

Response to Referee RC2:

We would like to thank the referee for the review of this manuscript and their constructive comments. Our response to each comment is below with the referee's comments highlighted in bold typeface. Where appropriate, we have also included relevant changes in the revised manuscript in italic typeface.

The main focus of these specific comments is to highlight text that could benefit from being more quantitative rather than subjective. In addition, there are some specific suggestions for clarifications aimed at improving the overall flow and clarity of the manuscript. Before starting the specific suggestions, one other general remark I have is that the figure axes and color bar labels are very small and hard to read.

We will improve the readability (i.e. increase size of labels) of all figures in a revised version.

Abstract

The abstract provides an overall view of how well ACCESS-ESM1 performs; however, they provide this information using terms such as “good” and “performs well”. How well? Good according to whom? Is your good the same as my good? First off, the authors state ACCESS-ESM1 overestimates the seasonal amplitude of LAI, but do not attach any quantities to this statement. Rather than providing a quantitative assessment of this statement, the reader is left to wonder if this is a substantial bias, perhaps even being prohibitive of using this model in the future, or merely a relatively minor difference that is offset by other positive features. The paper then continues with the statement that the oceanic and land fluxes “show good agreement with the observations”, but again, no metric is used. How big are the differences, what is the error, or how closely correlated, both spatially and temporally, are they to observations and/or other CMIP-5 models? In the last overview comparison, the authors state that the seasonal cycle is “close to the observed seasonal cycle”, but as the reader I have to wonder how close is close? I believe that putting quantifiable metrics on at least some of these statements will strengthen the concluding remark that ACCESS-ESM1 is indeed a useful tool.

We will include more quantitative statements in the abstract in a revised version.

Observations

The model evaluation (later in the text) proceeds through a straightforward succession of comparisons; however, the order of these comparisons is different than the order the data is listed in. For clarity and consistency, it would be an easy fix to reorganize this section to present the data in the order that they are used.

We will reorganise section 2.4 (Observations) in a revised version, so that data are presented in the order they are used later in the comparison.

Land Temperature and Precipitation

This section presents a time-series comparison of temperature and precipitation, but contains very little quantitative statements. For example, the temperature anomalies are “close to the observed anomalies through most of the period”. Again I have to wonder, how close is close? Perhaps a difference plot showing the errors would be useful? Or perhaps a correlation coefficient that may or may not be

significant? Some sort of metric on this statement would be much more enlightening to the reader. For example, as I look at Fig. 1 myself, I see the author's point that ACCESS-ESM1 is lower than the observations 1965-2005; however, to my eye I also see the model looks quite a bit lower in the 1940s. I think that a difference plot would help identify these areas, rather than relying on the reader to have to assess the differences between the models and the observations on eyesight alone. Another example later in this section occurs in the precipitation anomaly discussion: the authors state the differences are "generally small," but provide no values to suggest what is meant by small. This is followed by the statement that "the simulations compare well with observed rainfall anomalies until about 1950," with no supporting metric such as minimal differences or significant correlation coefficients to back up this statement. Overall, I found that this entire section had very few quantitative comparisons, instead relying heavily on subjective terminology, making me believe that at least some measurable metrics would improve the paper.

We have attached an updated Fig. 1 showing differences for temperature and precipitation anomaly (simulated – observed). However, we might not include the difference plot in the manuscript, because we don't think it adds a lot more information. Instead, we will focus on decadal mean differences and refer to the values in the text as appropriate:

Both ACCESS-ESM1 simulation scenarios (PresLAI and ProgLAI) show similar temperature anomalies over most of the historical period, being close to the observed anomalies through most of the period (decadal mean difference smaller than 0.2 K), apart from the 1940s where the PresLAI scenario shows a larger negative anomaly (decadal mean difference of about 0.37 K), which will be discussed later. From about 1965-2005 anomalies are by up to 0.4 K (decadal mean difference) lower than observations for both scenarios.

ACCESS-ESM1 simulations compare well with observed rainfall anomalies until about 1960 (decadal mean difference smaller than 8 mm/yr), with the exception of the period 1911-1920 for PresLAI (decadal mean difference of about 12 mm/yr) and the period 1951-1960 for ProgLAI (decadal mean difference of about 17 mm/yr). After that, observed anomalies are mostly higher than the simulation results (decadal mean difference of up to 41 mm/yr), a feature also seen in the ACCESS1.3 historical ensemble (Fig. 6a, Lewis and Karoly 2014).

The anomaly in the 1940s is already discussed in the manuscript on page 6, lines 26-30.

