# *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)."

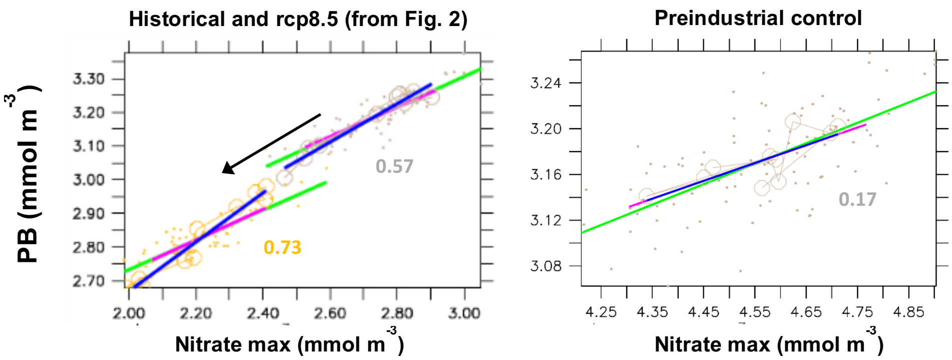
*4.) Section 2.6: I found this confusing. Why not just use the control run of the models to define interannual variability without climate change effects – that after all is exactly what the control run represents. I found the mathematical gymnastics in this section confusing and perhaps unnecessary – 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):




*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 21$^{st}$ 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.

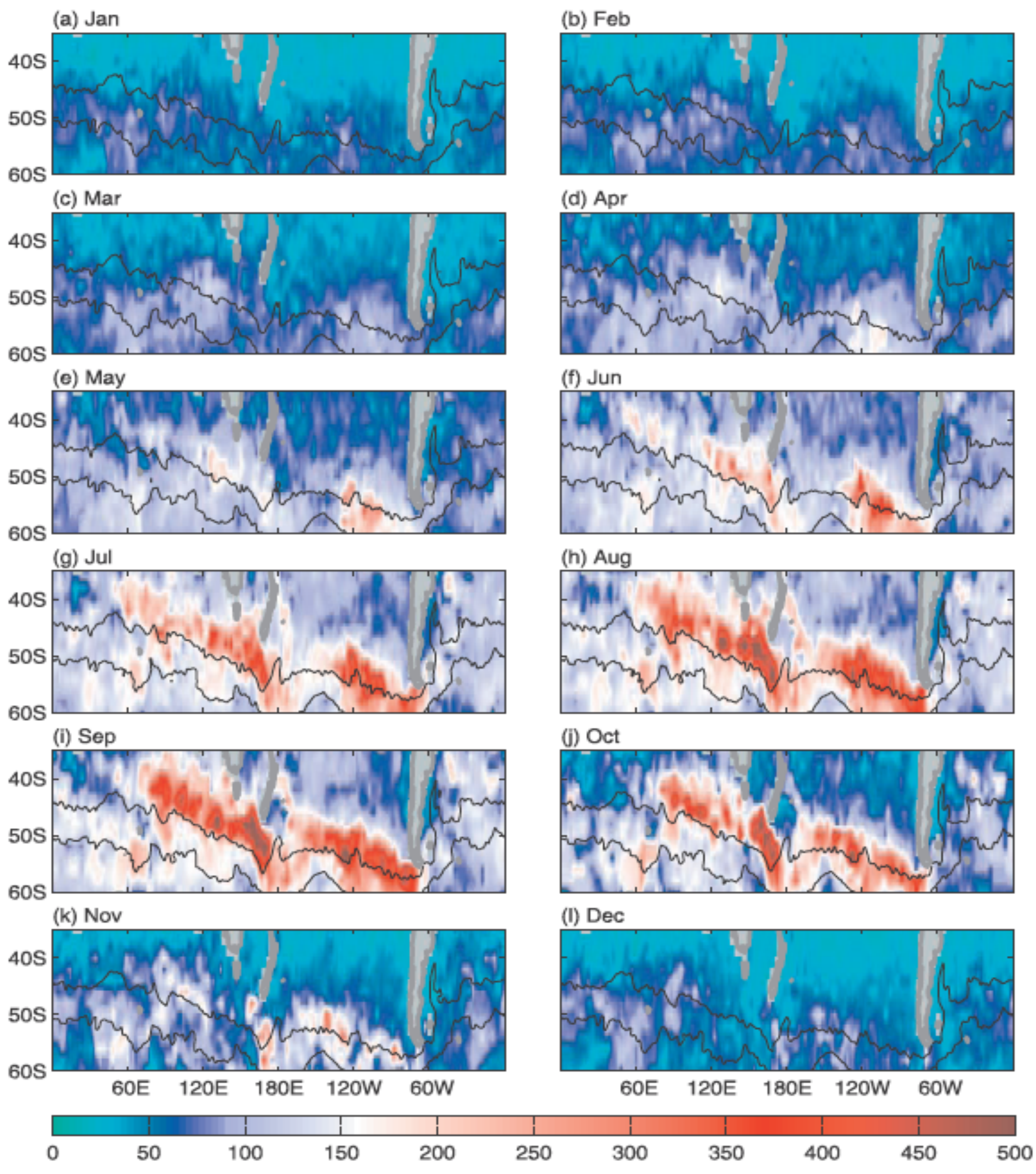



**Figure 3.** Objectively mapped monthly MLD (meters) from density criterion ($\Delta\rho = 0.03$ kg m$^{-3}$) for (a) January through (l) December. The black curves are the monthly mean SSH contours of $-0.4$ m (north) and $-1.2$ m (south), which are used to define the ACC. Quantitative data are found in auxiliary material.

*13.) Page 8179, lines 15-24: surely the observations should be compared to the same period in the* *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 21$^{st}$ 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.