

Interactive comment on “A latitudinally-banded phytoplankton response to 21st century climate change in the Southern Ocean across the CMIP5 model suite” by S. Leung et al.

Anonymous Referee #2

Received and published: 7 July 2015. Referee comments in italics. Our comments in regular text.

General comments:

1.) First, while the authors provide a correlative justification for mechanistic causation in each case, it is unclear how robust each of these explanatory factors might be... without conducting sensitivity tests with each model, how do the authors know for sure that their inferred drivers are true the causative one?

It should be made more explicit that 1) these inferred linkages are speculative based on correlation, 2) the magnitude of changes in the proposed factors are hypothetically large enough to drive most of the change, and 3) discussion of potential alternative explanations.

Good points.

We added to the text in Section 3.2.1: “It is important to note here that these inferred linkages are based only on correlations, but in all cases are also supported by model equations and previous studies.”

And in Section 3.2.2: “Importantly, the magnitude of 100-year changes in the chosen variables of interest are also hypothetically large enough to drive most of the 100-year change in PB. We note, however, that in the real ocean, phytoplankton adaptation and evolution could alter the driver-response relationship observed at the interannual scale within these models.”

And in the Conclusions section: “It is important to note that the relationships between phytoplankton responses and their potential drivers discussed here are based on correlative analysis and thus do not perfectly prove causation. It is promising, however, that in all cases the significant and most strongly correlated phytoplankton and potential driver relationships matched with expectations based on both previous studies and model equations.”

2.) Second, the use of ‘sign of change’ is potentially very misleading to a community that is accustomed to changes being expressed as integrated anomalies - I am very concerned that readers will interpret a null result that 50% of pixels increase and 50% of pixels decrease as alternatively that the mean value increased (or decreased) by 50%... there needs to be a more thorough statistical justification for this approach based on the result of a null test where nothing changes.

We see now that we did not explain the bootstrap technique and figures associated with it sufficiently.

To address this, we began by rewriting our methods section describing the bootstrap technique. In this

new text, we also discuss why we believe that the bootstrap method is a good technique.

Our rewritten longer bootstrap methods section is as follows: “

“To quantify the significance of multi-model mean 100-year trends, we calculate the percentage of simulated model realizations that agree on the sign of a predicted trend for a given variable, using the statistical technique known as bootstrapping. We built 1,000 realizations of the 100-year trend by randomly selecting n models (where n is the number of models with data available for any given variable) with replacement among the n available models. Within a single realization, one model may be represented more than once, while other models may not be represented at all. We take into account interannual variability by randomly selecting one of the 20 years from the present-day *historical* scenario (1980-1999) and one of the 20 years from the future *rcp8.5* climate change scenario (2080-2099) for each selected model. For every variable of interest at every spatial grid point, we then create a realization of the 100-year trend by finding the difference between the two randomly chosen years. We then obtain the multi-model significance of this trend at each grid point by calculating the percentage of 1,000 realizations that predict a positive change. Thus, the higher (lower) the bootstrap percentage above (below) 50%, the greater the significance of the positive (negative) trend at a given location. This bootstrapping procedure provides a more robust measure of significance than simply calculating the percentage of models that agree based on single model runs alone because it both takes into account interannual variability and greatly increases the number of permuted realizations. See Cabré et al. (2014) for further details on application of the bootstrapping method to the CMIP5 dataset.”

The next thing we did to clarify bootstrapping was to make our old Fig. S16 into a main figure (new Fig. 5). These maps show the percentage of model realizations in agreement with one another at each grid point.

We now describe this figure in the text as follows: “To get a wider sense of spatial agreement among models throughout the SO, we look at maps of intermodel consistency in projected SO phytoplankton trends and their proposed drivers across all 16 CMIP5 models with ocean biogeochemistry in Fig. 5 (complementary to Fig. 1). The maps in Fig. 5 detail the fraction of model realizations (via the bootstrap technique explained in Section 2.2 above) that predict a positive trend in the listed variable at each grid point. Thus, the closer the fraction to one at a given location, the greater the intermodel agreement on a positive trend at that point, and the closer the fraction to zero, the larger the intermodel agreement on a negative trend at that point (0.5 denotes the greatest amount of intermodel disagreement, where 50% of model realizations predict an increasing trend and 50% predict a decreasing trend).”