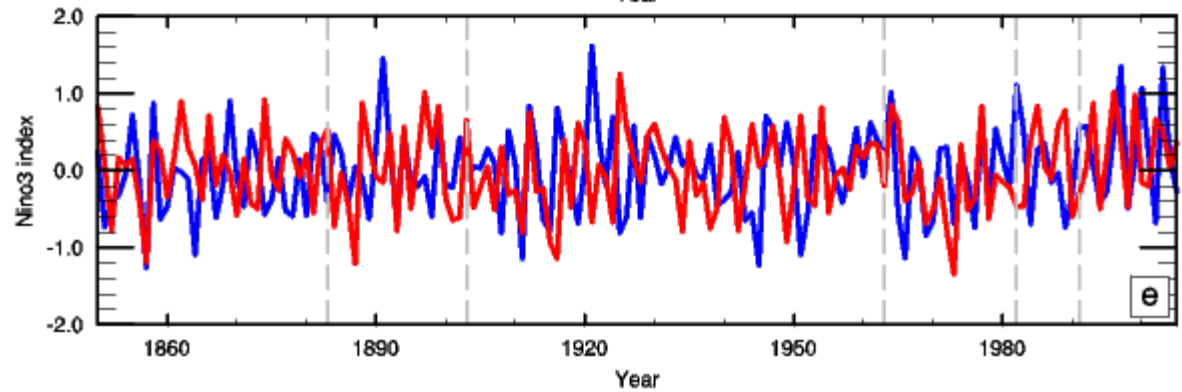
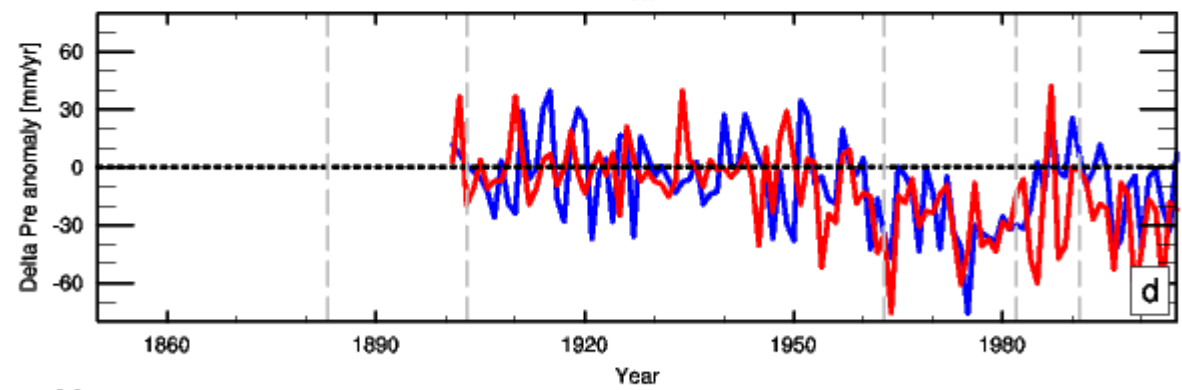
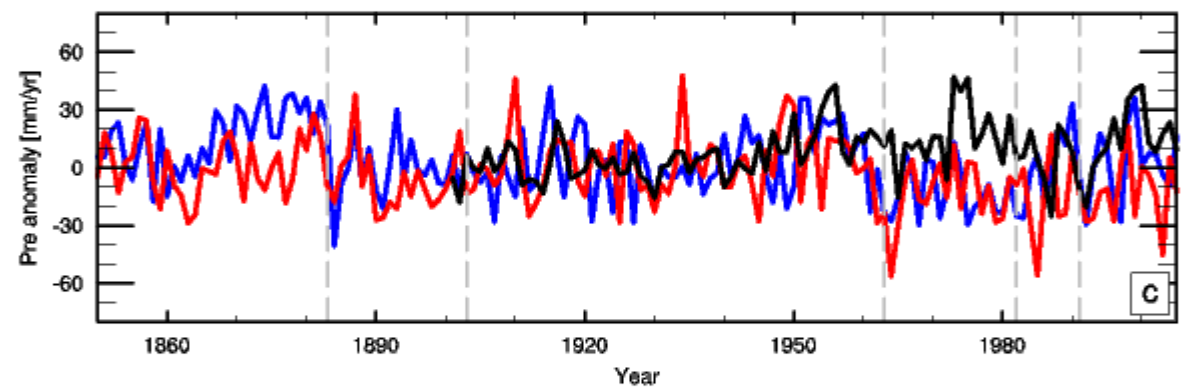
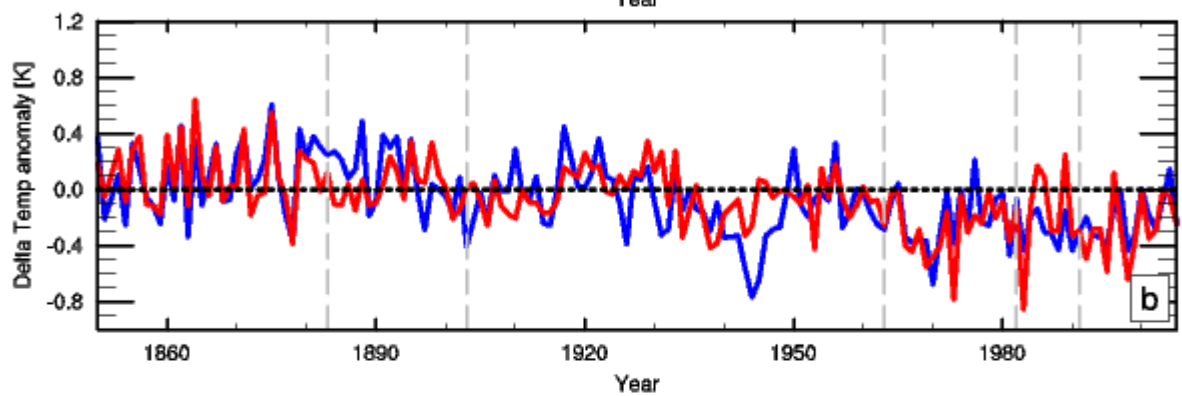
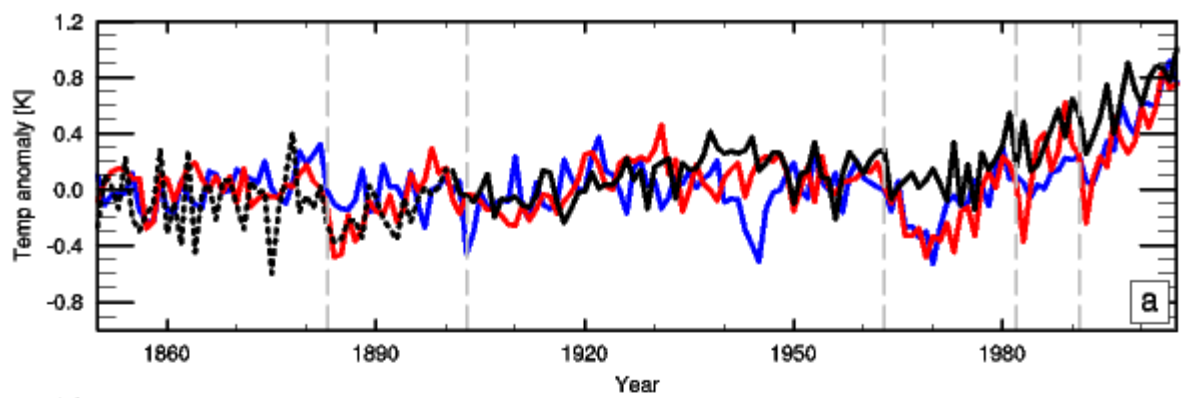
We do not think that a correlation coefficient would be a very meaningful metric to assess the errors between simulated and observed temperature and precipitation. According to Anav et al. (2013), there is no reason to expect models and observations to agree on the phasing of internal interannual variations. We therefore calculate the model variability index (MVI) to analyse the performance of ACCESS-ESM1 for temperature and precipitation. The MVI compares the models variability at every grid cell, which is then averaged for the globe. Perfect model – observation agreement would result in an MVI value of 0. For example, for temperature we calculate an MVI of 0.3 (PresLAI) and for precipitation an MVI of 1.7 (PresLAI) over the period 1901-2005. We will include this information in a revised version of the paper:

