We thank Reviewer #4 for their constructive comments. We have listed their comments in bold below and our responses in normal formatting.

***Reviewer #4***

**First, I am concerned that biases in the circulation model might skew the results and have not been properly acknowledged. Can the authors state whether they are using a new version of the MITgcm 2.8degree circulation model, or the same one that has been used since the early OCMIP era? Previous studies (Dutay et al. 2002; Doney et al. 2004) have identified some significant shortcomings of this circulation model that might impact the relative importance of different regions in the current study. Not least, the model does not produce deep water along the Antarctic coastline as it should, and instead produces deep water at around 50S. This would shift deep water formation from the Antarctic to the Subantarctic regions defined in the current study, and give the Subantarctic region unrealistic leverage over interior nutrient distributions. It could be the case that the circulation has been reformulated since those studies and this bias corrected. If that's the case it is important for the authors to demonstrate this, to reassure readers like myself who have reservations about that model. The simplest way to show this would be to calculate ideal age in their model, and plot a meridional cross section through the Pacific. They should be able to show a tongue of young water subducting right along the Antarctic coastline and spreading northwards along the seafloor (not a tongue of young water penetrating the deep ocean at 50S). If they are indeed using the old, biased circulation model, this should be acknowledged in the text where the significance of the Subantarctic region is discussed. Either way, a figure like the one I suggested should be included as a supplementary figure either to demonstrate that the circulation model is robust, or to make readers aware of potential biases introduced by the Southern Ocean wartermass structure.**

Thank you for highlighting this important caveat. We have plotted the meridional cross section (Figure 1) and the model does subduct water around 50°S. In response to this, and to comments from other reviewers, we have included this figure, a comparison of where the densest surface waters are versus observations, and a water-mass analysis in the supplementary material. We have also updated the Discussion in the manuscript to make clear that this is an important caveat. We have kept the circulation model for a number of reasons: 1) the Subantarctic regions do not dominate the sensitivity at a global level, 2) the circulation model is likely to over-estimate the sensitivity of $CO_2$ to remineralisation in the Subantarctic regions, 3) it is widely used for modelling biogeochemistry.

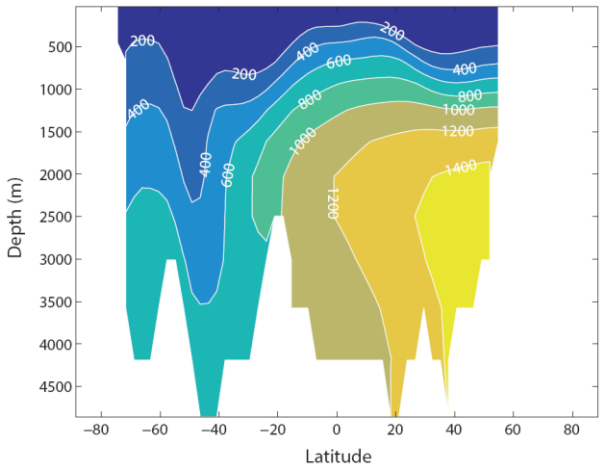


Figure 1. Meridional cross section of ideal age in the Pacific (224°W).

**Second, I am confused as to why a paper focused on sensitivity of atmospheric CO2 does not use a model that resolves the carbon cycle. Instead, they model only the phosphorous cycle and relate it to carbon cycle changes using a relationship derived from prior modeling studies. The authors state that this is to avoid the computational expense of simulating the carbon cycle. But that would only require the addition of two tracers – DIC and Alk, and a single value for a well-mixed atmospheric CO2 concentration. This should therefore only double the computation time, and given that transport matrix method is being used (where efficient Crank-Nicolson time-stepping methods can be applied), this does not seem preclusive. And even then, they needn't include the carbon cycle in all 200 of their simulations, only enough to redefine the statistical PO4 vs. CO2 relationship from their own model. This would at least keep their study self-consistent, rather than relying on previous results from different models. If the authors are not able to do this in the current study (which would be preferable), they should again acknowledge more clearly the caveats of their chosen method. In Fig. A1, it is obvious that different models yield different relationships between these properties. Fitting just the Marinov et al. results would lead to a much shallower relationship, but those results are not strongly weighted because they contribute fewer data points than others. It would seem more reasonable to fit the relationship for each previous study separately, and propagate that uncertainty into their CO2 estimates**

In response to this comment, and to comments from other reviewers, we have added the carbon cycle to the model. We have redefined the preformed $PO_4$ and $CO_2$ relationship and have used this to calculate the change in $CO_2$ for the Latin hypercube ensemble.