We then explain in the text that the percentages listed above each zonal band in old Fig. 5 (new Fig. 6) are just areal averages of the percentages shown in new Fig. 5 over each zonal band. We also clarify the meaning of the arrows as follows:

“To also get a better idea of how well models agree with one another within each zonal band, Fig. 6 shows zonally averaged all-model mean projected trends (zonal averages of Fig. 1) and zonal band averaged intermodel agreement percentages (areal averages over each zonal band of Fig. 5, listed above each zonal band accompanied by an arrow indicating the direction of the trend agreed upon by the majority of model realizations). Only percentages for variables which are most important within each zonal band (as determined by Figs. 1-4 above) are listed and as such, represent a summary of the important drivers of projected phytoplankton change discussed here. Agreement among models is highest at the center of each zonal band (Fig. 5), but decreases towards the edges due to offsets in the

precise boundaries of water masses among the models. These slight offsets lower the zonal band-average agreement among models shown in Fig. 6, such that if one were able to perfectly compare water masses among models, consistency in predicted trends within each zonally-banded biome would likely be even higher.”

We also clarified Fig. 6’s caption with the following red text:

“Summary of predicted phytoplankton responses and their drivers within each zonal band. **Here each colored line represents normalized 100-year all-model mean zonal changes in the listed variable.** Each variable was normalized by first computing the all-model mean zonally-averaged 100-year change at every latitude and then dividing by the absolute value of the largest of these changes occurring south of 30°S. **Listed above each band is the spatially averaged percentage of model realizations that agree on the sign of the change (based on a bootstrap analysis – see Sect. 2.2 and Fig. 5) in each variable over that band. The colored arrows denote the direction of the trend agreed upon by the majority of models.** The number of models (n) and the total model weight (w) taken into account for each variable are listed in Fig. 5.”

3.) Third, how strong are any of these changes? There should be a figure quantifying the magnitude of the change in each band to compliment figure 5 that currently gives the percentage of pixels that agreed on the sign of the change.

Old Fig. 5 (new Fig. 6) did not give the percentage of pixels that agreed on the sign of change, but rather the areally averaged percent of model realizations that agreed on the sign of change over each zonal band. See the clarifications discussed in point 2 above.

We added Fig. S17 showing the magnitude and sign of zonally averaged 100-year changes in each variable alongside their *historical* means. For a wider spatial view of 100-year changes, see Fig. 1. For a wider spatial view of *historical* means, see Figs. S1-10.

4.) Finally, the purpose of the section on SAM observations seemed out of context and potentially in conflict with another paper in preparation. I recommend the section be better integrated into the current paper or removed.

We find this section to be central to our study because we would like our work to be linked to and put into the context of recent observational work on SO phytoplankton trends and potential climate change effects. Please note that this section discusses SO phytoplankton and physics observations in general, not just those specific to SAM. In particular, we discuss only 2 sources in regards to SAM correlations. All of the other studies we cite/analyze look at observations of SO phytoplankton trends over different regions and time periods.

As we mention in the text, the SO satellite chl record is not yet long enough to separate the effects of climate change from those of interannual processes driven by the leading modes of shorter-term variability in the SO, in particular SAM. For these reasons, many observational studies have looked at the effects of SAM and other modes of variability, rather than climate change, on phytoplankton abundance and productivity. These types of studies can, however, still provide essential insight into the mechanisms driving possible longer-term changes. We show in this paper that - at least within the CMIP5 models - mechanisms responsible for changes in phytoplankton biomass on interannual and five-

year time scales are also responsible for projected 100-year trends within the SO. Thus, understanding the effects of a more positive SAM on SO phytoplankton may help predict the direction of phytoplankton changes in a warmer future climate.

Indeed, one of the main points of the paper is that 21st century trends in phytoplankton productivity predicted by the CMIP5 models go in the same direction as observed trends over the last couple of decades and tentatively agree with the sign of previously established SAM-driven changes.

