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 #1

Received and published: 29 June 2015

Referee comments in italics. Our responses in regular font.

1.) Page 8161: Is all of the output used annual mean values only? Or is monthly data used too? Somewhere it would be good to include a list of the CMIP5 variable names used (e.g. intpp, nit etc.). Also, do all models include all variables examined? I guess not, but the information wasn't easy to extract.

To clarify whether monthly or annual output was used, we added to the methods section: "For all model analyses conducted here, we use yearly time series, which was sometimes calculated from CMIP5 monthly output and sometimes taken straight from CMIP5 yearly output depending on availability."

We also added the text in red: "For example, whenever we analyze individual models, we show PB because we frequently only have monthly model output (with which to generate maximum, minimum, or average annual data) ..."

We also added Table 2, which has a list of the CMIP5 variables we downloaded and how we treated them to take them from monthly into yearly time series, and Table S1, which shows which models had data on which variables.

In Figs. 2-3, we added whether it was the annual min, max, or average of the variable that was being shown on the x-axis to clarify that we had used yearly time series for these analyses.

2.) Page 8163, line 1-2: presumably this should be "positive or negative change" – it's the sign of the trend that's important for the bootstrap analysis I think?

We see that our initial explanation of the bootstrap analysis was not clear, so we rewrote Section 2.4 (now Section 2.2).

It now reads: "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 permutated realizations. See Cabré et al. (2014) for further details on application of the bootstrapping method to the CMIP5 dataset."

To address the referee's particular point of confusion, we added: "Thus, the higher (lower) the bootstrap percentage above (below) 50%, the greater the significance of the positive (negative) trend at a given location."

3.) Page 8163, line 25-27: just because 2 models show the same sign of the trend in PP or PB doesn't mean that you are comparing the same water masses. I think this is a fallacious argument and the authors would do better to just use the 'increasing the signal to noise ratio' to justify this decision.

Good point.

We took out: "for two primary reasons: first of all, masking enables us to objectively compare the same water masses across the different models even if latitudinal boundaries of these water masses differ among the models (assuming that water masses across models behave in dynamically and biologically similar ways)."
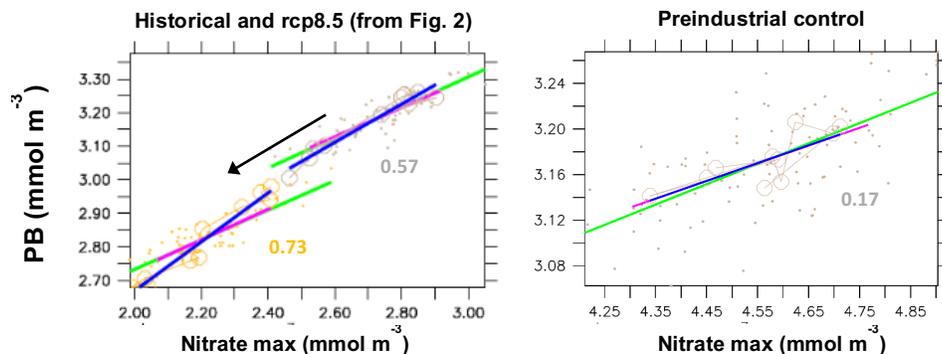
4.) Section 2.6: I found this confusing. Why not just use the control run of the models to define interannual variability without climate change effects – that after all is exactly what the control run represents. I found the mathematical gymnastics in this section confusing and perhaps unnecessary – surely the "control and climate change time series" of the section title are the control and RCP8.5 runs of the models?

Good question.

Added the following to the text to address this: "Here we purposely chose to use detrended *historical* scenario time series rather than *preindustrial control* scenario time series (forced with constant preindustrial CO_2 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):

2

Historical and rcp8.5 (from Fig. 2)   **Preindustrial control**

5.) Page 8165, lines 5-11: In the calculation of trends (here particularly because monthly data is used, but also elsewhere with annual data), are the effects of autocorrelation accounted for? Calculating a trend on monthly data without accounting for autocorrelation will result in spurious correlations and too-small p-values.

Good point. In any analyses of model output, we always use yearly time series, which should not contain any significant autocorrelation. For all of our observations besides chl, we had also used yearly data. With chl, however, we had previously used monthly time series, but we have now changed that to yearly as well. Spatial patterns of trends and significance were not noticeably altered. See new Fig. 7c.