The interannual variability in temperature is well reproduced by both ACCESS-ESM1 scenarios, showing an MVI of 0.3 (PresLAI) and 0.4 (ProgLAI) for the period 1901-2005.

According to Anav et al. (2013) only a few CMIP5 models show an MVI of lower than 0.5 (although their calculation is based on present day, i.e. 1986-2005).

For precipitation we calculate an MVI of 1.7 (PresLAI) and 1.8 (ProgLAI) for the period 1901-2005, which suggests that the IAV is not well represented in ACCESS-ESM1. However, according to Anav et al. (2013) none of the CMIP5 models had an MVI close to the threshold of 0.5. Also note that for the calculation of the MVI for precipitation we had to exclude 60 land points (mainly coastal points) due to inconsistencies in the regridding.

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., and Zhu, Z.: Evaluating the Land and Ocean Components of the Global Carbon Cycle in the CMIP5 Earth System Models, J. Climate, 26, 6801–6843, 2013.



The last paragraph in this section discusses the timing of precipitation anomalies versus volcanic eruptions. The authors point out a reduction in precipitation following eruptions, with the one exception of El Chichon; however, when I looked at the time-series, I did not see a decrease after the Santa Maria eruption in addition to El Chichon. In fact, looking at the time-series, the decrease after volcanic eruptions did not stand out to me, especially when the authors note the reduction following the Mt. Pinatubo eruption, which does stand out, is too far away from the eruption date to be related. I think a more quantifiable analysis of the magnitude and the timing of this decrease, in days or weeks or some stated time-scale, would be helpful.

We agree with the referee that there is no decrease in precipitation visible following the eruption of Santa Maria. We will include this statement in a revised version of the manuscript.

However, we did not say that a reduction in rainfall following Mt. Pinatubo is too far away from the eruption date. In fact, along with Krakatao and Mt. Agung, the Mt Pinatubo event shows a significant reduction in precipitation anomalies immediately after the eruption.

Sea Surface Temperature and Mixed Layer Depth

The second full paragraph of this section discusses spatial patterns of sea surface temperatures, except no time period for the analysis is provided, not even in the caption for Figure 3. Are these differences over the entire simulation or a selected time period? Shifting to the text, the authors state that ACCESS-ESM1 “produces very heterogeneous differences from observations.” Reading this, it was unclear to me what was meant. When I turned to look at the figure I expected to see random errors; however, in my opinion the differences are spatially coherent in latitude bands. Then in this same discussion they state “there do not appear to be strong seasonal biases,” but with this terminology I have to wonder are there or aren’t there? Then they state the exception of the North Atlantic, which has a coherent bias towards cooler temperatures, but to me it looks like this is a year-round bias more than a seasonal bias. It perhaps does vary in magnitude with season, being a larger bias in August, but it looks like the sea surface temperature here is underestimated year round. Further looking at Figure 3, I also see a flip-flop in errors in the southeastern Pacific Ocean, with positive differences in August and negative differences in February. To me, having differences that vary with time of year makes it seasonal, but this region is not mentioned. I think this section could benefit from more careful wording and analysis.