To clarify the role of SAM in changing the physical drivers of phytoplankton ecology, we have added the following red text to Section 3.1:

“These abovementioned factors are proximate physical and biogeochemical drivers of predicted phytoplankton responses within the models, but what is the ultimate driver of all of these physical and biogeochemical changes? **Historical and projected 21st century increases in the strength of the principal mode of variability in the SO, called the Southern Annular Mode (SAM), due to a combination of elevated CO₂ concentrations and ozone depletion could be one explanation.** One highly agreed upon dynamical change captured within all of the CMIP5 models analyzed here is an intensification and poleward shift of the SO westerly wind belt (Fig. 1h; Fig. S9) associated with an increasingly positive phase of the SAM with future warming, as seen both here (Fig. S11) and in previous work (e.g., Yin, 2005; Arblaster and Meehl, 2006; Russell et al., 2006; Gillett and Fyfe, 2013; Zheng et al., 2013). This highly consistent increase in wind stress (which is most pronounced in the summer – plots not shown) south of 50°S may explain the deepening of summertime MLDs south of 50°S, while the decrease in wind stress between 30°S and 50°S may explain the shoaling of summertime MLDs in that region (Fig. 1d; Fig. S5). These changes in MLD can then affect nutrient supply to the surface, perhaps leading to the large decreases in surface nitrate concentrations between 30°S and 50°S (Fig. 1c; Fig. S3).”

It is also important to discuss how to potentially apply global model predictions to better understand and confirm in-situ observations. We attempted to make this purpose and point of the observations section clearer by adding the following two blocks of text:

- 1.) “Because the same interannual mechanisms for phytoplankton growth hold on 5-year, decadal, and even longer-term timescales within the CMIP5 models, it is reasonable to compare recent observations to future model projections if it is also assumed that short-term drivers of observed phytoplankton variability propagate up to longer-term timescales in the real ocean as well. However, it is out of the scope of this paper to compare recent observations to *historical* model output from the same period. Instead, we would like to understand how our modeled 21st century SO predictions compare to observed mechanisms and trends thus far.”
- 2.) “In sum, the observed spatial distribution in trends of phytoplankton productivity, MLD, and cloud cover over the past few decades qualitatively matches the latitudinally-banded structure of the respective 100-year 21st century trends predicted by the CMIP5 models. We have found that (a) in CMIP5 simulations, interannual effects propagate up to 100-year timescales and (b) drivers for short-term biomass change are similar in models and observations within individual zonally-banded biomes. If the CMIP5 model mechanisms and projections are to be trusted, then this suggests that observations may already contain a climate change signal even though this signal cannot be teased apart from decadal variability and shorter-term noise just yet (e.g., Henson et al., 2010).”

Specific comments:

1.) pg 8161, ln 3 - *Distribution misspelled*

Fixed. Thanks!

2.) pg 8161 - *It is currently not clear how much of the results are novel relative to the Cabre et al 2014 paper. Given the overlap in topic and methods, the analysis overlap and distinct contribution of the present work should be made explicit.*

This paper is novel work not at all touched on in Cabre et al. (2014). What we take from our previous work in Cabre et al. (2014) is a general understanding of ecological modules in CMIP5 models, background knowledge of certain global model properties, the weighting scheme for intermodel comparison, and the bootstrap technique to assess significance of trends across models. The present study, however, focuses specifically on the Southern Ocean (while Cabre et al., 2014 separates the entire global ocean into traditional biomes thus ignoring the finer-scale, important banded patterns within the SO region discovered here) and looks at interannual variability and phytoplankton drivers that work on different timescales (while Cabre et al., 2014 ignores these topics entirely). Furthermore, comparison to observed trends and mechanisms related to biological change in the SO banded structure is also lacking in Cabre et al. (2014), while it is one of the main sections here.

To clarify what we take from the Cabre et al. (2014) paper, we added the following to the introduction: “To this end, we borrow some statistical methods developed in Cabré et al. (2014) (namely, the model weighting scheme and the bootstrap technique, both described in Section 2 below) to conduct our work.”

You can find our Cabre et al (2014) paper at <http://link.springer.com/article/10.1007/s00382-014-2374-3>.

