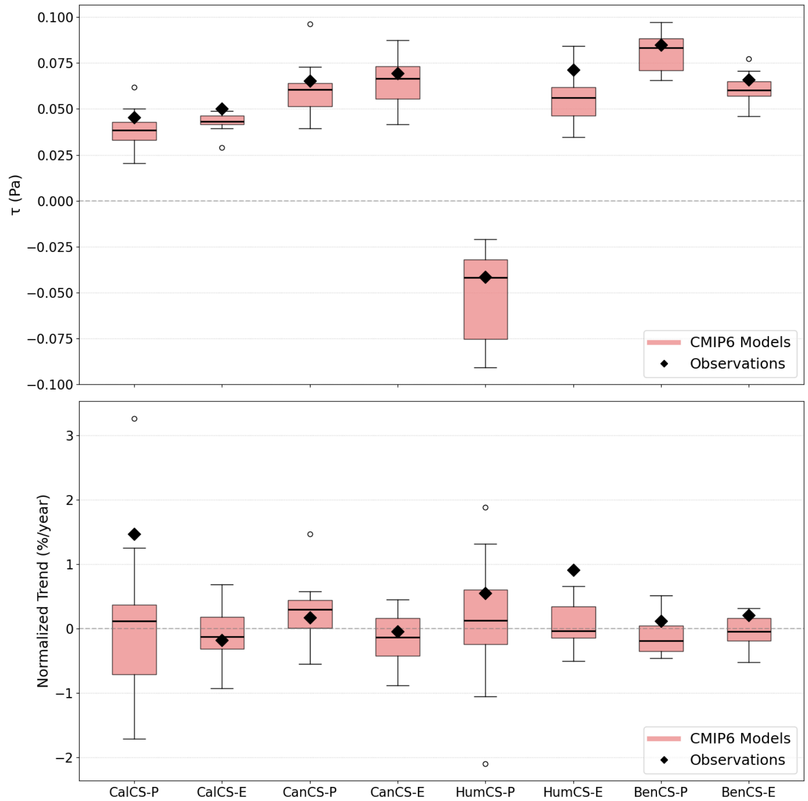

Fig. C) Box-and-whisker plots showing the distribution of observational data (ERA-Interim reanalysis) and CMIP6 model estimates of historical (1998-2014) τ mean values (upper figure) and trends (lower figure) across sub-EBUSs. For each region, boxes represent the interquartile range (IQR), whiskers extend to 1.5×IQR, and the median is shown as a thick black line, with the black diamond being observational data.

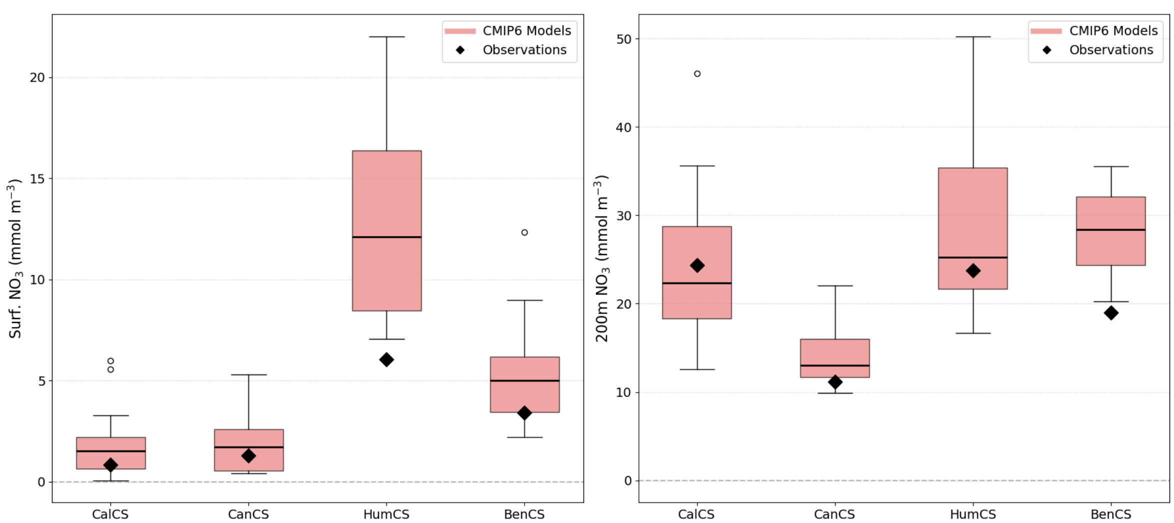

Fig. D) Box-and-whisker plots showing the distribution of observational data (from World Ocean Atlas 2018, Garcia et al., 2019) and CMIP6 model estimates of climatological (1900-2017 for observations, 1900-2014 for CMIP6 models) NO_3 at the surface (left figure) and at 200 m depth (right figure) across EBUSs. For each region, boxes represent the interquartile range (IQR), whiskers extend to 1.5×IQR, and the median is shown as a thick black line, with the black diamonds being observational estimates.