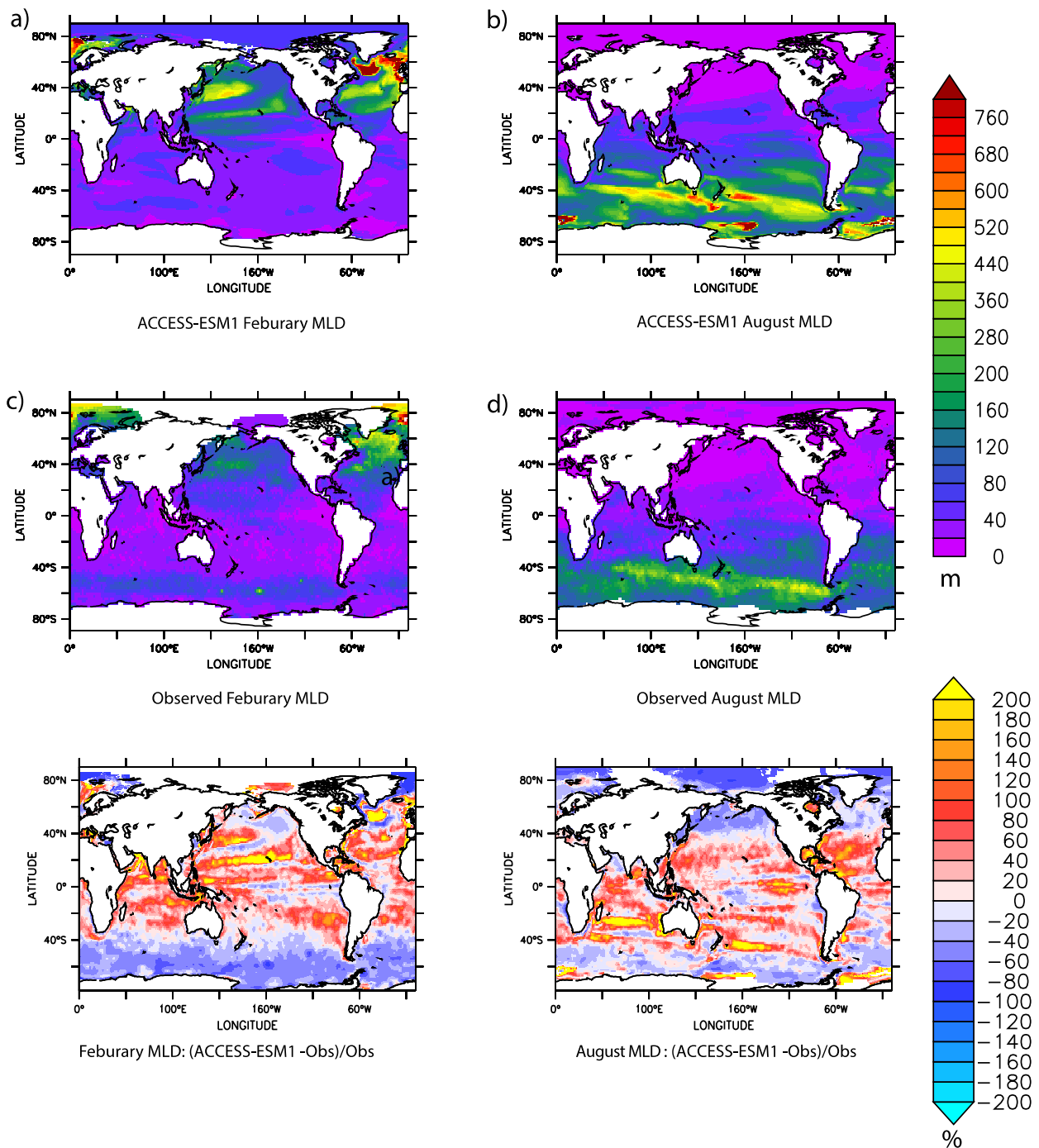
We apologize for the oversight and have now stated in the caption and text that the time period of the figures is the IPCC historical period 1986-2005. We have rewritten the text to better convey our intended meaning and analysis. Indeed the spatial patterns of the warming are perplexing better upon a much closer examination appear to be associated with the Met Office Unified Model (MetUM) which has known biases and which ACCESS-ESM1 utilizes as its atmospheric model, indeed similar biases are seen in HadGEM2 that also use the MetUM.

The ocean mixed layer depth discussion could also benefit from a more quantitative analysis, rather than using statements like “appears to slightly overestimate the

depth in winter” and “appears to underestimate the depths in summer,” with no statistical support for these subjective comments. Also, in Figure 4 in this section, the caption states that differences are shown, suggesting a difference plot; however, the figures just show the results from both the model and the observations. I would suggest rewording the caption to avoid confusion.

To address the reviewer’s comments we have rewritten the section on mixed layer depth comparison, and added an additional figure showing the percentage changes in mixed layer depth between the observations and ACCESS-ESM1. This allows the changes between the obs and model to be quantified relative to the total observed mixed layer depth. We have also updated the caption to reflect the reviewers concerns.