3.) pg 8162 - *I am not sure if the application of the bootstrapping is mechanistically justifiable. While I appreciate that this method bring correlation analysis to a more sophisticated statistical level, I have a hard time appreciating what the mechanistic significance of the signs being correlated 64% of the time and anticorrelated 36% of the time... Which of the two cases is more relevant to a causal mechanism can only be determined by either direct sensitivity tests or scaling via simplified idealization.*

We use the bootstrap technique to assess the significance of trends only, not of correlations.

As was mentioned above, we see now that we did not explain the bootstrap technique very well, so we rewrote the methods section to explain it better. Here is that text, which is repeated from general comments, point 1 above:

“To quantify the significance of multi-model mean 100-year trends, we calculate the percentage of simulated model realizations that agree on the sign of a predicted trend for a given variable, using the statistical technique known as bootstrapping. We built 1,000 realizations of the 100-year trend by randomly selecting n models (where n is the number of models with data available for any given variable) with replacement among the n available models. Within a single realization, one model may be represented more than once, while other models may not be represented at all. We take into account interannual variability by randomly selecting one of the 20 years from the present-day *historical*

scenario (1980-1999) and one of the 20 years from the future *rcp8.5* climate change scenario (2080-2099) for each selected model. For every variable of interest at every spatial grid point, we then create a realization of the 100-year trend by finding the difference between the two randomly chosen years. We then obtain the multi-model significance of this trend at each grid point by calculating the percentage of 1,000 realizations that predict a positive change. Thus, the higher (lower) the bootstrap percentage above (below) 50%, the greater the significance of the positive (negative) trend at a given location. This bootstrapping procedure provides a more robust measure of significance than simply calculating the percentage of models that agree based on single model runs alone because it both takes into account interannual variability and greatly increases the number of permuted realizations. See Cabré et al. (2014) for further details on application of the bootstrapping method to the CMIP5 dataset.”

4.) pg 8162 - When assessing sign of change, did the authors apply a magnitude threshold of significance, or just much at the sign independent of magnitude?

We did not apply a magnitude threshold of significance per individual model because the magnitude of change is taken into account in the bootstrap analysis. Using this technique, we are constructing a large distribution of 100-year changes and then asking what percentage of that distribution is greater than zero (indicative of a positive trend). You can then decide a threshold of significance from there by choosing what percentage of model realizations need to agree to meet your criteria. In Fig. 1, we chose to hatch areas where >80% of model realizations agree on the sign of the trend (both increasing and decreasing trends, so bootstrap percentage <20% or >80%).

We have added and edited text as was described in general comments, point 2 and specific comments, point 3 to hopefully better explain the bootstrap procedure.

5.) pg 8163 - The discussion of all the figure specifications is typically limited to the discussion of the figures themselves either in the text or captions rather than separately in the methods before the figures are introduced. I recommend moving this section.

Good point. We moved sections 2.5-2.7 into the figure discussion paragraphs as suggested.

6.) pg 8166-8167 - The discussion of different mechanisms operating on different timescales is intriguing but speculative, incomplete and not well justified.

We tried as best as we could to thoroughly validate and explain each decision we made (masking, methods of detrending, CMIP5 scenario selection, variable selection, model selection, etc.). We think that it's important to show that if a particular mechanism associated with climate warming is thought to be the driver of some change on 100-year (or so) climate-driven timescales, then it should operate on decadal and shorter timescales as well. In other words, the statement that a particular change in a variable is contributing to or causing the change in another in the context of climate warming becomes more robust and believable if this cause-and-effect relationship holds on shorter (and perhaps even longer) timescales, too.