We also clarified use of yearly data in our observational analysis by adding the following to the text: "The linear trend in Fig. 7c was calculated from yearly-averaged monthly chl anomalies, which ensures minimal autocorrelation. To look at trends in observed summertime MLD, monthly ocean temperature and salinity reanalysis products from the Met Office Hadley Centre's EN3 dataset (http://www.metoffice.gov.uk/hadobs/en3/) were used to calculate minimum monthly MLDs for each year from 1950-2013. To look at trends in observed summertime cloud cover, synoptic monthly mean ERA-INTERIM (http://www.ecmwf.int/en/forecasts/datasets/era-interim-dataset-january-1979-present) reanalysis products of total cloud cover from December 1980 - February 2013 were averaged over the summer months (December to February) of each year to generate a yearly summertime cloud cover time series."

6.) Page 8168 – discussion of sea ice reduction. A figure or reference to published work is needed here to show the extent of sea ice retreat.

Good point.

We added the following citations:

1.) Turner, J., Bracegirdle, T. J., Phillips, T., Marshall, G. J., and Hosking, J. S.: An initial assessment of Antarctic sea ice extent in the CMIP5 models, Journal of Climate, 26, 1473-1484, 10.1175/JCLI-D-12-00068.1, 2013.

2.) Mahlstein, I., Gent, P. R., and Solomon, S.: Historical Antarctic mean sea ice area, sea ice trends, and winds in CMIP5 simulations, Journal of Geophysical Research: Atmospheres, 118, 5105-5110, 10.1002/jgrd.50443, 2013.

We also added Fig. S10, which shows 100-year all-model changes in sea ice area fraction.

7.) Page 8170, lines 14-19: the authors state that within each band the drivers of PP response are the same in all-model mean as in individual models – I would say except in the 50-65S band where there is some model disagreement on drivers.

Sounds good.

We added the text in red: "In particular, within each zonally-banded biome, the proposed drivers of projected phytoplankton responses in the all-model means are the same ones driving phytoplankton responses within the individual models studied here (with the possible exception of the 50-65°S band, where iron appears to play a role within IPSL-CM5A-MR, but not in the all-model mean)."

8.) Page 8171, lines 1-5: unmasked comparisons should be shown for selected correlations (similar to Figs 2 + 3), rather than the unselected ones, so that the reader can more easily assess the impact of using the masking. Also, a table with the slopes calculated in Figures 2 and 3 would make comparison easier.

Great suggestions!

We added Table 3, which summarizes the slopes calculated in Figs. 2 and 3.

We also added Fig. S13 (the unmasked analog of Fig. 2) and Fig. S15 (the unmasked analog of Fig. 3).

9.) Page 8171, lines 24-29: Although the authors conclude that the same mechanisms act on timescales of interannual to centennial, in reality on long time scales phytoplankton adaptation could alter the driver-response relationship which they observe at the interannual scale. This should be included in the discussion.

Good idea.

We added to the text: "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."

10.) Page 8172, lines 22. . .Page 8173, line 17: I think this section should refer to figure 4? (not figure 3) Worth noting that although various relationships are discussed, only the relationship with max yearly winter nitrate is significant. The boxes are also artificially chosen to draw the eye in some cases, with lots of points falling outside the boxes, e.g. green and purple on Fig 4c.

We fixed the figure reference.

Good points on the boxes and significance.

We added the following text in red to clarify significance and to discuss points that fall outside of the boxes: "Nitrate emerges as the driver for changes in PP within the 30-40°S band across all models (i.e., all red points lie in the third quadrant and within the red box in Fig. 4b). Models with greater relative decreases in wintertime surface nitrate concentrations undergo significantly ($p<0.05$) greater decreases in average production within the 30-40°S band. It is worth noting that this is the only significant, highly

