Response to Referee RC1 (Chris Jones):

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.

My main concern with this model is the length of time taken to achieve a spun-up state. Law et al document this nicely, but I think it requires more discussion in this paper too what the implications are. The drifts in carbon stores are still non-egligible even after 800 years of spin-up (your start point here). I think this should be laid out explicitly before the analysis starts. You do, in the land carbon section, acknowledge this and subtract the control run drift. But unless a reader has been through Law et al they would not know how big this drift is. For the ocean it is more important still, and the ocean section does not mention this at all. The drift of circa 0.7 GtC/yr (figure 11a in Law et al) is of similar magnitude to your historical fluxes (I assume these are corrected for the drift). If a reader hadn't seen that figure then they would not realise from this paper the size of drift being subtracted.

We will include a paragraph in the Simulations section 2.2 to acknowledge the drift in carbon stores more explicitly for land and ocean. We will also make it clear in the ocean section that the historical fluxes have been corrected for the drift.

As noted in Law et al. (2015) the net carbon fluxes for land and ocean did not equilibrate to zero. At the end of the control run (i.e. year 800 to 955), global NEE is 0.3 PgC/yr for PresLAI and 0.08 PgC for ProgLAI. The net autgassing from the ocean is about 0.6 PgC/yr at the end of the control run. We take this drift into account when we calculate the net uptake of carbon for land and ocean.

In CONCENTRATION-driven runs like this, you can of course force the correct CO2 and correct for the drift after the simulation. But that is not possible in an EMISSION-driven run, and such a drift would cause a massive drift in atmospheric CO2 render-ing an emissions-driven historical run meaningless. This would, at present preclude use of this model in C4MIP for example which would be a great shame. The latest C4MIP protocol (Jones et al 2016, GMDD - CMIP special issue) recommends a maximum acceptable drift of 10 GtC per century. I therefore throughly recommend that ACCESS modellers attempt to find accelerated means to derive a spun-up state in time for CMIP6. There are numerous options, such as running offline (for either land or ocean), or using reduced turn-over time techniques as per Koven et al for CLM (http://www.biogeosciences.net/10/7109/2013/bg-10-7109-2013.pdf). as a final word on this, lack of carbon conservation would also be more of an issue for E-driven runs.

We are currently exploring various spin-up options (including offline simulations) to improve the drift for ocean and land. If successful, this will then be implemented in future versions of ACCESS-ESM.

My second concern is the lack of land-use change as a forcing. You already know this, and acknowledge it in the paper, so no revisions to the manuscript are required, but I just take this opportunity to stress that simulations of contemporary and future climate/carbon cycle are very much reduced in usefulness if they lack the

very large land-use forcing of the land carbon cycle. Implementing this for CMIP6 is also therefore a priority I would say.

We agree with the referee and it is a high priority for us to include land use change in future versions of ACCESS-ESM.

Having quickly read Law et al before I reviewed this one I was struck that there was not an evaluation there of (land) carbon stocks. I do feel that the land carbon modelling community have become fixated on evaluation against fluxes to the detriment of stocks and residence times. This is beginning to change and I was pleased to see some discussion of carbon stocks in this paper. It would be nice to see the time changes in these though as well - could table 2 be extended to show pre-industrial and present day stocks? In the discussion on biomass you mention that your results are higher than observational based estimates - but of course you lack land-use change as a driver. So it could easily be expected that your biomass is of the order 100-150 GtC too high due to this. If you masked for present day agricultural regions you would probably get a much closer fit to expected global totals. So your simulation is actually not bad.

In a revised version we will extend Table 2 to show both, pre-industrial and present day stocks. We thank the referee for the comment about the size of our biomass pool and we will include this statement in a revised version.

	Pre-industrial						Present day						Historical change C	
	PresLAI			ProgLAI			PresLAI			ProgLAI			PresLAI	ProgLAI
Pool	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
$\sum$	1762	88.6	10.0	2067	93.3	12.3	1846	90.5	10.5	2187	96.4	13.0	97.6	136.5

Abstract - you can say that aerosol forcing is large or larger than other models. But don't say "over-sensitive" as we simply don't know. Maybe this is correct ...

