Summary of Major Changes

The major issue raised by the reviewers is that, the original manuscript largely treats the models in aggregate, lacking in-depth analyses on individual models and change of environmental drivers. In response to this concern and other questions and comments raised in the constructive comments, we have performed additional analyses and closely examined the behavior of some individual models.

As a result, we have completely rewritten the discussion section, and added six Figures. We also included ten supplemental figures. Specifically, the following items have been changed.

1. Abstract. The abstract has been modified for clarity.

2. Introduction. Amplitude figures for Mauna Loa and Point Barrow CO2 are updated. Clarification on why the amplitude figures are different from Graven et al. (2013) is given. Several new references including Randerson et al. (1997) and Peylin et al. (2013) are added, including some descriptions. Minor textual changes are also made.

3. Results.
Sect. 3.1. Compared our results, with Gurney and Eckels (2011). Merged some results from Sect. 3.2 into this section, added relative amplitude change for the Northern Hemisphere (25-90N).

Old Sect. 2.2 and Figure 3 are removed.

Sect. 3.2 (old Sect.3.3). Redesigned Figure 4 (now Figure 3) to include curves for individual models and detrended CO2 growth rate. As a result, the text is lengthened, and the features of several models are discussed, and cited Anav et al. (2013).

Sect. 3.3 (old Sect.3.4). The description of Figure 4c (originally Figure 5c) is rewritten for better clarity. Add description for new Figure 5 showing zonal amplitude of all models, which includes different phase. Also discuss further on individual models and add Figure S1. Our new figures and analyses should answer some major questions raised by both reviewer, on the importance of tropical regions and interpretation of Figure 4c.

Sect. 3.4. New section. Moves in description for sensitivity experiments, and displays sensitivity experiment results for the GFDL and IPSL models (Figure 6 and 7). Also analyzes the major mechanisms, the caveat and reason for not including the IPSL model.

Sect. 3.5. Rewritten the first sentence.Changed some wording and added statement for uncertainty. Change caption of Figure 6 to Figure 8.

4. Discussion. In this new section motivated by helpful comments from the reviewers, we include the following:

To address the concern that ensemble patterns may be dominated by only a few models, we show the spatial patterns of individual models in Figure S2, which display high consistency and support our original conclusion. Also added discussion on the choice of two time periods.

In response reviewer 2's interesting questions on the amplitude change of Mauna Loa and Point Barrow, we first compared the model simulated CO_2 at the locations of the two stations with observation. We present our new findings in Figure 9 and 10, and discuss the difference of our findings and the results in Graven et al. (2013).

We then stated the amplitude increase at Mauna Loa and Point Barrow by 2081-2090. Then we discussed the caveats of treating this result, commenting on over-represented and missing mechanisms in CMIP5 models and the uncertainty in CO_2 emission.

Next, we discussed on the difference of climate simulations by the CMIP5 models , showing changes of soil moisture and near-surface temperature for each individual model in response to reviewer 1's comment on the change of drivers (Figure S6-S9).

We followed this by presenting the change of ecosystem types in some models (adding Figure 11 and 12), and discussed the implication.

We wrap up our discussion with effect of land use/cover change, both showing crop cover change (which addresses reviewer 1's comment on the change of drivers) and stressing the missing representation of agricultural intensification.

5. Conclusions. The original "Discussions and Conclusions" are significantly shortened due to the separation. We also removed the last paragraph as suggested by both reviewers, and added our suggestions on future work.

Specific point-by-point responses to each reviewer are given below. Reviewer comments are in bold.

Reviewer #1

The increase of CO_2 seasonal cycle in the northern hemisphere has drawn considerable interests in the recent times although the attribution of causes remain unclear. This paper explore the effect of terrestrial biosphere fluxes and its change with time in relation with climate change. The topic is of interests to the earth science community. The manuscript is generally well written, but I have some concerns which should addressed before accepting for final publication in ESD.

Many thanks for your constructive comments. In our revised manuscript, we have added six figures, a new discussion section and many other changes. Please see the beginning for a list of major changes, and we hope they can address your concerns. After our work in the revision process, we are now even more confident in our

original main conclusions, and are excited about our new findings. We hope you find the revised paper interesting and informative.

p780, l13-15: Why only "temperate and boreal"? Recently Wang et al. (Nature, 2014) suggested the tropical biosphere also respond to climate change. One might think that the terrestrial biosphere in the tropics will respond to the climate change in one direction only. But given that there is dry and wet seasons in tropics (analogical with cold and warm temperature in the temperate/boreal region?) we might expect some seasonally sensitive response to the CO_2 flux. And there are changes in seasonal atmospheric transport and magnitude of fossil fuel emissions. I am not suggesting you to solve all of these problems in this manuscript, but some thoughts should be interesting to discuss.