linear intermodel relationship within any of the zonal bands. In the rest of the bands, we mostly interpret only the sign, rather than the linearity, of PP changes related to the driving variables of interest across models. Within the 40-50°S band, in general, models with increases in relative iron concentration and decreases in summertime MLD also experience relative increases in PP (Fig. 4a, c, purple boxes). Models NorESM1-ME and IPSL-CM5A-LR are exceptions to this, however, in that PP still increases while iron concentrations decrease (Fig. 4c, purple unboxed). In these models, increases in light availability due to shoaling of summertime MLDs (Fig. 4a) and decreases in cloud cover (Fig. 4e) are large enough to cancel out the PP-suppressing effects of iron concentration decreases (Fig. 4c) between 40-50°S. Further solidifying the importance of climate-driven changes in light availability within the 50-65°S band, models predicting relative increases in summertime MLD or average annual cloud cover, along with decreases in maximum annual IPAR, also predict relative decreases in PP in this region (Fig. 4a, d, e, green boxes). Iron also emerges as a potential driver of PP decreases within the 50-65°S band, but not across all of the models (Fig. 4c). In models which undergo PP decreases concurrent with iron concentration increases (GISS-E2-R-CC, GISS-E2-H-CC, HadGEM2-CC, HadGEM2-ES, IPSL-CM5A-LR, NorESM1-ME, and MPI-ESM-LR; see Fig. 4c, green unboxed), reductions in light availability tend to be relatively large such that they win out in determining overall PP change. For example, GISS-E2-R-CC exhibits the largest relative iron increase between 50-65°S out of all the models (Fig. 4c), but also the greatest relative summertime MLD deepening (Fig. 4a), leading to vast reductions in light supply to phytoplankton during the most productive time of year."

11.) Page 8173, lines 20.. – what is it in the model formulation that makes a particular model's pp more or less sensitive to iron?

Great question. This is a pretty complicated subject.

We added the following to the text to address this: "Iron cycling within the ocean remains poorly characterized and is typically crudely parameterized (if at all) compared to the macronutrients. These models also differ considerably in many aspects of their treatment of iron including but not limited to the magnitude and location of sources (from both the atmosphere and the sediments), ligand dynamics, scavenging losses, and iron to carbon biomass ratios (Moore et al., 2013b). It is out of the scope of this paper to assess all of these differences, but at first glance, it appears that the models with more complex iron cycling dynamics have phytoplankton that are more sensitive to iron changes. For example, the more iron-sensitive GFDL-ESM2, CESM1-BGC, and IPSL-CM5A models have variable iron to carbon ratios and include sedimentary sources of iron (however crudely parameterized) (Dunne et al., 2013; Moore et al., 2013b; Aumont and Bopp, 2006), while the less iron-sensitive NorESM1-ME, HadGEM2, GISS-E2, MPI-ESM models do not (Assmann et al., 2010; Collins et al., 2011; Gregg, 2008). Models within the more iron-sensitive group tend to exhibit less well-defined latitudinally-banded 100-year phytoplankton changes, while the other models tend to exhibit a more obviously banded PB and PP change structure (see Figs. S1-2)."

Here we added the following references:

1.) Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-system model–HadGEM2, Geoscientific Model Development, 4, 1051-1075, 10.5194/gmd-4-1051-2011, 2011.

2.) Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., and Misumi, K.: Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 Scenarios, Journal of Climate, 26, 9291-9312, 10.1175/JCLI-D-12-00566.1, 2013.

12.) Page 8174, lines 25-29: the summertime MLD decreases, but wintertime iron increases. I assume that the atmospheric deposition of iron does not have an increasing trend in the models? So this implies that wintertime mixing must be deeper to result in increased winter time iron. In turn, this suggests that there will be a more pronounced seasonal cycle in MLD. What mechanism might cause this?

Great, but highly non-trivial questions.

First of all, it is true that atmospheric deposition of iron remains constant, but there are other possible causes of increased wintertime iron concentrations than wintertime MLD changes alone.

We added the following to the text to address this (at the end of Section 3.1):
"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)."

In the above text, we added the following reference:

Misumi, K., Lindsay, K., Moore, J. K., Doney, S. C., Bryan, F. O., Tsumune, D., and Yoshida, Y.: The iron budget in ocean surface waters in the 20th and 21st centuries: projections by the Community Earth System Model version 1, Biogeosciences, 11, 10.5194/bg-11-33-2014, 2014.

As for MLD max, many models do show an increase around 40-50°S (Fig. S4, especially the bootstrap percentages map). A possible mechanism to explain a more pronounced seasonal cycle in MLD is as follows. First, MLD min likely shoals because of a summertime-intensified southward shift in westerly wind stress, which leaves behind more stratified waters in the 40-50°S band during the summer months (as was briefly discussed in Section 3.1 within the text). Second, MLD max deepens a lot in the winter just north of the ACC front in both the real ocean (this is clearly seen in Fig. 3 of Doung et al. 2008,

copied here for reference) and the CMIP5 models (Fig. S4, all-model historical mean); with climate change, however, this band of deep MLD shifts slightly southward (compare Fig. S4 all-model historical mean to Fig. S4 all-model mean change), due perhaps to southward movement of the ACC front in turn again driven by the southward shift in the westerlies.

