Author's response:

We thank the reviewer again for the useful comments, and have modified the document following the suggestions. Below are the authors' point-by-point responses to the comments, with reviewer's comments in bold and author's responses in normal font.

This new version of the paper by Cioffi et al. represents a significantly improved version relative to the first version. I thank the author for considering my suggestions and implementing them, in particular for expanding the model validation section. I have a few remaining minor comments:

-validating wind stress and nitrate in the models is very good; however, I don't think using ERA Interim and WOA 2018 is appropriate. These products have both been replaced by improved products for years, the first by ERA5, the second by WOA 2023, so these should be used instead. ERA5: https://climatedataguide.ucar.edu/climate-data/era5-atmospheric-reanalysis WOA 2023: https://www.ncei.noaa.gov/products/world-ocean-atlas Perhaps the choice to use ERA-Interim and WOA18 was because of available products when the CMIP6 models were produced, and/or were products used to calibrate the models? Regardless, the paper considers these products as the "truth" (similar to NPP) so the most realistic products should be used regardless of which products were used to constrain the CMIP6 models.

We agree on the use of ERA5 and WOA 2023 for the wind stress and nitrate validation, and have updated our analysis with these datasets. Results are comparable to the previous version.

-Regarding my previous suggestion to look into narrowing the projection uncertainty by considering the accuracy of each CMIP6 model in representing historical NPP, I understand that a more detailed analysis would be beyond the scope of the paper and not be fully in line with the objective of your study. I would still suggest adding a few sentences summarizing why restricting models to the most realistic ones during the historical period is not straightforward and/or does not help constrain the projection uncertainty, as other readers may have the same question (especially since Rev #1 had similar questions such as "For instance, does CanESM5 already exhibit biases in simulating NPP in the CalCS during the historical period?"). As a suggestion, without doing any additional work, it could be worth pointing out that which models are outliers for projections is strongly dependent on the EBUS being considered (Fig. S5 – eg CanESM5 for CalCS and CanCS, vs CanESM5-CanOE for HumCS and BenCS, etc). It would also be interesting to evaluate if the outliers in Fig. F and G (Fig 2, S2 and S3) correspond to the models being outliers in Fig. S5, and state the result in text to make the point that constraining projections based on performance during the historical period is not straightforward.

We agree that it would be helpful to explain why it is not straightforward to constrain the ensemble based on historical performance to narrow projection uncertainty. Thus, we have

added a paragraph in the ‘Model evaluation’ section 3.1 to justify our choices. We have also updated all box-and-whisker plots, adding markers and a legend to make the outliers visible.

-l. 295 there is one “nitrates” remaining

- We have modified the manuscript.