Second, I think there is a missed opportunity to narrow the projection uncertainty by considering the accuracy of each CMIP6 model in representing historical NPP (related to point (1) in the previous paragraph). I realize that most CMIP analyses only consider model ensembles, and this approach is valid for physical parameters that are well constrained by first principles equations. However, BGC model formulation can be highly variable and, even if BGC models similarly perform globally, their performance can be variable in specific regions (here EBUS) and for specific variables (here NPP). Moreover, the different models have an horizontal resolution varying from 0.25° to 1.5° in the 13-model subset, and resolution has a strong impact on upwelling representation and model accuracy in EBUS (see above). It would be very useful to define a "score" for each CMIP model that defines how well they perform during the historical period, and to see if the standard deviation of future projection decreases by only keeping the top X% performing models (half?). This is conceptually similar to the method used by Ryan-Keogh et al. (2025) although a simpler ranking scheme could be used based on model evaluation.

We believe that the proposed approach may prove useful in reducing uncertainties in NPP projections. Following the reviewer's suggestion, we conducted a preliminary analysis to verify whether there is a relationship between historical trends and SSP5–8.5 projections across the model ensemble (Fig. E of this document). However, we found no significant emergent relationship in the vast majority of sub-EBUS regions. This suggests that the historical NPP projections of these models are not linearly reflected in future sensitivities in a way that allows for a simple classification scheme, which prevents us from pursuing this specific study further. Although previous studies have succeeded in reducing uncertainties regarding NPP, for example by using emerging constraints (Kwiatkowski et al., 2017) or a model classification system based on the similarity of linear responses (Ryan-Keogh et al., 2025), any such approach would require significant additional work in this case, and we are not certain that this is fully in line with the objective of the present study, namely to understand, across the entire CMIP6 scale, the consistent future changes in NPP and its determining factors. Furthermore, given the high variability of observational data products at the EBUS scale, the selection of a subset of models could inadvertently introduce arbitrary choices that may not necessarily be optimal.