Doung et al. (2008), Southern Ocean mixed-layer depth from Argo float profiles, JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, C06013, doi:10.1029/2006JC004051, 2008.
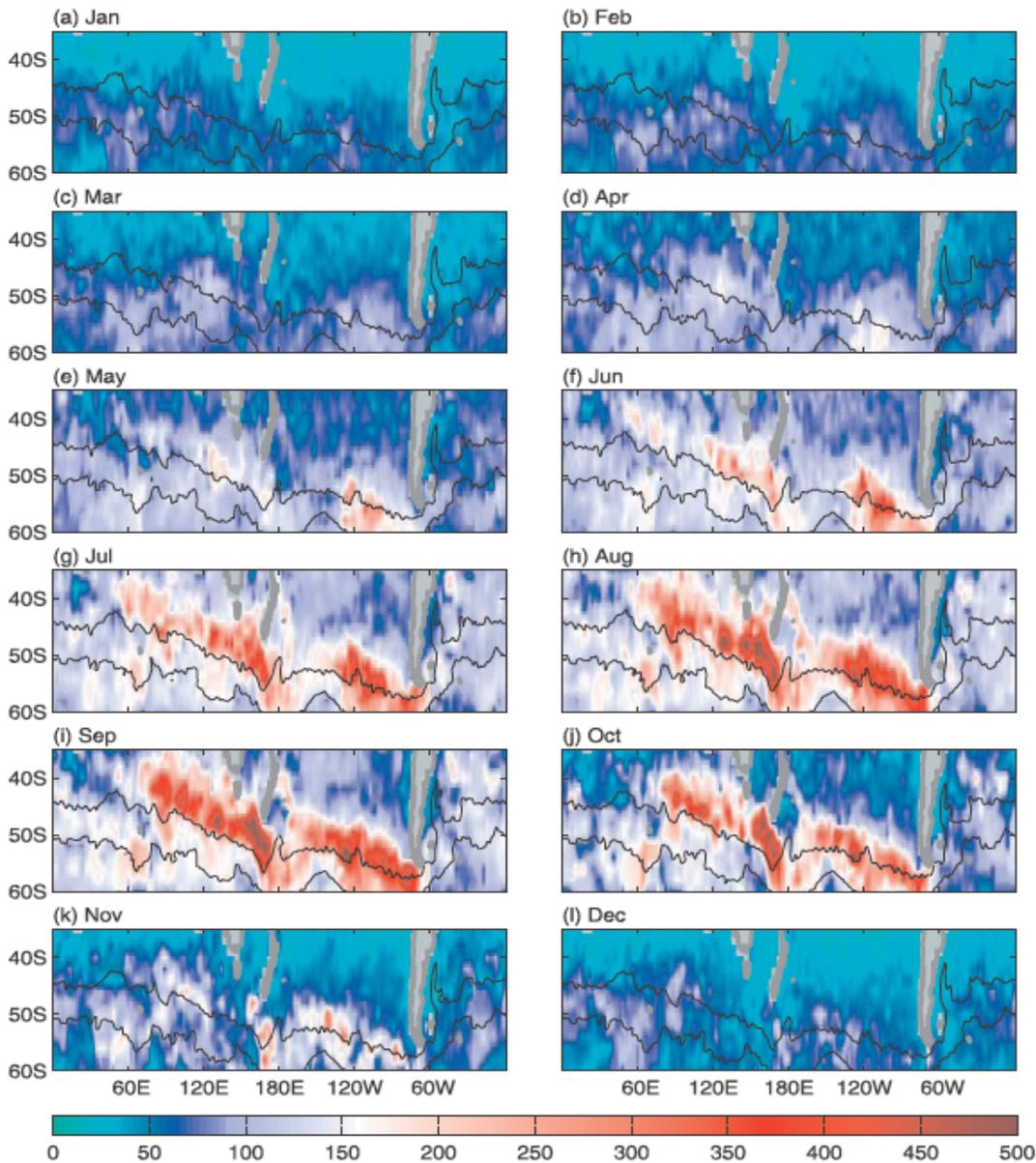
Figure 3. Objectively mapped monthly MLD (meters) from density criterion ($\Delta\rho = 0.03$ kg m^{-3}) for (a) January through (l) December. The black curves are the monthly mean SSH contours of -0.4 m (north) and -1.2 m (south), which are used to define the ACC. Quantitative data are found in auxiliary material.