We feel that past work has not done enough to approach the climate change attribution problem from this necessary angle. Many studies have looked at drivers either on interannual timescales alone or climate change timescales alone. To form a more complete picture as well as a more coherent argument (as the interannual and longer-term timescales inform each other), we believe that the two should be

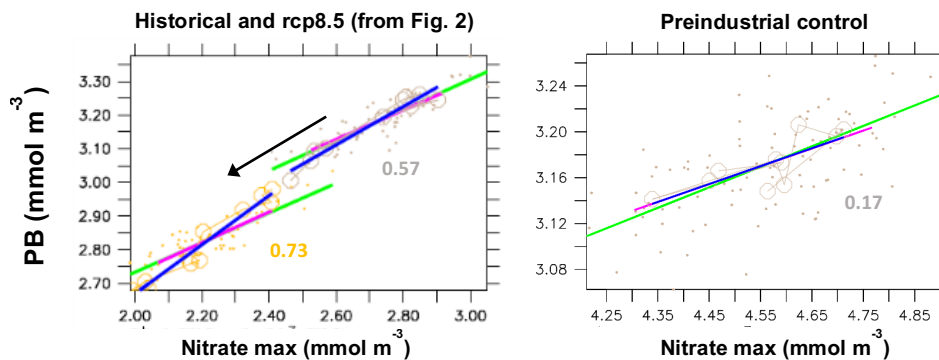
done in tandem. We further argue that this methodology should be applied more often, both in model and observational analysis.

We recognize that an investigation of preindustrial control short-term variability was missing from our discussion, however, so have added a sentence about this, as follows:

“Here we purposely chose to use detrended *historical* scenario time series rather than *preindustrial control* scenario time series (forced with constant preindustrial CO₂ concentrations) for practical reasons (not all the models provided all the necessary variables in the *preindustrial control* experiment). We did, however, prove that in at least model GFDL-ESM2G, the interannual drivers affect phytoplankton biomass in the same direction and with a similar magnitude in the *preindustrial control* case and the detrended *historical* and *rcp8.5* cases, as expected.”

Below is an example of one of the plots we looked at to conclude that detrending *historical* and *rcp8.5* time series would give us similar answers compared to looking at the *preindustrial control* time series.

Within GFDL-ESM2G’s 30-40°S band (see Fig. 2 legend and caption):



7.) pg 8167 - The iron supply mechanism appears suspect - do the authors know that supply increased. Or are they inferring this from concentrations. At least some (all?) of the models considered assumed fixed climatologies of iron supply. My guess is increased salinity stratification leading to shallower mld south of 50.

Typically, it is only atmospheric iron supply that is fixed (though sometimes the sedimentary source is fixed as well), which is not related to iron supply from the deep ocean or lateral advection.

We added the following to the end of Section 3.1 to clarify these possibilities:

“As for the ultimate driver of increases in surface iron concentrations, which contribute to increases in PB and PP in the Transitional (~40-50°S) and Antarctic (south of 65°S) bands, there may be other complicating factors at work. Parameterizations of the marine iron cycle differ from model to model and include processes such as atmospheric dust deposition, phytoplankton-community dependent biological uptake and remineralization, vertical particle transport, scavenging, and the release of iron from sediments (e.g., Moore et al. 2013b). While atmospheric dust deposition is kept constant in the CMIP5 simulations, other processes listed above may change, thus altering surface iron inventories. For example, the increase in iron in the 40°-50°S Transitional band can be explained by enhanced vertical supply due to deeper wintertime mixed layers (Fig. S4) or by increases in summertime water column stratification, which can trap and concentrate iron deposited from the atmosphere closer to the surface. On the other hand, Misumi et al. (2014) showed that in the CESM1-BGC model (*rcp8.5* scenario), a

southward expansion of the subtropical gyre and changes in low-latitude iron utilization resulted in increased lateral advection of iron into the SO over the 21st century. Another potential iron enhancing mechanism in the SO is increased release of iron from sediments, a mechanism important within at least the GFDL models (J. Dunne, private communication).”

8.) pg 8168 - *In the interannual variability discussion, the authors should note that the models are 'perfect' integrators in their annual time averages while observations are a collection of snapshots and the implications for interpretation of the bootstrapping.*

We agree that it should be noted that observational data may be highly discontinuous and filled with gaps, but felt that it was more appropriate to add this point to the observations section.