**Figure 3. I think a bar plot would be better suited to show this data. This is a key result of the paper, and a quantitative comparison between regions and production methods would be simpler in a bar format.**

Thank you for this suggestion. We have added additional panels to the figure showing the sensitivity estimates in a bar format (Figure 2).

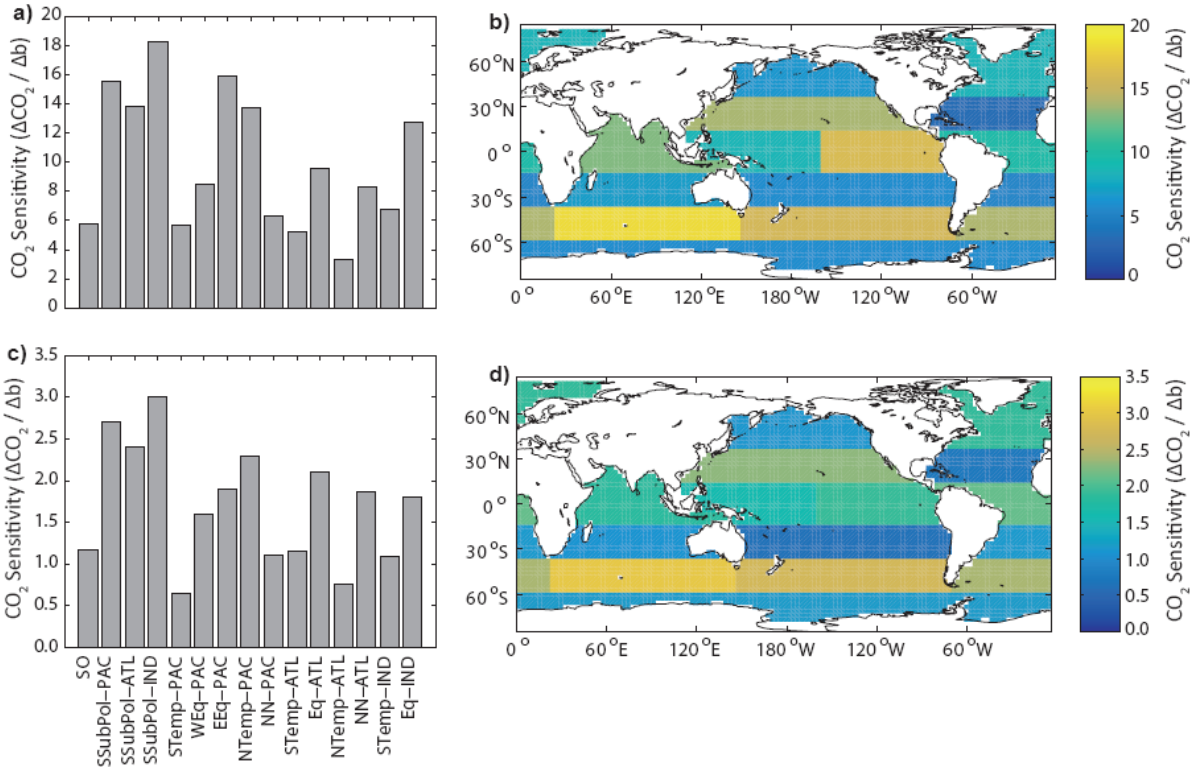


Figure 2. Regional sensitivity ($\Delta CO_2$ / $\Delta b$) of atmospheric $CO_2$ (ppm) to changes in Martin curves for the constant-export scheme (panels a & b) and restoring-uptake scheme (panels c & d). Atmospheric $CO_2$ is inferred from modelled preformed $PO_4$ using empirical relationships.