We will change this statement as suggested in a revised version.

section 2.3 - I like the comparison vs CMIP5 models. This is a nice way to put model-data discrepancies in context. But why do you only include 5 CMIP5 models (counting the 2 IPSLs as one model). Anav et al used more than twice that - any reason not to use the full set?

We agree that ideally we would have been able to use more Earth System Model (ESM) results like Anav et al., though unfortunately, at the time we were doing the analysis not all the output was readily available from the ESG servers. After careful examination of our final results and those of Anav et al., we are confident that our figures still capture the spread of CMIP5-ESM results and more models may only serve to broaden the spread of the models.

in a couple of places (e.g. start sec 3.1) you mention the variability of the land

sink and/or the atmospheric CO2. You could go further and use this as an evaluation metric. Both Cadule et al (2010, GBC) and Cox et al (2013, Nature) show the power of the C-cycle sensitivity to ENSO on inter-annual timescales as a really strong evaluation metric and constraint on the sensitivity of ESMs

We thank the referee for his suggestion and will consider this in future evaluation papers using the next version of ACCESS-ESM.

p.7, line 21. I don't understand why you attribute your slower warming to the initial warm bias - how do you know the warm bias causes this? Could just be under-sensitive SSTs not related to a bias.

The discussion was intended to be descriptive rather than mechanistic. We clarify now the text so as not to attribute the reduced warming to the warm bias early in the historical period.

p.8 line 3. I was amused to see that having an error different from other models was "encouraging"! not sure why! Can you say why it is better to have the opposite MLD bias from other models?

Encouraging in the sense that simulating well winter mixed layer depths are critical for setting interior ocean properties supplying nutrients to the upper ocean to fuel the biologically active growing season. We have now augemented the manuscript to say:

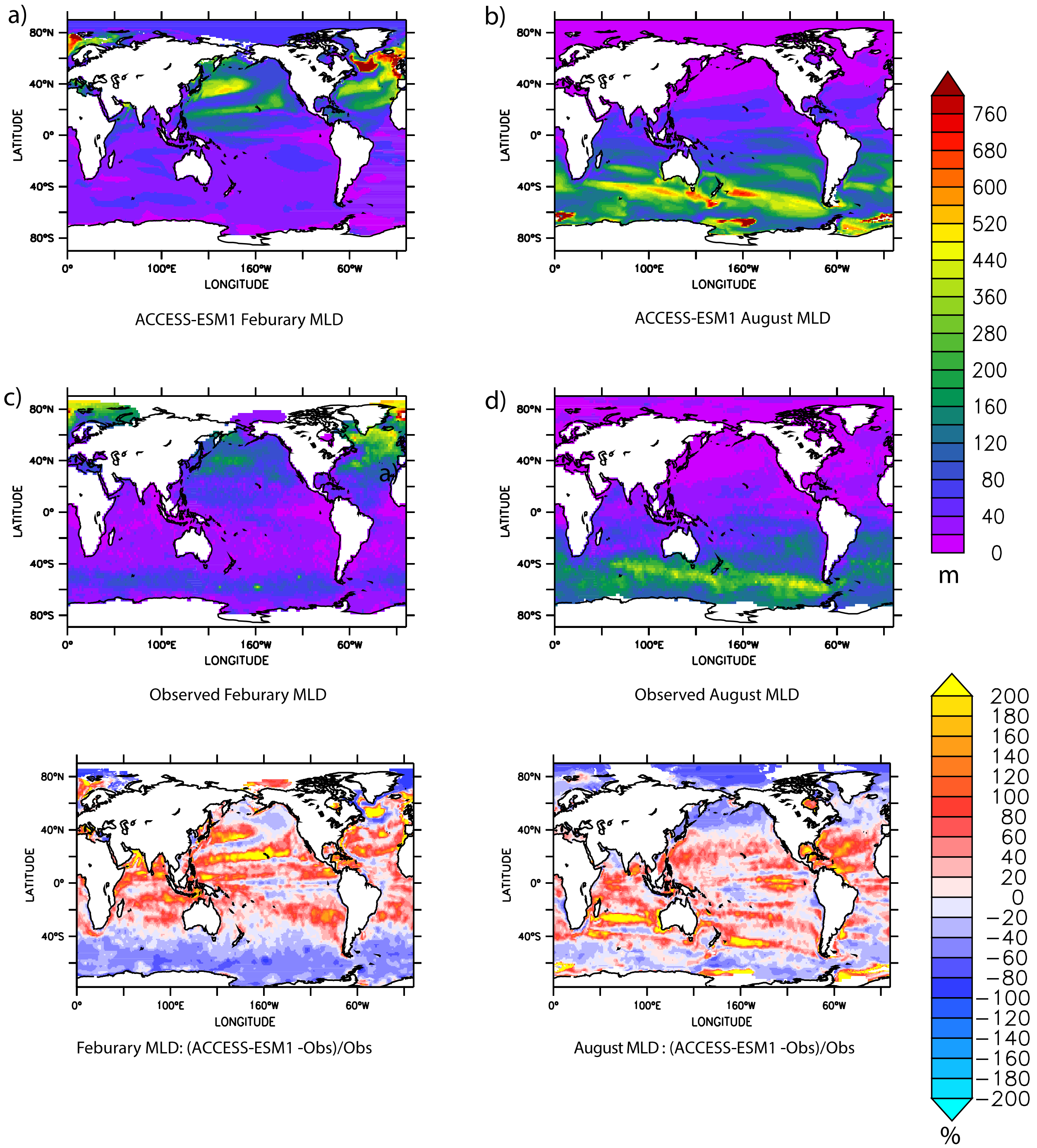
In the higher latitude the winter mixed layers are well captured by ACCESS-ESM1 Figure 4. This is encouraging given that many ocean models tend to underestimate winter mixed layer depths (Sallee et al (2013) and Downes (2015)). Simulating winter mixed layers correctly is critical for setting interior 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).

in general I thought this MLD section (3.2) was a bit superficial - can you be a bit more quantitative in your comparison and description? the figure shows the data but it can be hard to tell from there if the differences you describe are of the order of a few % or 10s of % or factors of 2 or more or what? And can you mention your confidence in the ons? presumably global maps of MLD are not directly observed but must have certain extrapolation uncertainties and so on. Are some areas/seasons of the oceans better sampled than others etc …

To better illustrate the chages we have added a lower panel to Figure 4 that is the percentage changes of the difference between the (ACCESS-ESM1 – obs/obs) *100. This illustrates the relative changes as suggested by the reviewer. Also please see above for an enhanced discussion.

Regading the number of samples and coverage of the obs we have added the following paragraph:

Ocean mixed layer depths are compared with the observations following DeBoyer Montegut et al (2004) based on more than 880,000 depth profiles from research ships and ARGO profiles, and based on a 0.03kg/m^3 density change from the surface. Significant advances in autonomous measurement platforms have allowed the mixed layer to be increasingly well constrained in all seasons across the global ocean.



p. 8 line 21. Can you define what you mean by IAV. Interannual Variability I know, but how do you turn this into a number? is it the standard deviation of a time series of annual means? in which case is the time series de-trended first? Etc

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.

p.9 When discussing historical changes in land carbon can you split into veg and soil changes (e.g. put in table 2). You could compare directly with the 2 models in Jones et al (2013) which also don't have land-use forcing (dashed lines in figure 2). You could probably also compare with model results from detection-and-attribution studies which ran with/without certain forcings. The are probably various no-land-use runs to look at.