We thus edited/added the following blocks of text to Section 3.4:

- 1.) “Furthermore, while models generate perfectly continuous data, observations tend to contain many more gaps, such that a longer observational time series is needed to detect significant trends compared to model data when the same threshold of significance is applied.”
- 2.) “With such short **and discontinuous** observational records, our model-observational data intercomparison is clearly only qualitative at this point in time. We advocate for **longer and more continuous** in-situ phytoplankton biomass and satellite chl data collection in this important but massively under-sampled region of the ocean to allow for the emergence of a climate change signal from short-term variability.”

We are not sure what is meant by “implications for interpretation of the bootstrapping”. We hope that the clarifications and improvements to the text discussed in general comments’ point 1 and specific comments’ points 3-4 have answered the reviewer’s concerns.

9.) pg 8171 ln 17, pg 8172 ln 9, pg 8172 ln 13, pg 8174 ln 21, and pg 8180 ln 11 - *Justification that something 'comes from theory' or 'accepted theory' or consistent with general expectation from theory' is totally unhelpful. The particular theory of relevance should be cited and described. The authors seem to think that one should trust that all theories are truth... They are not.*

pg 8171 ln 17 – changed “in agreement with both theory and previous studies” to “in agreement with previous studies”

pg 8172 ln 9 – changed “which could be logically predicted from theory or model equations” to “which could logically be predicted from previously discussed mechanisms or model equations”

pg 8172 ln 13 – changed “because we know from both accepted theory and model equations that an increase in cloud fraction would decrease light availability” to “because we know that an increase in cloud fraction would decrease light availability”

pg 8174 ln 21 – changed “agrees with the general expectation from theory and previous modeling studies” to “agrees with the general expectation from previous theoretical and modeling studies”

pg 8180 ln 11 – changed “in line with theoretical expectations and previous modeling studies” to “in line

with previous studies”

10.) Throughout - The nine instances of 'In order to' should be reduced to 'To'

Done.

11.) pg 8176 ln 4-5 - *The satellite interpretation should be prefaced by 'if the models are to be believed' before saying that the record isn't long enough to see a trend... There may indeed be trends in the satellite record, they just may not be attributed to climate warming as evidenced by the models. The authors are thus expressing a severe overconfidence in the models to the detriment of interpreting observations in an unbiased manner.*

It is clear that both models and observations have deficiencies and we need caution to interpret them, as we mention throughout the text. The one advantage models have (that we now specify in the text) is that they generate perfectly continuous data, while observations tend to contain many more gaps, such that a longer observational time series is needed to detect significant trends compared to model data when the same threshold of significance is applied.

To preface our interpretations of satellite data, we add the following text to Section 3.4: “If the CMIP5 model mechanisms and projections are to be trusted, then this suggests that observations may already contain a climate change signal even though this signal cannot be teased apart from decadal variability and shorter-term noise just yet (e.g., Henson et al., 2010).”

12.) pg 8177 ln 3 to pg 8179 ln 14 - *Instead of reviewing the literature, the authors should only bring in past observational studies to specifically describe how these studies support or refute the expected behavior of the models. In this line, it seems like the authors should be interpreting the models on the same timescale as the observations to illustrate the scope of the signal to noise relative to the 100 year trends.*

We review the literature specifically for the purpose of describing how these studies support or refute the behavior of the models. It would be dishonest and incomplete to purposely leave studies out, so we did our best to assess as many as we could find against our model predictions and mechanisms. Within each zonal band, we summarize all the relevant observational studies that we could find (since many of them only focus on specific regions rather than the entire Southern Ocean) and say how they relate to our model findings. For example, at the end of each paragraph detailing the observational studies conducted within each zonal band, we discuss whether and how the mechanisms proposed in each of the cited observational studies coincided with the ones we discussed here in our CMIP5 model analysis.

We realize that we did not do a good job of explaining why we wanted to compare future predictions to past observations, so we added the following to the text:

At the beginning of Section 3.4:

“Because the same interannual mechanisms for phytoplankton growth hold on 5-year, decadal, and even longer-term timescales within the CMIP5 models, it is reasonable to compare recent observations to future model projections if it is also assumed that short-term drivers of observed phytoplankton variability propagate up to longer-term timescales in the real ocean as well. However, it is out of the scope of this paper to compare recent observations to *historical* model output from the same period.