Thank you for pointing out the importance of tropical biosphere. Previous analyses on CO_2 seasonal amplitude change mostly focused on the temperate and boreal regions. Our analyses are however performed globally. Our figure 4 shows the spatial pattern and latitudinal distribution of ensemble NBP amplitude change. The zonal mean –NBP change plot does suggest some seasonally sensitive response from the tropical/subtropical regions. However, the tropical area's net effect on global NBP amplitude increase is small: tropical regions north and south of equator seem to cancel out (showing yellow and green shades, respectively).

We realize that we have not fully explained this issue, and our original Figure 5 (now Figure 4) may be confusing since tropics has a different seasonal cycle phase which is not reflected in our figure. Thanks again for your helpful comment, we have now largely expanded the original section (now section 3.3) and added two figures (Figure 5 and Figure S1), and we hope you find the revised version informative.

We have cited Graven et al. (2013) in our original manuscript, whose analyses suggest atmospheric transport and magnitude of fossil fuel emissions may only contribute a few percent to the CO_2 amplitude change. We also added CO_2 detrended seasonal cycle of each individual model and their ensemble alongside with NBP seasonal cycles in Figure 3.

p782, l25: Need a reference here? e.g. Peylin et al. (BG, 2013)

We have added the suggested reference, thank you!

p784, l23: Clarify whether the seasonal amplitude calculation is applied to zonal mean or individual grid or global?

We have used two definitions of seasonal amplitude and one approximation in our text. The first one is computed from CCGCRV package, which we use only for average CO_2 or total NBP over global and (NEW) large regions in Figure 2 and 9, and for (NEW) Mauna Loa and Point Barrow CO_2 (Figure 9,10, S4,S5).

Our second definition is simply maximum minus minimum of seasonal cycle in each year/decadal average, especially if we compare the 1961-1970 and 2081-2090 periods. Examples are global/regional mean/total (Figure 3) or zonal total (Figure 5).
These first two definitions are quite close for analyses of monthly values.

We also have one approximation of global seasonal amplitude change, using the difference between the changes of dormant season (October-December) averages and peak growing season (May-July) averages. This approximation is aimed at examining the spatial and zonal contribution to global NBP amplitude change only. While the amplitude of this approximation is smaller than those of the two amplitude definitions, the spatial patterns should be highly related. We believe this is a reasonable approximation judging from Figure 3c.

Hope you find this explanation helpful!

p785, l25: How is the trends in FFE is considered?

We focused on the trends of NBP in our study. For FFE, we stated this in our original Methods section:
"Previous studies have limited the impact of FFE and FOA on trends in CO2 amplitude to less than a few percent change (Graven et al., 2013). Therefore we focused on examining the seasonal cycle of NBP in this study."
Our comparisons in Figure 2 and 3 seem to support that NBP change is the main driver for trends in CO2 amplitude.

p785, l2 : Not clear how the different fitted components are utilised here?

A detailed description that we cited is http://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html, which did a superb job in explaining the algorithms. For our study we just use the amplitude and detrended seasonal cycles (which we use to calculate the amplitude based on the means of two periods) computed by the CCGCRV package.

p786, l12: Will some of the differences between GFDL and CanGCM arise from the method employed for NBP calculation.

The NBP given by the models all represent the net carbon flux from atmosphere to land. We have carefully checked the global carbon budget of our equation (1) to make sure the models are self-consistent (which we stated in our Methodology section).

p786, l25ff: What if some models have peak uptake outside the months from May to July? please comment how is that going to affect your analysis using the ensemble means.

Only CanESM2 has peak uptake outside May-July (please see our new Figure 3, showing the seasonal cycles of individual models). Following your suggestion, we excluded it and BNU-ESM (for which October-December is not dormant season) and got very similar spatial patterns.

p786, l27: I am glad that you spelled this out. I am most curious how the model's behaviour differ depending on the differences in timing of the seasonal minima? Then of course taking 3-monthly mean may also have some implications for the model to model comparison.

The models in general are quite consistent in timing of the seasonal minima (except for CanESM2), and all models have their minima in June in the future period (2081-2090). We use the difference of 3-monthly means to determine the relative spatial concentration. From the graph we showed in figure 3c, this selection is reasonable for most models. We have also tried to just use June and October, and our main conclusions still hold (Figure R1, R2).