We will included the change in land carbon for vegetation, litter and soil for both scenarios (prescribed LAI and prognostic LAI) in Table 2 and also compare these values against simulation of the two models without land use change from Jones et al. (2013) in a revised version.

	Pre-industrial						Present day						Historical change C	
	PresLAI			ProgLAI			PresLAI			ProgLAI			PresLAI	ProgLAI
Pool	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
$\sum$	1762	88.6	10.0	2067	93.3	12.3	1846	90.5	10.5	2187	96.4	13.0	97.6	136.5

When we calculated the change in land carbon based on the pool sizes rather than using the net flux, we noticed that the total land carbon uptake over the historical period is actually much smaller than stated in the paper: 98 PgC instead of 128 PgC (PresLAI) and 137 PgC instead of 154 PgC. This is because we used an earlier section of the control run (years 325-480) to calculate the drift. However, the historical runs we describe in the paper were started at year 800 from the control run and this section (i.e. year 801-955) shows a much smaller drift for both scenarios. Using the correct drift applied to the net flux over the historical period provides now about the same total land carbon uptake as calculated via the pool sizes. The numbers for the total uptake of carbon by the land will be corrected in a revised version.

We have also updated the NEE time series plot (Fig. 5c) using the correct drift, although changes are hardly visible, particularly for the 5yr running mean.

Sec 4.2. Despite the large drift your historical ocean sink does look a very close match to the obs. Can you also quote a cumulative uptake here?

We have added the following text to the paper:

The cumulative uptake of carbon by air-sea CO_2 fluxes in the period 1959-2005 from ACCESS-ESM1 is 83 PgC which is good agreement with the GCP value of 82PgC over the same period.

sec 5.1.2. I was curious to see that your prescribed LAI didn't match your obser-

vations. Can you explain why not? You illy that this is because there are differences between observed datasets - which is of course true. But then why do you choose one dataset to prescribe LAI to the model, and then a different one to evaluate against? If one is better than the other can you use it for both?

The prescribed LAI we use in ACCESS-ESM1 is based on MODIS observations and has no interannual variability. We decided to compare this with a more recent LAI product (MODIS/AVHRR combination) to (a) investigate if there are significant differences in the mean seasonal cycle for present day (as shown in Fig. 10) and (b) to investigate what the interannual variability in the observations looks like (not show in the manuscript).

We don't think that one LAI product is better than another. Historically, CABLE uses a prescribed LAI with no interannual variability. However, in future CABLE/ACCESS versions we might update our prescribed LAI to a product with interannual variability.

As mentioned in the paper, LAI products differ from each other because different sensors and algorithms are used. In addition, the LAI products shown here are also based on different observing time periods.

p.13 discussion of different carbon stocks for the two configurations. Given both have very similar GPP, the large difference in biomass is presumably due to residence times? Having GPP further north in compared to the tropics means that for the same global GPP you have a higher biomass? You might consider next time an evaluation of turnover times - e.g. as per Carvalhais et al (2014, Nature)

We agree with the referee, that the difference in carbon stocks for the two scenarios (prescribed LAI and prognostic LAI) can probably be explained by a difference in residence time. We thank the referee for pointing this out and we aim to include an evaluation of turnover times in future work.