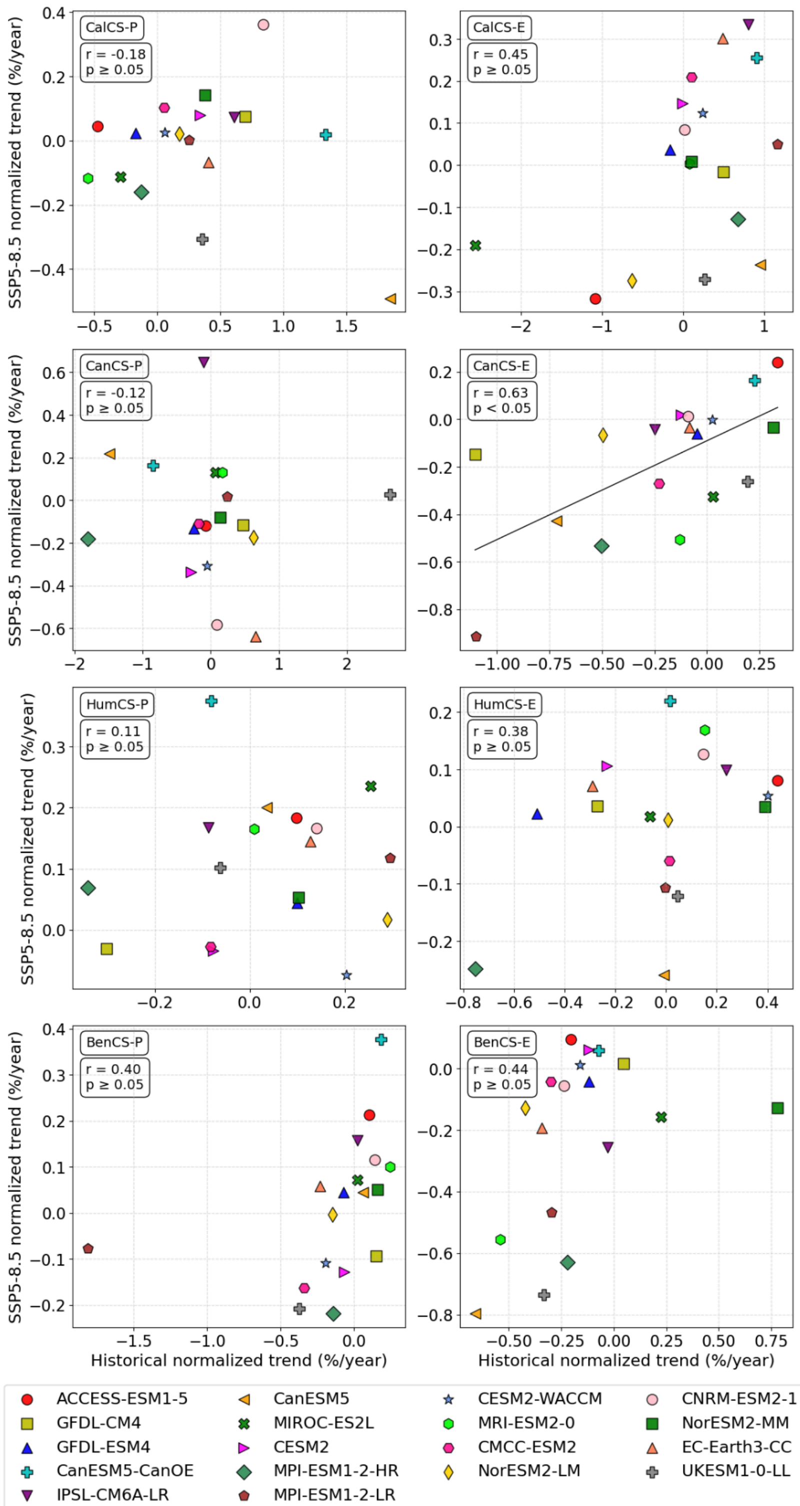

Fig. E) Relationship between historical and future normalized NPP trends across sub-EBUS for CMIP6 models. In each sub-EBUS, individual markers represent models, the x-axis indicates the historical normalized trend (1998-2014), the y-axis shows the corresponding future trend under the SSP5–8.5 scenario (2015-2100). The Pearson correlation coefficient (r) and its significance level are displayed in each panel, and a linear regression line is shown only where the relationship is statistically significant ($p < 0.05$).

Third, some of the hypotheses in the discussion could be easily tested or at least explored further. For instance, if geostrophic transport and wind-stress curl are important contributor to NPP changes, there should still be agreement between w60, surfNO3 and NPP in regions where there is no agreement between tau, surfNO3 and NPP (section 4.2). Similarly, agreement between subsurface nitrate, surfNO3 and NPP could be tested (section 4.3).

We will revise the Discussion section to improve both its clarity and completeness. In particular, following the reviewer's suggestions, we have undertaken the following:

- We will expand Section 4.2 to improve its integration within the overall analysis. Specifically:
    - We will provide a more detailed explanation of the processes already illustrated in Fig. A12 (Appendix), whose current description is quite concise.
    - To strengthen the analysis, we will include an additional "four-color" EBUS map (similar to Fig. 4 in the manuscript), showing the relationships between projected anomalies in Ekman and geostrophic transports (Fig. F of this document). This additional analysis shows that in the central HumCS and BenCS-P regions, Ekman and geostrophic transports tend to vary in opposite directions, likely contributing to the weak correspondence between wind changes and vertical velocity in these regions. This helps explain the projected changes in wo_{60} in the BenCS-P region, but it provides limited additional insight for the HumCS, where vertical velocity remains highly variable and noisy.
- We have analysed combined changes in NPP, surface and subsurface (200 m) NO_3 to test further potential relationships and investigate whether there are consistent concurrent changes in such variables. However, the results (Fig. G of this document), show only limited regions of consistent changes (CanCS-E and BenCS-E). We therefore prefer not to include this additional figure in the Manuscript, as we believe that this result does not add much to our current analysis and consider it sufficient to retain only Figures A13–A16.