13.) Page 8179, lines 15-24: surely the observations should be compared to the same period in the

7

historical runs, not to the full future runs? Also, I think the satellite-derived trends are calculated on monthly data, but model output is annual? So not directly comparable. (Also see earlier comment about accounting for autocorrelation).

We see why this was confusing.

We addressed the autocorrelation concern and the monthly/yearly data confusion in points 5 and 1 above, respectively.

We added the following two blurbs to the text to clarify the reasoning behind our method of model-data comparison:

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."

And at the end of Section 3.4:
"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)."

14.) Page 8179, lies 28-29: I disagree that the comparison with observed data "suggests that the effects of climate change on SO phytoplankton may have already become detectable". The authors have only shown that there may be some similar sign/pattern of trends and possibly some consistent mechanisms in 2 time series which have substantial difference (as in above comment). For detection to occur, much more rigorous analysis than this is required.

Good point.

We changed the sentence

"This potentially suggests that the effects of climate change on SO phytoplankton may have already become detectable."

to the following:

"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)."

15.) Figure 1: the latitudes need to be marked on one of the figures so that it's easier for the reader to relate the text referring to various latitudinal bands to the figures.

We added latitudes to many of the figures. We had forgotten that other people don't stare at these maps for hours like we do…

16.) Figure 2: It's really hard to distinguish the different symbols here, particularly the grey ones.

We purposely wanted the background dots to be a little bit faint, so that the brighter-colored best-fit lines could be emphasized. We played with different ways to plot the dots and circles, but darker colors/bigger symbols made the plot look really distracting/confusing and caused too much overlap. We recommend zooming in to better distinguish the different symbols.

17.) Figure 3: are these best fit lines statistically significant?

Good question. They are not because it's difficult to test for significance in this type of spatial correlation. Grid points close to one another may be highly correlated, so we would always overestimate the significance. Thus, Fig. 3 is really just meant to be a qualitative comparison and we add the lines in simply to visualize the slope and be able to compare that slope to those from Fig. 2.

To make this clearer, we added to the text: "Least squares best-fit lines are drawn for each scatter plot to help visualize the slopes and enable comparison with the corresponding slopes in Fig. 2. Because it is difficult to accurately test for significance in this type of spatial correlation (neighboring grid points are likely highly correlated, leading to large significance overestimates), these regression lines may or may not be statistically significant. Thus, the lines are meant only to serve as a qualitative visual guide."

18.) Figure 4: How were these potential drivers chosen? PPmax is used throughout as a metric – if the authors are trying to assess overall productivity, the annual total (integrated) PP would be a much better measure. Also, the authors are then comparing a single event (the max PP) to an annual average, e.g. cloud fraction. Again, it would be more consistent to compare annual average cloud fraction to annual average or total PP.

Figure 4 actually does have average annual integrated primary productivity (PP) as the y-axis variable, but the caption and labels were not clear (sorry about this), so we did actually compare annual average cloud fraction to annual average PP.

To clarify how these potential drivers were chosen, we added to the text: "These variables were chosen by first plotting all of the potential drivers of interest (listed in Table 2) and then choosing the ones which showed the strongest correlations or most consistent directions of changes across the models, guided by the relationships found in Figs. 1-3."

In this study, we are not necessarily trying to assess overall productivity changes alone, but rather are trying to understand how SO phytoplankton characteristics in general may change with future warming along with the drivers behind these changes. To clarify this, we added the following text in red at the beginning of the "Results and Discussion" section:

"In this study, we attempt to understand how the general characteristics of SO phytoplankton may change with future warming by investigating biomass and productivity at both peak bloom times and

averaged over the entire year. To this end, we choose to study the following two variables: (1) maximum annual surface phytoplankton biomass (henceforth PB, representative of phytoplankton biomass at the peak of an annual bloom) and (2) average annual primary production vertically integrated down to 100-m depth (henceforth PP, representative of average yearly water column integrated conditions). We conducted all of our analyses with both of these variables, but only show results for the variable which made the most sense to use in the context of the analysis. For example, whenever we analyze individual models, we show PB because we frequently only have monthly model output (with which to generate maximum, minimum, or average annual data) at the surface of the ocean (i.e., monthly NO_3, iron, and light output are only available at the surface) and want to keep the variables we are cross-correlating spatially consistent whenever possible (either all variables at the surface only or all vertically-integrated only)."