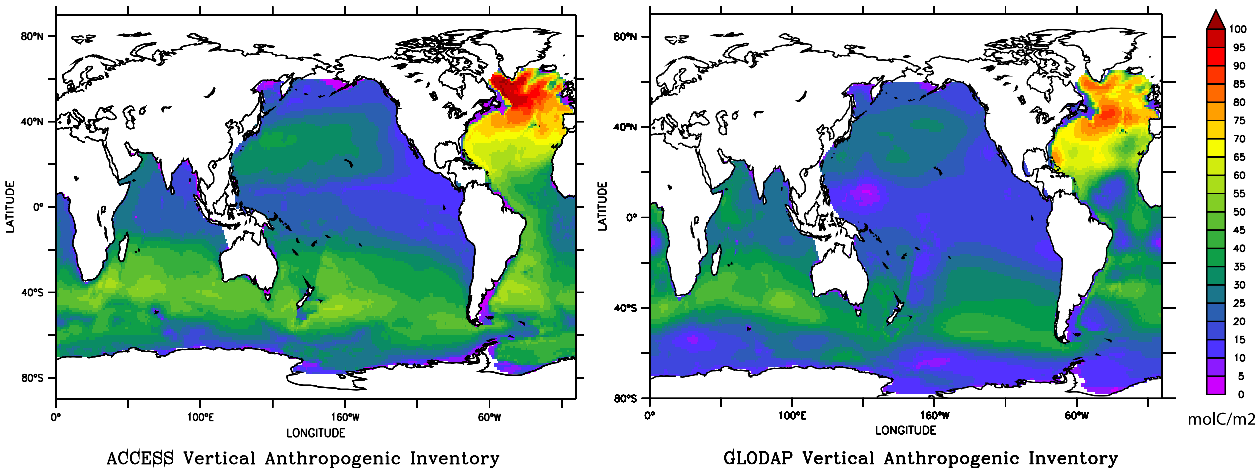
section 5.2.1 / 5.2.2. Can you swap the order of figures 12 and 13? you describe 13 first.

The order of these figures will be changed in a revised version.

sec 5.2.3. You compare global totals between model and GLODAP, but the map shows missing areas in the ons dataset. So should you first mask out the model to match the same area before you compare totals? (or quote a full AND a masked number for the model). The values in the Arctic for example are very high but missing in the obs (although of course the area is smaller than this projection makes it look).

We have redrawn the ACCESS-ESM1 map with only ares that observations exist. Furthermore we now quote the number from comparison with GLODAP and for the entire domain:

The inventory for the period 1850-1994 in ACCESS-ESM1 is 132 PgC, which is close to the estimated value from GLODAP of 118 +/- 19PgC (Sabine et al, 2004) over the same domain. This suggests that despite a somewhat limited representation of the seasonal cycle of sea-air CO_2 fluxes in key regions of anthropogenic uptake such as the Southern Ocean, ACCESS-ESM1 is doing a very good job, spatially and temporally, of capturing *and storing anthropogenic carbon. If the entire domain (including the Arctic Ocean) is integrated the anthropogenic uptake is 143 PgC over the same period.*



sec 5.3. It seems reasonable that the land drives most of the seasonal cycle at your ons stations except the south pole. I like the way you have labelled the land and ocean CO2 separately to be able to diagnose this. But I couldn't work out why the contribution of the ocean at the south pole looks different for your two LAI configurations. If I read the figure correctly then the role of the ocean is given by the blue solid minus the blue dashed line and the red solid minus the red dashed line. These are quite different by eye - e.g. the blue lines are quite far apart in December and the red lines quite close. So why does your LAI treatment affect your December ocean fluxes so much?

The LAI scenario should not affect the ocean carbon fluxes and we thank the referee for pointing out this inconsistency. By mistake, we showed the total flux using two different ocean configurations in Figure 16 (mean seasonal cycle ot atmospheric CO2). Using the same ocean configuration for both scenarios (prescribed LAI and prognostic LAI) solves this issue and the dashed lines (land carbon flux) and solid lines (total flux) are much closer together (see figure below). Figure 16 will be updated in a revised version accordingly.