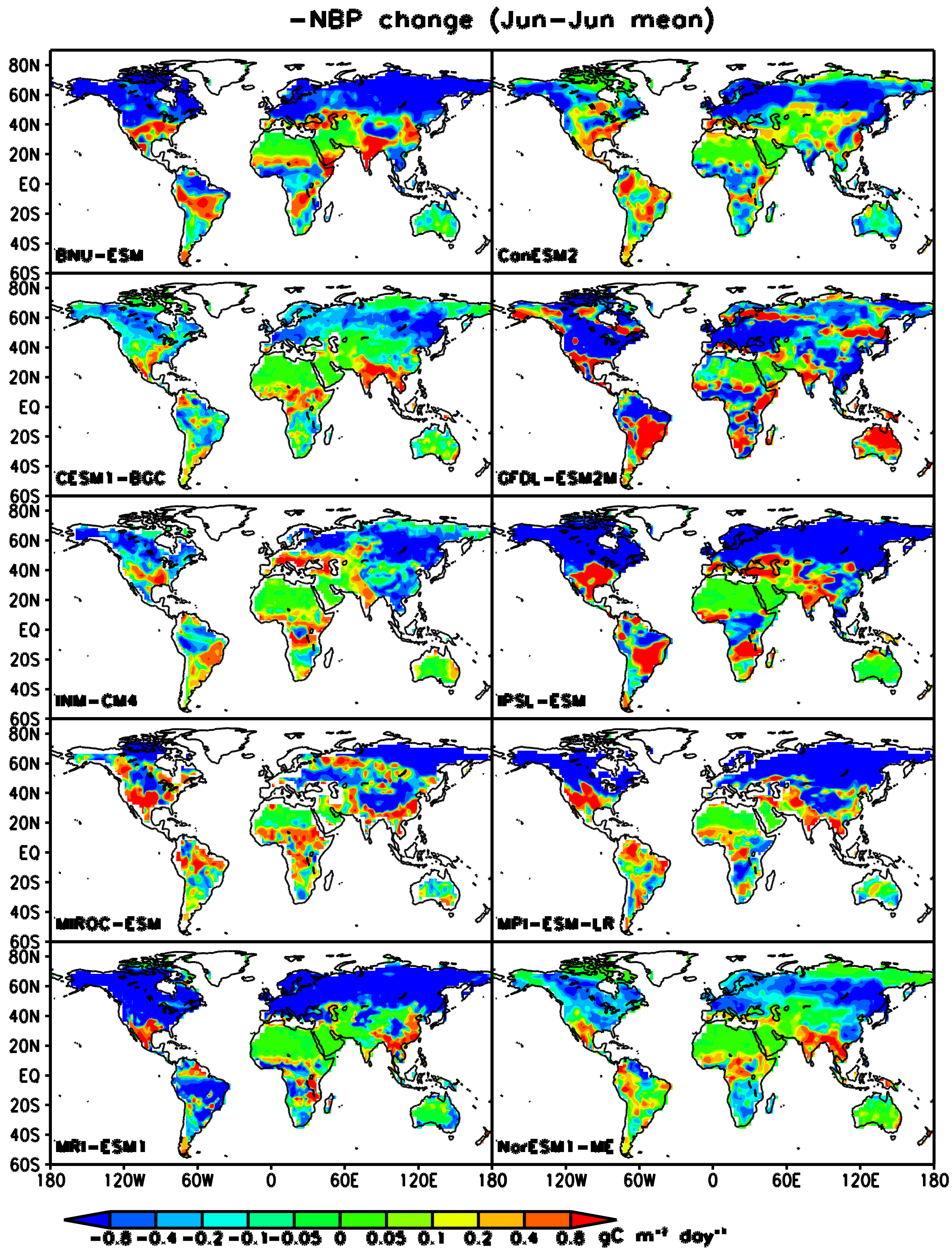

Figure R1. Spatial patterns of -NBP changes in June for the 10 models.