For further clarification to the reviewer, we chose to try both annually averaged PP and maximum annual PB in all of our analyses (not all results are shown in our manuscript) not only to preserve temporal/spatial consistency, but also in order to amplify the strength of correlations and enhance our ability to detect existing mechanistic relationships between phytoplankton responses and their drivers. In some cases, PB max was mostly strongly correlated with annually averaged variables, while in others it was most strongly correlated with either max or min annual variables. This was also the case for avg annual PP, in that sometimes it was most strongly correlated with an annually averaged variable and at other times it was most strongly correlated with an annual max or min variable. Rather than viewing these avg annual–max/min annual mixed correlations as inconsistent, we see them as informative and perhaps even helpful in illuminating specific mechanisms. For example, a stronger correlation between avg annual PP and max annual iron than between avg annual PP and avg annual iron suggests that it is wintertime iron supply changes that are driving overall annual PP changes. Another example mentioned in the caption of Fig. 2 is directly related to the question the reviewer brings up here, in that PB max was significantly correlated with average annual cloud cover but not summertime cloud cover on all three studied timescales in GFDL-ESM2G between 50-65°S. This suggests that there may be some temporally integrative effect of cloud cover on PB max or that there is perhaps a first-order ultimate mechanism driving cloud cover change at one time of year and phytoplankton change at another time (which is why we also test IPAR max and other summertime light availability variables for summertime light effects). Either way, the information gained is interesting and potentially useful for future studies looking more carefully into seasonal effects and changes with impacts lasting over several months or more.

19.) Figure 5: For band 50-65S, PP is reported as having 41% decrease, but by eyeball looks like mostly increasing. Also, for this band and for 40-50S, only half of the models agree on the sign of the trend for PB, which should be noted.

We added the following red text: "The majority of model realizations predict an increase in PP (59%), while 55% predict a decrease in PB."

We also added the following red text: "Within the transitional (40°S to 50°S) band, most of the model realizations predict an increase in PP (70%) while only around half of the models predict an increase in PB (55%)."

20.) Figure S11: Add a colour legend to this figure to stop the reader having to flick back and forth.

Done.

21.) Figure S12: What does the star indicate in the bottom left plot? (and for Fig S13-15 too).

It indicates the variable chosen for display within that zonal band and that model in Fig. 2 or 3.

To clarify this, we did the following:

Fig. S12 caption: Changed "starred" to "chosen plot indicated by a star."

Fig. S14 caption: Changed "starred" to "chosen plot indicated by a star."

22.) Figure S16: W is not reported in some sub-plots, e.g. c,d,i

Fixed.

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

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 drives 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 hypotheitically 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

11

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 permutated 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 CO_2 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 anticorrelared 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 permutated 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.
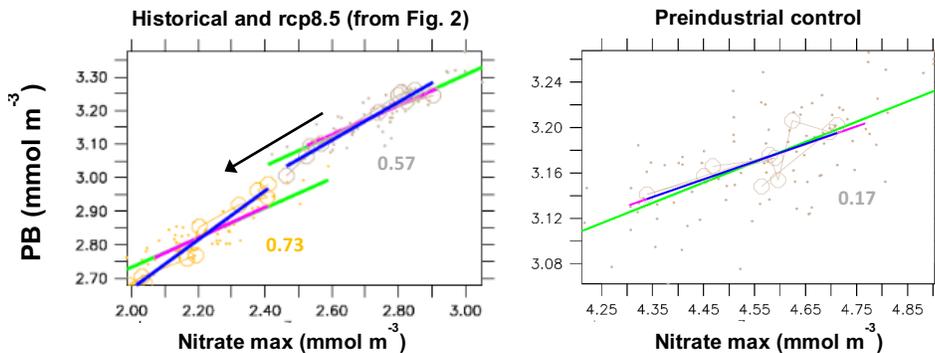
We feel that past work has not done enough to approach the climate change attribution problem from this necessary angle. Many studies have looked at drivers either on interannual timescales alone or climate change timescales alone. To form a more complete picture as well as a more coherent argument (as 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 CO_2 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 discsussion, the authors should note that the models are 'perfect' integraters 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 this 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.