**Figure 5. This figure is a little overcomplicated, as evidenced by the fact that a fair amount of the text (not just the caption) is devoted to explaining what it means. How about showing these results as a color matrix instead? Region in which b is varied down the rows, region in which we are looking at the preformed PO4 along the columns (or vice versa), color shows the regression coefficient. I know this is contrary to my previous comment about bars being more precise, but I think precision is less important here than the need to reduce complexity. The color matrix would allow the reader to pick out "bright" rows or columns as indicative of important regions.**

Thank you for this useful suggestion. We have reformatted the figure as a matrix (Figure 3). This does highlight spatial patterns more clearly and allows both constant-export and nutrient-restoring results to be shown.

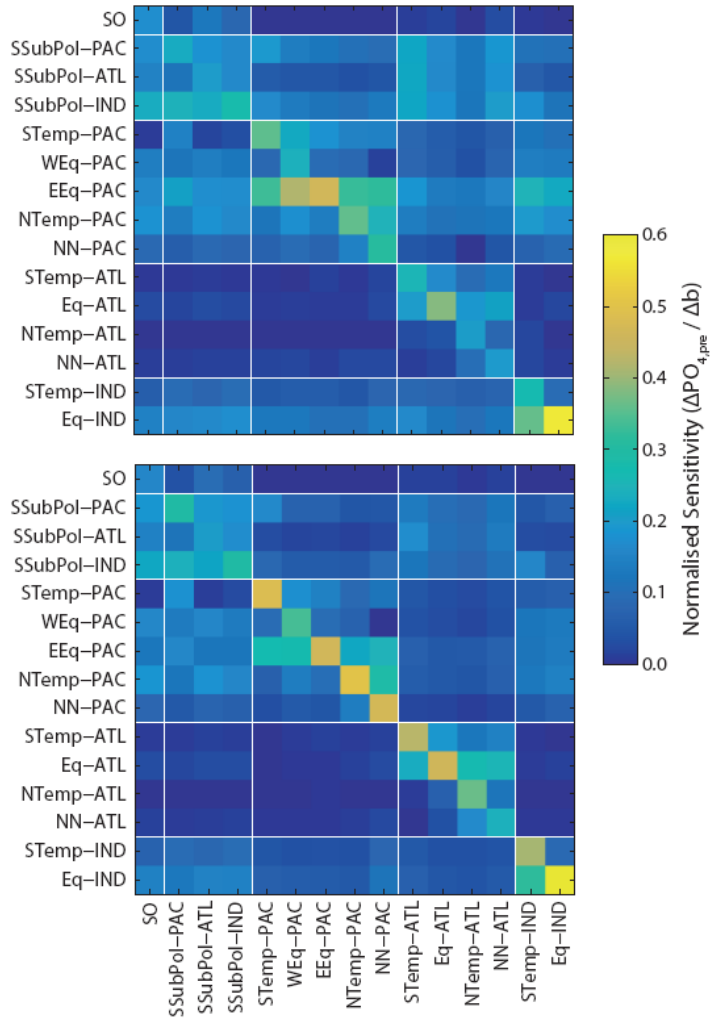


Figure 4. Sensitivity of steady-state normalised mean preformed [$PO_4$] exported from each region. The preformed [$PO_4$] from each region is expressed as a function of b using linear regression. Preformed [$PO_4$] is normalised to the range of values in the ensemble to account for large differences in preformed [$PO_4$] between regions. The regression coefficients are arranged such that each row shows the impact of changing b in that region on preformed [$PO_4$] across other regions. Results from the constant-export and nutrient-restoring schemes are shown in the top and bottom panels respectively.

**Section 3.2 and Figure 6. I'm not sure why the results in Figure 6a are shown? The authors acknowledge that simply averaging the b values is not the correct way to quantify the global-mean remineralization profile, and then attempt to correct for it in panel b. But why show an obviously incorrect result in the first place? It seems like the correct way to define a "global mean b-value" would be to construct a global-mean (area-weighted) organic matter flux profile, and then fit the Martin relationship to that.**

We have moved panel a to the discussion of averaging in the supplementary material. We have kept the averaging approach as this has been used previously, e.g., Henson et al., (2012), and so provides useful context.