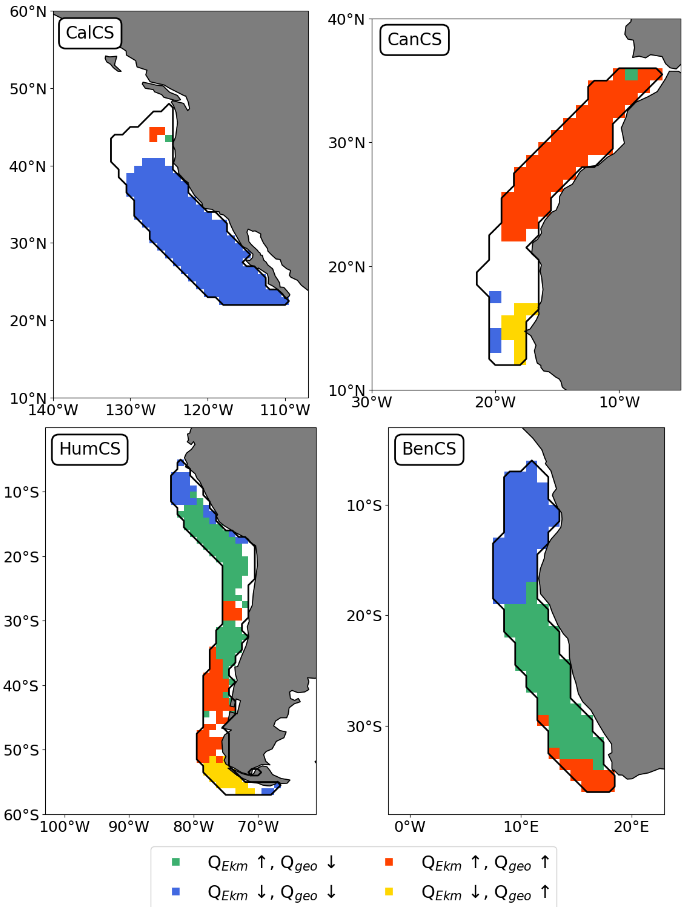

Fig. F) Ekman and geostrophic transport changes are reinforcing each other in some regions, offsetting in others. Signs of twenty-first century colocalized anomalies in Ekman and geostrophic transport under SSP5-8.5 (2071-2100). Grid cells are colored if there is ≥ 50 % model agreement.

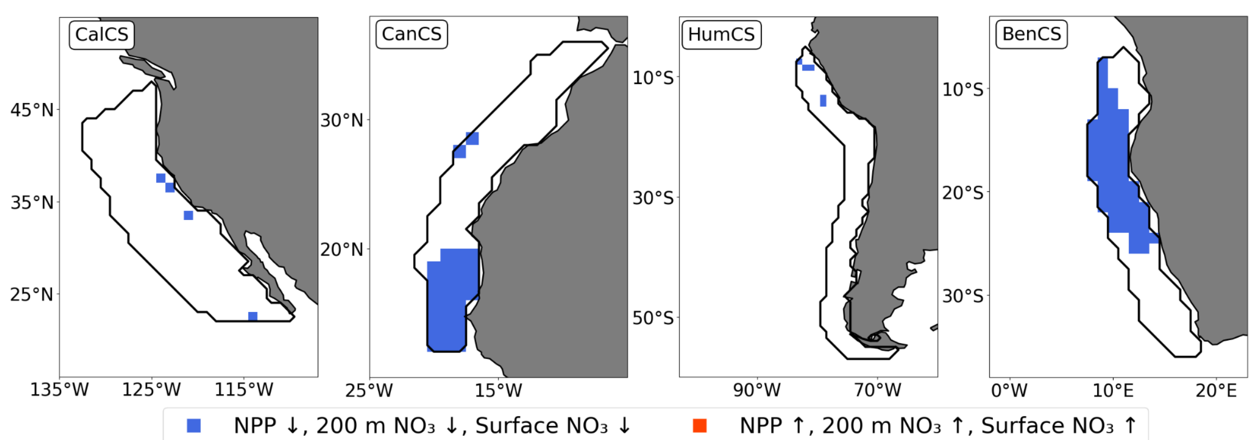

Fig. G) Signs of twenty-first century co-localized anomalies among key variables under SSP5-8.5 (2071–2100). Grid cells are colored if there is ≥ 50 % model agreement on the relationship between anomalies of NPP, surface and 200 m NO_3 combined.

Minor comments

Abstract “This uncertainty partially results from compensating positive and negative NPP anomalies within individual systems, with consistent multi-model responses only emerging at subsystem scales.” and similar statements elsewhere (l. 159-172): I don’t really understand what you mean. Are you referring to anomalies compensating between models (some models are positive, other negative) or regions (some subregions are positive, other negative)? In either case, that suggests higher variability in subregions than when considering entire EBUS, so I am unclear why this explains part of the uncertainty in projections. Also consistent multi-model responses are not very common according to Fig. 3 (1/3 of the coastal area perhaps? Should be quantified) so “consistent patterns projected across models” (caption) is contradicted by the figure, from what I see.

