We thank Reviewer #3 for their constructive comments. We have listed their comments in bold below and our responses in normal formatting.

Reviewer #3

While I understand why the two model scenarios (restoring or fixed export) are presented as end-members, the fixed export run nevertheless takes its export from a restoring run. It is true that output from the run giving the closest fit to observations is used as baseline but it should still be acknowledged that the 'end-members' are far from independent models.

We have amended the text as follows: "These two schemes represent two end-member scenarios, strictly within the context of this model, where organic matter production either depends entirely on macronutrient concentrations…"

The description of tracking preformed phosphate needs more detail. The decomposition described in Appendix B gets phosphate away from surface only – it still needs to be tracked in the interior. How is this done?

We have changed equation B1 in Appendix B to make the operation clearer:

$$PO_4^{pre} = (A_i^I A_e^I - I)^{-1}((A_i^I B_e + B_i)PO_4)$$

The authors should show the scatter plot of predicted vs observed values for the relationship described in page 5 lines 12-13 as it is fundamental to the manuscript. It should show predicted and observed changes in PO4 as this is the predicted field.

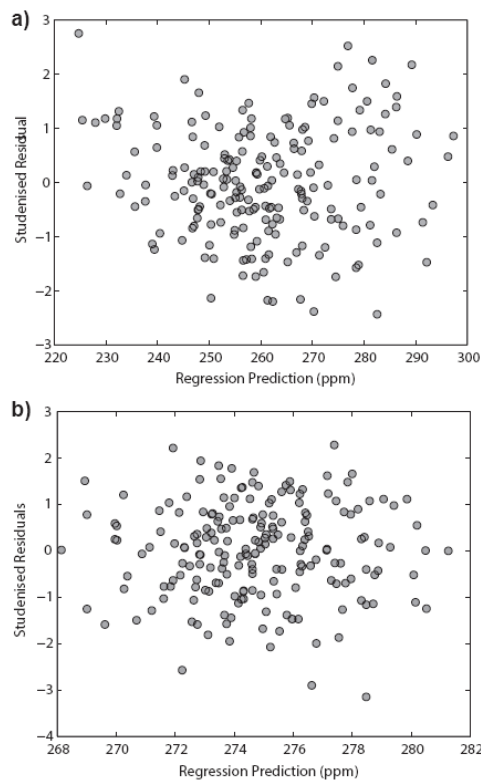We have added this figure to the manuscript, (see Figure 1 here).

Figure 1. Residuals for the linear regressions that estimate sensitivity of CO_2 to spatially varying Martin curves for (a) constant-export and (b) restoring-uptake schemes.

Fig 6 and section 3.2 – there is a sound argument for geometric mean so just show geometric mean and give the argument in the methods. It is not necessary to show arithmetic mean results in Fig 6a

We have moved this figure panel to the discussion of calculating the geometric mean in the supplementary material.

As a more informative second panel for Fig 6 show the same as current 6b but with regression taken out to show variability due to regional variability more clearly. The authors should also acknowledge in the text that the random sampling leads to undersampling of highest and lowest global b values.

Thank you for this suggestion. Because the global mean of responses track the globally-uniform responses closely and we could not find any evidence that the variability was associated with changes in b in any specific region, this additional plot did not provide much additional information so we have kept the original panel.

We have added the following text: "Note that b in each region is varied within the full parameter range but that because Latin hypercube sampling varies all parameters across their parameter range simultaneously the global mean does not reach the highest and lowest global b values."

How independent in structure are the 3 models used for the PO4 vs pCO2 relation?

In response to other reviewer comments we have replaced the statistical relationship between preformed PO_4 and CO_2 with one calculated specifically for this model using a carbon cycle.

Fig A1 should be in the main body of the paper

We have moved the equivalent plot to the Methods section of the manuscript.

Consistency needed in terminology: in Subantarctic (text) and subpolar (fig)

We have changed any use of 'subpolar' to 'Subantarctic' throughout.