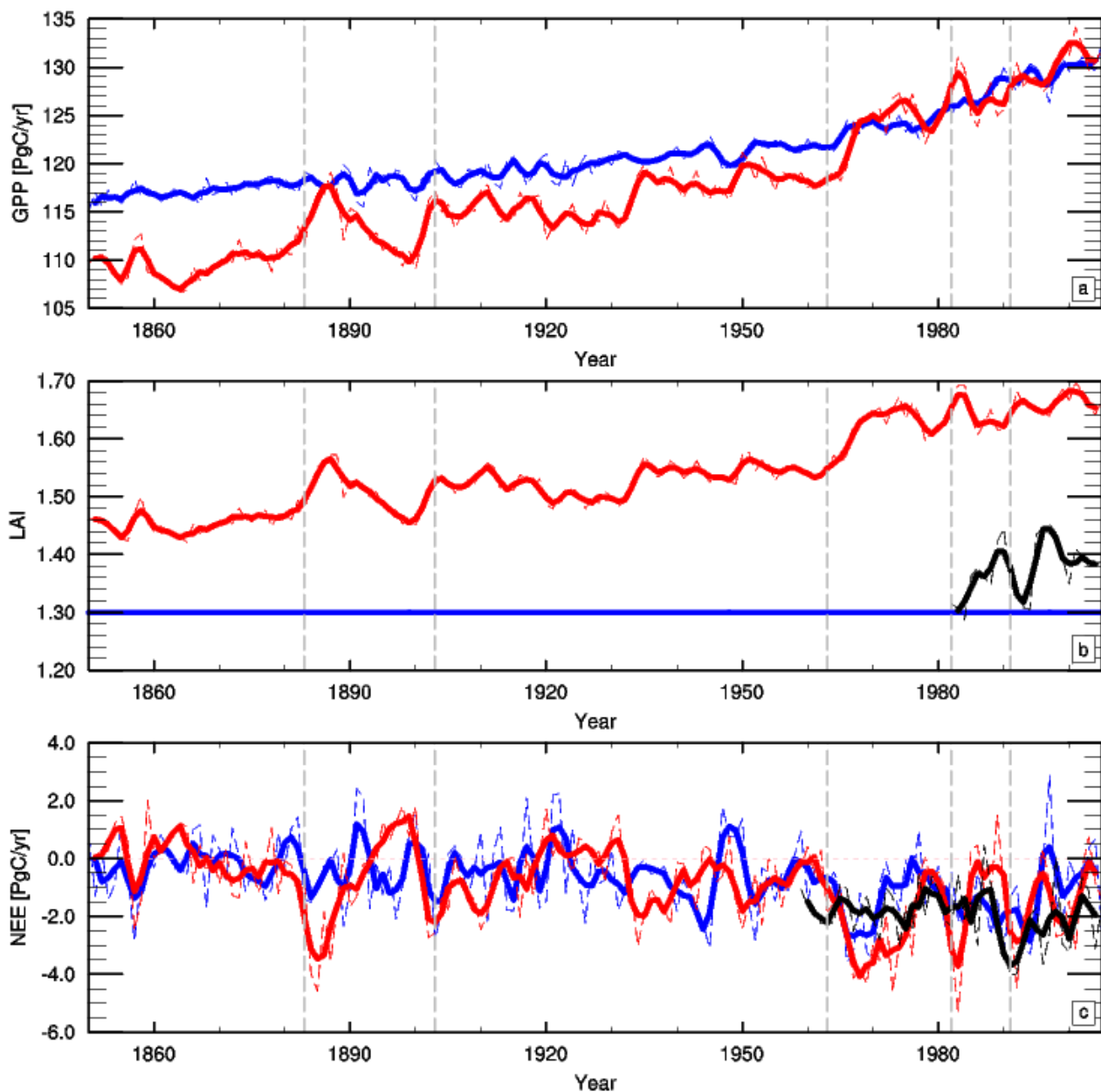
In the higher latitude that the winter mixed layers are well captured by ACCESS-ESM1 Figure 4. This is encouraging given that many ocean models tend to underestimate winter mixed layer depths (Sallee et al (2013) and Downes (2015)). Simulating winter mixed layers correctly is critical for setting interior ocean properties supplying nutrients to the upper ocean to fuel the biologically active growing season (Rodgers et al, 2014). However in contrast to the winter, ACCESS-ESM1 appears to systematically underestimate mixed layer depths in the high latitude ocean in summer, ~60% (or 30-40m) in the Southern, Pacific and Atlantic Oceans. In the Southern Ocean, in particular, the underestimation of summer mixed layer depths is consistent with Sallee et al (2013) and Haung et al (2013) who showed that most CMIP5 models underestimate summer mixed layer depths. Haung et al (2013) attributed this to a lack of vertical mixing in CMIP5 rather than sea surface forcing related to individual models, this is consistent with Downes et al (2015) who showed that these biases are also present in the ocean only simulations of ACCESS-ESM1 (ACCESS-O).



Land Carbon Response

I think it would be helpful to include the MODIS/AVHRR LAI data in Figure 5b. I realize this would only be for the last few years of the simulation, but the benefit is that it would provide a reference to the simulated LAI values. After reading it through, I discovered there is a section on LAI where it is discussed in more detail, but this figure comes first, and when I read it I was wondering how it compared.

Observation based LAI data (MODIS/AVHRR) have been included in Fig. 5b for the period 1982 to 2005. For comparison we have also included the prescribed LAI used in ACCESS-ESM1, which has no interannual variability. The updated figure will be included in a revised version of the paper



I found the discussion in the last two paragraphs of this section confusing. I loved to see values and uncertainties; however, it was unclear to me what values were comparable. The section starts with a discussion on carbon uptake, which at first I assumed represents total uptake, or GPP. But from what I could tell, the same values provided for ACCESS-ESM1 were then used in the following paragraph, which talks about NEP. I read the section several times, but only found the one value of 154 Pg C for prognostic ACCESS-ESM1. I was then confused when it was stated that this value is “at the low end of the CMIP5 range,” when that range is estimated to be from -59 to 18 PgC according to Shao et al. (when outliers are not included) or from

-124 to 50 PgC from Jones et al. Based on these numbers, doesn't ACCESS-ESM1 take up and store much more carbon (+154 Pg C) than the CMIP5 models? It doesn't help that the signs make this analysis even more confusing. Since NEP and uptake were discussed, I assumed a negative value was a source of carbon. I know this confusion on signs is difficult to handle, but I just wanted to raise awareness that it contributed to making this section difficult to follow. I apologize if I got these comparisons incorrect, but that indicates that more careful discussion and use of terminology would be helpful.

The section "Land carbon response" discusses the impact of the historical forcing on some carbon related variables, i.e. gross primary production (GPP), leaf area index (LAI) and net ecosystem exchange (NEE) and their interactions.

We actually do not discuss absolute values for GPP in this section, we focus mainly on interannual variability (IAV) and trend. Absolute values of GPP (i.e. mean GPP for present day) are discussed and compared against observations in section 5.1.1. "GPP".

The last two paragraphs analyse the total land carbon uptake over the historical period, which is the sum of the net ecosystem production NEP (opposite sign to NEE, i.e. $NEP = -1 \times NEE$) from 1850 to 2005. Throughout the paper we consistently analyse the flux to the atmosphere (i.e. land to air and sea to air) which is commonly used for analysing CMIP5 modelling results. However, in order to calculate the uptake by land and ocean we need to reverse the sign. We will clarify this in a revised version of the paper.