We acknowledge that this explanation could be made clearer, and we will revise the results section (l. 159–172) to improve clarity. The key point is that, while the high uncertainty at the EBUS scale is primarily linked to differences between models in certain regions (e.g., CalCS-E), in others (BenCS-E, BenCE-P, CalCS-P), it is also associated with contrasting and compensating trends within individual systems that are not apparent at the aggregated EBUS scale. When examining projected time series at the sub-EBUS scale, these spatially heterogeneous but internally consistent signals emerge and uncertainty is reduced (e.g., in BenCS, there is an offset between positive (in the poleward portion) and negative (in the equatorward portion) NPP anomalies).

Statements l. 36 “Most past studies have focused on (…) wind-induced upwelling” and l. 75 “we reevaluate the standard perspective that perturbations to upwelling favorable winds (…)” are somewhat contradicted by l. 45-66 that highlight a rich literature supporting other drivers (stratification, geostrophic transport, subsurface nitrate etc). These statements should be rephrased.

We will rephrase these sentences.

l. 125-127: did you use the products by Ryan-Keogh et al. (2023), or calculated your own? In the first case this paper needs to be cited. In the second case, more details are needed since the authors only cite OC-CCI which, as far as I know (and found on the website) is only chlorophyll.

We used the Ryan-Keogh et al., 2023 products, we will modify the manuscript by citing the paper.

l. 280: do all CMIP6 models include iron limitation? If not it could be useful to see if model performance in regions where it may matter most (CanCS) depends on iron inclusion in BGC models.

Except for two models (MRI-ESM2-0 and CanESM5), all CMIP6 members include a representation of the iron cycle (Séférian et al., 2020), although the level of complexity varies across models. We agree that a deeper investigation of how these different representations influence primary productivity would be valuable, but we recognize that such an analysis would require a substantial additional effort that goes beyond the scope of this study. For this reason, we consider our current approach, which focuses on nitrate as the primary limiting macronutrient, to be appropriate for the objectives of this work.

l. 120: the 30m approximation has only been shown to be valid in the CalCS system; is it valid for other systems too? Also please note that this is defined from the mixed layer depth (as described in Jacox et al., 2018) rather than Ekman layer depth.

We recognize that using a fixed 30 m depth is a simplification, as the Ekman layer depth varies spatially and seasonally and does not strictly correspond to the mixed layer depth, from which this approximation has been defined (Jacox et al., 2018). As pointed out, this approximation has been used in the California Current System (Jacox et al., 2018; Ding et al., 2021), but has also been applied to the BenCS (Veitch et al., 2010). However, there is limited consensus on how Ekman layer depth across different EBUSs. To account for potential differences (and considering that the 30 m approximation is derived from the MLD), we also computed geostrophic transport using spatially and temporally varying MLD fields from the models (Figure H of this document). This approach yielded results that are broadly consistent with those obtained using a fixed 30 m depth, however, in the HumCS-P, the MLD is substantially deeper, which complicates the interpretation of the results. For these reasons, we opted to retain the fixed 30 m depth, as it provides a more interpretable framework for inter-system comparison.