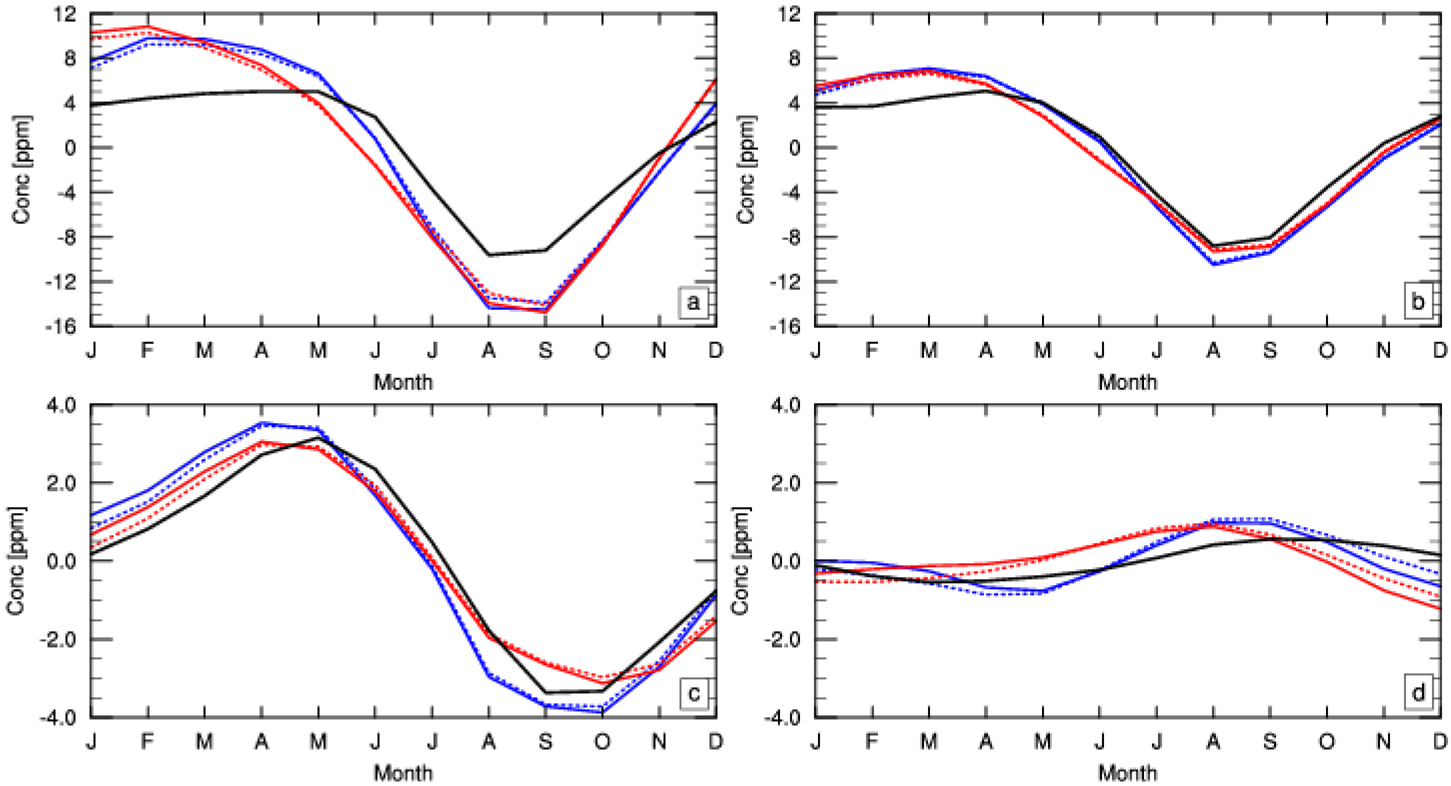


Table 2. Can you add a final row at the bottom of “total”

We will do this in a revised version.

	Pre-industrial						Present day						Historical change C	
	PresLAI			ProgLAI			PresLAI			ProgLAI			PresLAI	ProgLAI
Pool	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
$\sum$	1762	88.6	10.0	2067	93.3	12.3	1846	90.5	10.5	2187	96.4	13.0	97.6	136.5

general comment on figures - maybe a personal preference but please can you add a legend so that I can easily read which line is which. The text in the captions is very good - but I have to read a whole paragraph to spot what the red/blue lines are.

Would be great to simply see a legend with this in as well as the detail in the caption text.

A legend will be added to all relevant figures in a revised version.