The value of 154 PgC represents the total land carbon uptake over the historical period for the scenario with prognostic LAI (i.e. cumulative NEP). Note, this value will be corrected in a revised version (see also reply to reviewer 1).

We currently do not consider disturbances such as land use and land cover change (LULCC) in our simulations, which means that our land carbon uptake is simply calculated based on NEP. The majority of CMIP5 models include LULCC in some form or the other, which makes it difficult to compare our calculated uptake against land carbon uptake from CMIP5 models. We tried to do this in two ways:

(a) we compare our results against cumulative NEP with values reported in Shao et al. (2013) with NEP ranging from 24 to 1730 PgC, which means ACCESS-ESM1 is at the lower end of this range.

(b) we compare our results against observational based estimates of land use emissions which are thought to be 108-188 PgC for the historical period. This means we get an almost neutral behaviour by accounting for LULCC in this way. CMIP5 models that include disturbances also estimate a neutral behaviour by providing an estimate of land carbon uptake of -59 to 18 PgC (Shao et al., 2013) and -124 to 50 PgC (Jones et al., 2013).

We will revise the whole section accordingly.

Ocean Response

I thought the second half of the discussion in this section was clear and informative; however, I had a two questions on Figure 6. First off, the caption states that it's "Integrated Primary Production," but doesn't define what that is. I assume that's the same as NPP? I'm unfamiliar with that terminology, so it might be worth

clarification. Second, the values in the text and the figures don't match up. In the text the global mean ACCESS-ESM1 NPP is 46 PgC/yr, but from what I can tell this must be for the entire simulation. This is then compared to SeaWIFS value of 52 PgC/yr for 1998- 2005. Upon reading this, I expect ACCESS- ESM1 to be lower than SeaWIFS in Figure 6; however, looking at it, ACCESS-ESM1 is higher than SeaWIFS for these years. I personally think it would be better if these comparisons were values representing the same time period, both to have a fair comparison and to match up the values with what is seen in Figure 6.

We apologise for any inconsistency, it should have been 51 PgC/yr, and we have now clarified the text to replace integrated primary production with net primary production and added a section explaining how NPP is calculated.

GPP

I would consider removing the first sentence in the second paragraph, as it is subjective and is not needed with the supporting text. I would then combine the first and second paragraphs to have a complete discussion. I would also remove or modify the final sentence in the second paragraph that states that containing nitrogen and phosphorous “ensures a more realistic simulation.” While I think there is evidence that including nitrogen and phosphorous is beneficial in many circumstances, that alone does not ensure a model outperforms one where these are not included but instead has more realistic representations of other important processes.

As suggested by the reviewer we will remove the first sentence of the second paragraph and combine the first two paragraphs in a revised version. We will change the final sentence of the second paragraph to:

ACCESS-ESM1 contains both nitrogen and phosphorus limitation, which may provide a more realistic simulation of carbon cycle uptake by the terrestrial biosphere.

The second section in this section discussing the mean annual cycle of GPP is again quite subjective, and perhaps it wouldn't be too difficult to provide a few quantitative statements.

We will include more quantitative statements in the discussion of the mean annual cycle of GPP in a revised version.

For the final discussion in the section on IAV, first off I was wondering how IAV is calculated? I can think of a couple of different methods, and it is not defined how they actually calculate the values that are shown. For this section, I suggest a PDF of errors, therefore when you say there is good agreement in the spatial pattern, it can be backed up with “x% of the globe has errors less than x kg C/m2.” I will also note that the labels in Figure 9 were particularly hard to read.

The interannual variability (IAV) is calculated as the standard deviation for the de-trended annual mean values. This explanation will be included in a revised version.

As suggested by the reviewer we have calculated the absolute error for present day mean GPP for each land grid point. For example, 95% of all land points have errors smaller than 0.5 kgC/m2/yr for the scenario with prescribed LAI (86% for the scenario with prognostic LAI). We will include those numbers in a revised version.

We will increase the size of labels in Fig. 9 in a revised version.

CNP Pool Sizes

The discussion on the HWSD and soil carbon I found to be quite subjective, again focusing on the comparisons being “good” or “generally good.”

For the nitrogen comparison, first off please clarify what the value reported, 85 Pg N, represents (i.e. global over entire simulation?) And just a thought: it might be interesting to show or state how this evolves in time, similar to the time-series shown for carbon.

The 85 PgN represent the mean soil organic (SOC) pool size for the last 20 years of the historical period (1886-2005). In the manuscript on page 13, line 2 we stated that all pool sizes are calculated over the last 20 years of the historical period.

We will include initial pool sizes (i.e. spun up pools from pre-industrial simulation) for CNP for both scenarios in a revised version in Table 2 so that they can be compared against present day pool sizes:

Pool	Pre-industrial						Present day						Historical change C	
	PresLAI			ProgLAI			PresLAI			ProgLAI			PresLAI	ProgLAI
	C	N	P	C	N	P	C	N	P	C	N	P	ΔC	ΔC
Biomass	611	5.7	0.31	731	6.15	0.33	670	6.2	0.34	807	6.8	0.37	69.5	87.2
Litter	117	0.85	0.04	149	1.02	0.05	126	0.9	0.05	163	1.1	0.06	7.6	12.3
SOC	1034	82	9.6	1187	86.1	11.9	1050	83.4	10.1	1217	88.5	12.6	20.5	37
Σ	1762	88.6	10.0	2067	93.3	12.3	1846	90.5	10.5	2187	96.4	13.0	97.6	136.5

For the phosphorus discussion, I would suggest either removing the “slightly” modifier used in the discussion on how the model results are lower than the estimated range, or give references on how this is smaller than previous modeling estimates to give a frame of reference for that adverb.

We will remove the word “slightly” as suggested by the reviewer in a revised version.