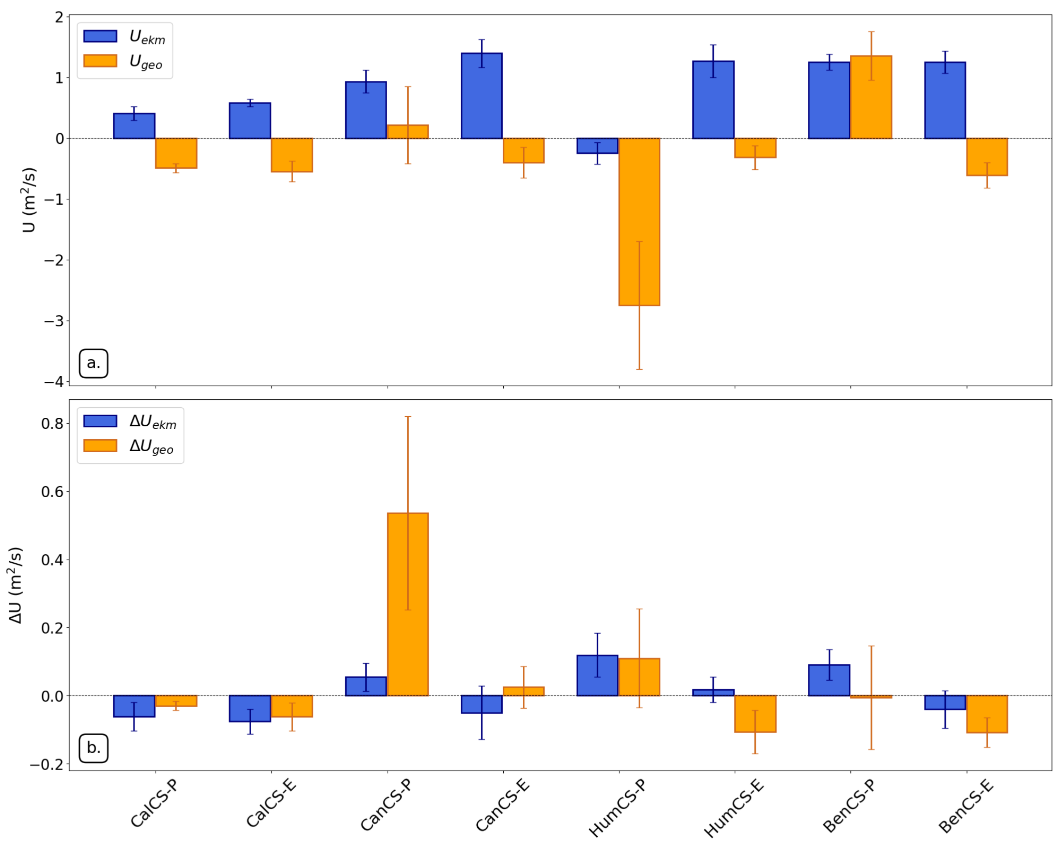

Fig. H) Relative contributions of Ekman and geostrophic transport in EBUS. Ensemble mean (13 models) Ekman and geostrophic transport in the sub-EBUS computed according to equations (1) and (2) respectively (using MLD as Ekamn depth rather than the fixed 30 m), for historical (1985-2014) (a), and anomalies (2071–2100 relative to 1985–2014) (b).

Technical comments

l. 105 what do you mean by “equatorward and downward component of surface wind stress”? Did you just mean equatorward?

In the model outputs, tauv and tauu are officially defined as ‘surface downward northward stress’ and ‘surface downward eastward stress’, where the specification ‘downward’ is used to specify that the stress is applied from the atmosphere to the ocean.

l. 110, 207, 209 and elsewhere “nitrate” is a more common spelling than nitrates.

Yes, we will modify the manuscript.

l. 227-229 it would be useful to refer to Fig. 5 right column (that supports this statement)

Yes, we will modify the manuscript.

l. 354 cite the Biogeoscience paper rather than Biogeoscience Discussions

Yes, we will modify the manuscript.

<u>References:</u>

Ryan-Keogh, T. J., Thomalla, S. J., Chang, N., & Moalusi, T. (2023). A new global oceanic multi-model net primary productivity data product. Earth System Science Data, 15(11), 4829-4848.

Garcia, H. E., Weathers, K. W., Paver, C. R., Smolyar, I. V., Boyer, T. P., Locarnini, R. A., ... & Reagan, J. R. (2019). World Ocean Atlas 2018, volume 4: Dissolved inorganic nutrients (phosphate, nitrate and nitrate+ nitrite, silicate).

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Laufkötter, C., ... & Séférian, R. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. Nature Climate Change, 7(5), 355-358.

Ryan-Keogh, T. J., Tagliabue, A., & Thomalla, S. J. (2025). Global decline in net primary production underestimated by climate models. Communications earth & environment, 6(1), 75.

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., ... & Yamamoto, A. (2020). Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. Current Climate Change Reports, 6(3), 95-119.

Jacox, M. G., Edwards, C. A., Hazen, E. L., & Bograd, S. J. (2018). Coastal upwelling revisited: Ekman, Bakun, and improved upwelling indices for the U.S. West Coast. Journal of Geophysical Research: Oceans, 123, 7332–7350.

Ding, H., Alexander, M. A., & Jacox, M. G. (2021). Role of geostrophic currents in future changes of coastal upwelling in the California Current System. Geophysical Research Letters, 48(3), e2020GL090768.

Veitch, J., Penven, P., & Shillington, F. (2010). Modeling equilibrium dynamics of the Benguela Current System. Journal of Physical Oceanography, 40(9), 1942-1964.