Remineralisation depth is defined (page 2, lines 8-9) assuming exponential profile (decrease by 63%) but models use Martin curve

The reviewer is correct that the definition assumes exponential decay whereas the Martin curve is a power-law. This was used previously by Kwon *et al.,* (2009) who used Martin curves but also expressed them as e-folding depths. Our purpose was to introduce the term 'remineralisation depth' as this allows for more clear and concise discussion of changes in Martin curves as the terms 'shallower' or 'deeper' can be used rather than changes in the dimensionless exponent b.

We have changed the text to better reflect this comment:

"In this paper we use the term 'remineralisation depth', defined as a depth at which a defined % of POC has been remineralised. Previously, this has been defined as an *e*-folding depth: the depth at which ~63% of POC has been remineralised (Kwon et al., 2009) (although note the Martin curve is not exponential)."

Does the misfit function used to carry out the comparison to WOA (page 4, lines 22-23) take volume into account?

Yes. We have amended the text to state that it is volume-weighted.

Explain the maximin Matlab option for hypercube sampling in Matlab (page 4, line 31)

We have added the following text to clarify:

"…with 'maximin' sampling (an additional constraint that helps reduce clustering of samples, by maximising the minimum distance between points, in order to give a well-spread distribution of points across the parameter space)."

page 4 line 25: not sure that "reference" is appropriate

We have updated the experiment description with headings to separate the description of the control run, global and regional sensitivity runs. "Reference" has been removed from the text.

Fig 3 caption needs rewording. All values are positive.

The caption has been reworded to: "The sensitivity value reflects the increase in CO_2 (preformed PO_4) for an increase in b (shallower remineralisation)."

The authors' definition of the Subantarctic boundary makes it a little difficult to compare results to Kwon's paper where the Southern Ocean was defined as south of 40S. Given that the Kwon paper provides such strong motivation for this manuscript this deserves comment.

We had added an additional row to Table 1 in the manuscript describing metrics for the Southern Ocean as defined as >38°S for comparison with Kwon et al., (2009). We have also noted this comparison in the Discussion.

Page 6, lines 4-6: It should be explicitly acknowledged that there is a rather weak relationship between export and sensitivity for the restoring runs (Fig 4b)

We have added correlation coefficients to help demonstrate the weaker relationship between export and sensitivity. The following text has been added:

"Similarly, we find a general positive correlation between sensitivity and regional export production (r=0.79, p<0.01 for constant export, r=0.47, p=0.07 for restoring uptake), as measured by the mean annual average export production across the 200 ensemble runs (Fig 4). The correlation is much weaker with nutrient restoring uptake compared to the constant-export production."

Use notation that distinguishes regional and global means of PO4_pre

We have updated the text with notation to distinguish between regional and global means of preformed PO_4.

Both constant export and nutrient restoring should be shown in Fig 5.

In response to other comments from reviewers, we have updated the format of this figure and have added panels for both constant export and nutrient restoring.

Page 7, Line 3: "sensitivity"

Fixed.

Page 8, lines 28-30: "As such, the global mean change in potential future and past changes in remineralisation depth may be larger than the uncertainty associated with spatial variability." B changes discussed less than current observed range" The changes being discussed here are substantially smaller than the current range of observed values. Even if, as this paper argues, the global ocean may not be overly sensitive to spatial variation in b, it is worth noting that the current uncertainty in a global value of b still has very large uncertainty partly because of the confounding effect of under-sampled spatial variability

We have added the following text: "However, we note that the modern global mean b is subject to uncertainty associated with under-sampled spatial variability."

Page 10, line 21: Which sea ice field is used?

The following text has been added:

"…scaled the fraction of seaice present (Fice, as monthly average fields from the original circulation model)."

Page 11, line 5: 1-v not 1-kappa

Fixed.

Appendix A: state that the bottom of the second grid box in the vertical is at 120m (presumably)?

Done.