Instead, we would like to understand how our modeled 21st century SO predictions compare to observed mechanisms and trends thus far.”

At the end of Section 3.4:

“In sum, the observed spatial distribution in trends of phytoplankton productivity, MLD, and cloud cover over the past few decades qualitatively matches the latitudinally-banded structure of the respective 100-year 21st century trends predicted by the CMIP5 models. We have found that (a) in CMIP5 simulations, interannual effects propagate up to 100-year timescales and (b) drivers for short-term biomass change are similar in models and observations within individual zonally-banded biomes. If the CMIP5 model mechanisms and projections are to be trusted, then this suggests that observations may already contain a climate change signal even though this signal cannot be teased apart from decadal variability and shorter-term noise just yet (e.g., Henson et al., 2010).”

13.) 8180 ln 1-5 - If the authors are concluding that SAM is the central mechanism, the authors should be assessing the biomass and productivity changes as a function of SAM in the present paper rather than citing a manuscript in preparation. I wonder if this whole section on SAM (3.4 Linking CMIP5 model projections to observations and figure 6) should be removed.

Here we do not conclude that SAM is the central mechanism, but rather speculate that it could be important given that SAM is the main mode of interannual variability for Southern Ocean physics. We anticipate that readers will want to know about the ultimate driver of all of the biogeochemical changes we study here (as we did), which is why we mention SAM and associated westerly wind intensification/poleward movement. It is out of the scope of the current paper to analyze the SAM-driven effects on phytoplankton; rather this will be the focus of a future manuscript. Please see the detailed discussion under general comments, point 4 above.

14.) pg 8181 ln 7-10 - In concluding that there should be 'at a minimum, one or two representative time series' sites for each of the four Southern Ocean provinces they characterize, the author's are pointing out that the 4 provinces they discern would each require time series sampling for long term monitoring. However, the authors have a lot more specific guidance to offer on this than they currently provide, namely how one should robustly site a 'representative' location for each site... This information is critical for the recommendation to be actionable. Are these sites expected to be spatially fixed? In damping down the noise through the use of ensembles and extremely long time averaging, this work is more of a detectability study than an impacts study akin to what one would expect to derive from satellite or other observed fields with all their caveats. Are observations at time series sites demonstrably robust against all manner of spatial, temporal and causation caveats?

We agree that it is important to note that observations at time series sites may not be robust against all manner of listed caveats (but still do believe that they are one of the best ways to capture different physical/biogeochemical regimes throughout the ocean). We also do not necessarily believe that the sites need be spatially fixed, which was a good point to bring up.

To address these issues, we added the following red text to the Conclusions: “The main result of this study – a consistency of the model-projected phytoplankton trends within 4 distinct SO bands over the 21st century – suggests a framework for selecting a minimum number of sites for future SO biogeochemical observational time series stations **or repeat sampling campaigns**; at a minimum, one or two representative time series are needed from each of the 4 SO bands described here. **These datasets**

(and any observational datasets, for that matter) are subject to all manner of spatial and temporal caveats, but over time and in combination with larger-scale satellite observations, longer-term in situ time series will allow us to distinguish natural variability from the climate change signal and more readily compare observed mechanisms and trends with those predicted by our models.”

15.) pg 8181 ln 18 - The assertion 'Given the fragility of polar ecosystems' here requires more support. What makes polar ecosystems particularly fragile? Is there some particular fragility that is motivating this work? If so, this motivation should also be brought up in the introduction.

Good point. We removed this assertion due to it being not entirely relevant or necessary.

In the text, we changed “Given the fragility of polar ecosystems and the critical importance of the SO in driving global carbon and nutrient cycles as well as low-latitude productivity” to “Given the critical importance of the SO in driving global carbon and nutrient cycles as well as low-latitude productivity.”

16.) Figure 5 - why aren't all the lines plotted for all the latitudes?

It was originally for the sake of clarity so that there were not too many overlapping lines. We felt that it was justified to leave some variables out of some bands because many variables only matter in some small subset of regions (say, nitrate between 30-40°S or IPAR south of 50°S). However, it is more informative and probably less confusing if we leave all of the lines in at all of the latitudes, so we've added them back in.