Ocean Carbon

The figures in this section are out of order. I would recommend swapping Figures 12 and 13, rather than discussing Figure 13 first and then going back to 12.

We have now swapped the figures consistent with the suggestion of the reviewer.

Ocean NPP

This section shows and discusses mean seasonal cycles; however, after discussing seasonal aspects such as shifts in timing, the section ends with the disclaimer that the timing cannot be compared. I’m wondering, if that’s the case, might it be better to just look at amplitudes in a different format, such as a table or bar graph? If you leave the figures, I would recommend moving this statement to the beginning of the

discussion and making it clear the timing aspects are only being compared between ACCESS- ESM1 and CMIP5. With the figure as it is currently, I again found the accompanying text subjective, which might also support looking into a table or bar graph format for this section to provide more quantitative comparisons.

We regret any misunderstanding; the statement or caveat here refers to the challenge of comparing the response only in the tropical ocean given that ENSO cycles have a strong influence on ocean productivity. We have now reordered the text in this section to make it clearer for the reader and provided more insights into the mechanisms driving the differences between ACCESS-ESM1, observations and CMIP5. We are keen to stick with the plots to highlight the differences in both the magnitude and phase of the seasonal cycle. Additionally, we have also added that the plots and text refer to the observational period 1998-2005, consistent with Anav et al (2013). The manuscript now states:

To assess the seasonal anomaly of Net Primary Production (NPP), calculated as the anomaly of vertically integrated primary productivity through the water column, the global ocean is broken down into 5 regions, following Anav et al (2013). Figure 12 shows the NPP seasonal anomaly from ACCESS-ESM1, CMIP5 models and SeaWIFS over the (SeaWIFS) observational period 1998-2005. At the global ocean scale, seasonally we see that the magnitude of NPP from ACCESS-ESM1 is less than the amplitude of CMIP5 and SeaWIFS, with poor phasing. This likely reflects the biases in ACCESS-ESM1 toward lower latitudes, reflecting excess nutrient supply, and utilization, to the upper oligotrophic ocean (Law et al 2015) associated with deeper than observed mixed layers.

In the northern and southern subtropical gyres ACCESS-ESM1 (18N-49N and 19S-44S respectively) appears to overestimate the amplitude of the observed seasonal cycle when compared with SeaWIFS. Again this overestimate of NPP is associated with deeper than observed mixed layers which increase nutrient supply to the oligotrophic upper ocean. The phase of the NPP in these regions, where agreement between observations and CMIP5 is very good, is delayed by about three months. This delay may also be explained by a combination of higher (than observed) concentrations of nutrients and slower than expected biological productions associated with cool biases, particularly in the Atlantic Ocean allowing the bloom to occur later.

In the high latitude northern hemisphere, the magnitude of the seasonal cycle of NPP is not well captured in ACCESS-ESM1. While CMIP5 appears also to underestimate the magnitude of the seasonal cycle, ACCESS-ESM1 is lower again. In contrast in the Southern Ocean the amplitude of the seasonal cycle of NPP in ACCESS-ESM1 shows good agreement with observations. However in the high latitude oceans the phase of NPP is delayed by about 2 months. This delay may be attributed to the too shallow mixed layers that exist in these regions, which means that it is only when mixed layers start to deepen that biological productivity can start to occur. As a result the remaining growing season is shorter (than observed) leading to a reduced total productivity. This may in part explain why the total NPP northern hemisphere is much less than observed.

Interestingly, in the tropical ocean we see very good agreement in the amplitude of the seasonal cycle with CMIP5 and SeaWIFS. We note however, that comparing the phase of the seasonal cycle from ESMs (ACCESS-ESM1 and CMIP5) with SeaWIFS is not very meaningful in this region as they all simulate their own ENSO cycle with their own timing. Therefore any comparison over a 20 year period between models has the potential to be biased by the number of El Nino or La Nina events.

I also want to note on Figure 13 the Southern Ocean has different values on the x- axis, which was not mentioned. This was confusing because at first it appears the Southern Ocean does not have a seasonal cycle, but from what I can tell (which is difficult given the large range), the amplitude is comparable to the other oceanic regions.

Clearly it is desirable to have all of the plots on the same scale. However, in order to highlight all features, different scales cannot be avoided. We have now highlighted in the captions that we use different scales in Fig.13.

Ocean Sea-Air CO2 Fluxes

In this discussion, the authors state that the Southern subtropical gyres overestimate the observed sea- air flux; however, when I look at the figure, it looks to me the biggest uptake is occurring on the coastlines. I personally would like more of a mention of this, and maybe a discussion on why coastlines, particularly off the west coast of Australia, are taking up so much carbon compared to the broader ocean gyres.

I think that role of the coastal ocean in the global carbon cycle is very interesting and has been suggested by authors such as Borges et al (2005). However coastal oceans in ESMs is not well represented in terms of the key mechanisms in CMIP5 models (Bopp et al, 2013). Nevertheless, despite the claims of some authors e.g. Borges et al (2005) that the coastal oceans are playing a very large role, the role likely remains small relative to the gyres, primarily because the surface area of the coastal ocean remains very small relative to the size of the ocean gyres. Furthermore, the heterogeneous response of the coastal ocean means that observationally the response coastal ocean remains more poorly constrained than the ocean. Nevertheless, as more observations are collected, resolution improves in ESMs allowing a better the representation of the key processes in the coastal ocean well better it is likely these models will focus more on these regions, as this is where the impacts of climate change and variability are felt most acutely.

Borges, A. V., Bruno Delille, and M. Frankignoulle. "Budgeting sinks and sources of CO2 in the coastal ocean: Diversity of ecosystems counts." *Geophysical Research Letters* 32.14 (2005).

In the second paragraph of this section, the second sentence beginning with "Furthermore it appears that globally...." doesn't make sense to me, as I don't understand what it is that lies outside the range. Usually regional analyses reveal why global results occur, and it looks like the Southern Hemisphere is the main contributor to the ACCESS- ESM1 global seasonal flux anomaly? The text does go on to state this, so you may just consider removing or modifying that sentence. Later in that same paragraph, the text states that the Northern Hemisphere has fluxes larger than observed but within the range of CMIP5; however, to me it looks like there are several months when ACCESS- ESM1 is outside the range (i.e. Jan, Feb, May). Since the same disclaimer was put on this paragraph (which again I would suggest moving to the beginning of the paragraph if these figures are kept), I would consider a different format, such as a table or bar graph, or even keeping the seasonal figures but not showing the seasonal cycle in the observations if you don't believe them, and instead using a solid line.

We have rewritten this section to clarify our intended meaning, we have also used the regional plots to explain where and why the spatial biases exist. As above we have also chosen to keep the figures as these are both helpful in our explanation and consistent with previous CMIP5 assessment paper e.g. Anav et al 2013. The text now states:

The anomaly of the seasonal cycle of the sea-air CO₂ fluxes was assessed against observations of W13 and CMIP5, shown in Fig 14 for the period 1986-2005. Here we see that ACCESS-ESM1 has larger global amplitude of sea-air CO₂ fluxes than observed (W13) and simulated but close to the upper value of the range from CMIP5 models. We also see that globally the phase of sea-air CO₂ fluxes is not well captured in ACCESS-ESM1, lying outside the range of the CMIP5 models. To better understand why there are differences between ACCESS-ESM1, CMIP5 and W13 we separate the response of sea-air CO₂ into the same regions as for NPP, again following Anav et al (2013).

ACCESS-ESM1 appears to capture well the phase of sea-air CO₂ fluxes in the subtropical gyres. In the northern subtropical gyre in particular, we see that the amplitude and phase of the seasonal cycle in ACCESS-ESM1 shows very good agreement with W13, in contrast with other ESMs (CMIP5). In the southern subtropical gyres, while the ACCESS-ESM1 appears to overestimate the amplitude relative to the observations, we see very good agreement with CMIP5 models. As anticipated the tropical ocean shows very little seasonality, nevertheless we do see good agreement with CMIP5 models. However, the comparison of ACCESS-ESM1 against observations (while shown) is not very meaningful as W13 is based on values of oceanic pCO₂ from Takahashi et al, 2009 which does not include El Nino years.

The largest differences are seen in the representation of sea-air CO₂ fluxes in the high latitude ocean. In the high latitude northern hemisphere, we see that the magnitude is larger than either CMIP5 or W13 and shows poor phasing. While the magnitude of the seasonal cycle in the Southern Ocean lies within the upper range of CMIP5 again poor phasing is seen. That the seasonal cycle is out of phase suggests that during the summer the solubility response likely dominates over the NPP response, leading to an out-gassing in the summer and uptake in the winter, as discussed in Lenton et al (2013). Consequently, we see that the poor global phasing in global sea-air CO₂ fluxes is likely due to the solubility dominated response of the high latitudes during the summer.

Anthropogenic Inventory

The discussion is again quite subjective. Also, in Figure 15, there is a reference to Key et al. (2004) that is confusing and not in the text (I'm unsure if the figure appears in both papers?) Upon finishing the paper, I see the reference in the conclusions, but you might want to somehow clarify or include it in the text during this discussion.

Key et al (2004) refers to the data used in the paper to compare with ACCESS-ESM1 while the estimated anthropogenic CO₂ uptake is from Sabine et al (2004). We have ensured that the correct and appropriate references are used in the figure caption, section 5.2.3 and the conclusions.

Atmospheric CO2

I think it would be better to combine the first paragraph with the next sentence following it (starting “Therefore, our...”).

We will do this in a revised version.

Conclusions

I think the conclusions would be stronger if some quantitative assessments were provided. I also noticed the order of the conclusions was not the same as the order presented in the text, but neither was it combined in a more succinct fashion to group conclusions. Simply reordering the paragraphs would make the text more consistent.

We will reorder the paragraphs in a revised version and also include some quantitative assessments.

In the second- to- last paragraph, I was confused by the statement that “Seasonally the ACCESS- ESM1 appears biased toward the Southern Hemisphere.” I’m not sure how to interpret this statement, as to me it’s using both temporal and spatial references combined in an unclear fashion. This same type of statement occurs later in the sentence “Globally the annual mean is well captured but biased to low latitudes.” I recommend reworking those sentences to clarify the intent of the text. Also, a quantification of how it is “well captured” would strengthen the statement.

We have now rephrased this paragraph to better reflect our intended meaning and removed the very qualitative comments. We have not added numbers to the conclusions but instead have added more quantitative analysis to the text that is referred to in the conclusion, the text now states:

Globally integrated sea-air CO2 fluxes are well captured and we reproduce very well the cumulative uptake estimate from the Global Carbon Project (LeQuere et al, 2014) and our anthropogenic uptake agrees very well with observed GLODAP value of Sabine et al (2004). The spatial distribution of sea-air CO2 fluxes is also well reproduced by CMIP5 models and observations. At the same time global ocean NPP also shows good agreement with observations and lies well within the range of CMIP5 models. However seasonal biases do exist in sea-air CO2 fluxes and NPP, potentially related to biases in mixed layer depth and surface temperature that are present in ACCESS-ESM1; and will need to be addressed in later versions of ACCESS-ESM1.

Technical Comments

Overall I have very few technical comments. My main technical comments are on comma usage. I personally find well- placed commas can aid in reading the text, but am not a comma expert, so feel free to take or leave the suggestions.

All technical comments will be considered and corrected in a revised version.