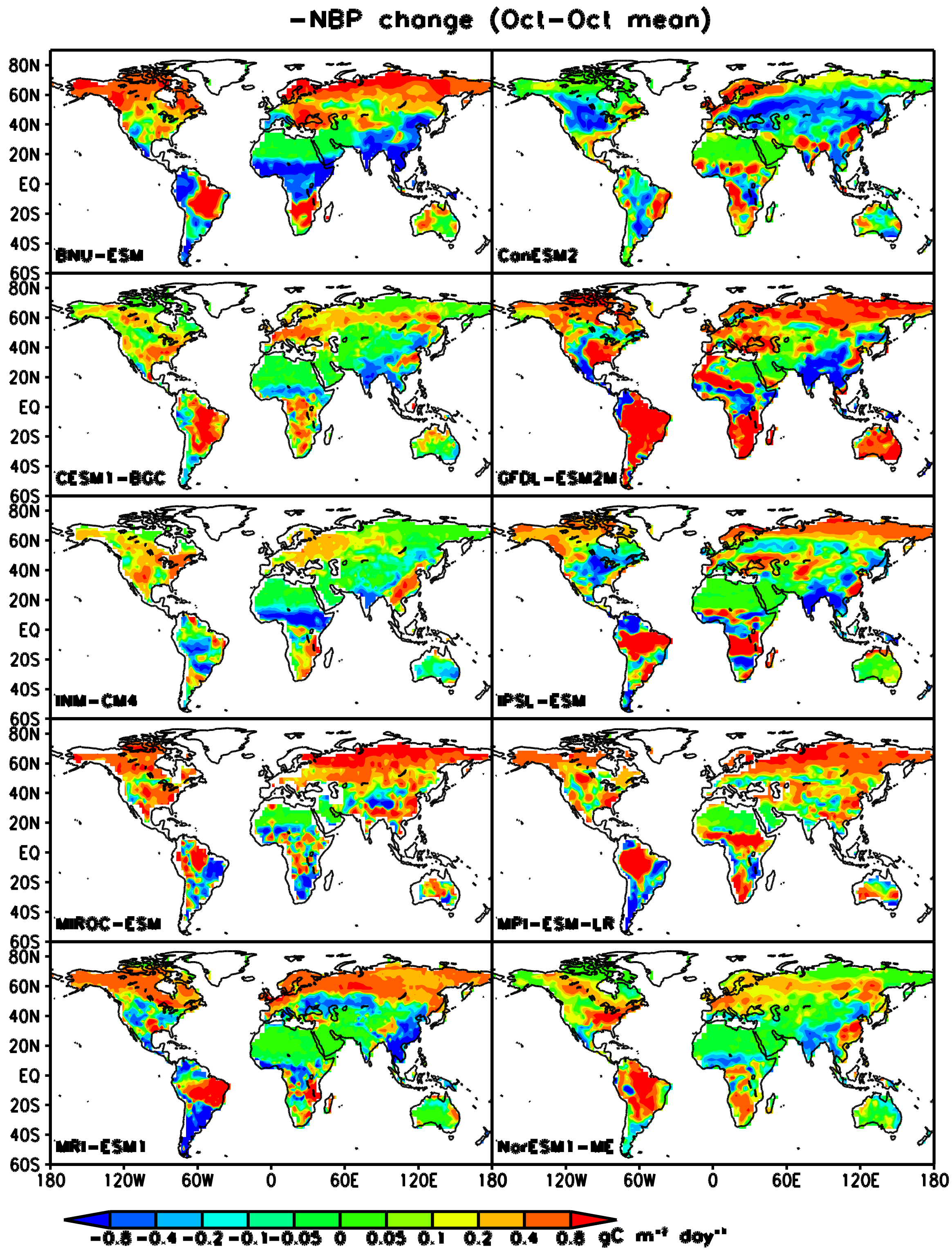

Figure R2. Spatial patterns of -NBP changes in June for the 10 models.

p787, l10: Replace minimums and maximums by minima and maxima, respectively, throughout the manuscript

We have made the suggested changes throughout the manuscript.

p787, l19: This package must be calculating peak to trough monthly values. Do you think these results are compatible with the three monthly means stated at the end of the methods section?

Yes, the amplitude calculation in this package yields similar (smoothed) results compared to using peak to trough values directly. If computing amplitude based on the three monthly means, the amplitude will be smaller, but the main features in the spatial patterns should be similar (we have tested).As we are only concerned about the relative spatial and zonal contribution instead of the absolute amplitude, we believe our method is reasonable.

p787, l24ff: Given the state-of-the-art/quality of the ESSMs I have reservations about the utility of this result. Please clarify. For example when the models do not represent the past well, what should be our take home message for the future projections.

In the case of CO2 seasonal amplitude change, we compared model results and surface observation, in which we found the model ensembles actually represent the amplitude change in the historical period well for both Mauna Loa and Point Barrow. This and the high model consensus on relative amplitude increase of global CO2 and NBP suggest the models have some skills.

Even so, when we closely examined the mechanisms, the current representation in CMIP5 models are probably not realistic. We have extensively discussed this interesting issue in our new discussion section, and advised extra caution when treating the results, which only holds if the mechanisms and future emission scenarios are correct. Specifically, we cited our recently accepted paper in Nature (Zeng et al., 2014) on a major missing mechanism-agricultural intensification, and we presented the change of ecosystem types, which is also mentioned by Graven et al. (2013), in our new Figures 11 and 12. We hope our findings allow us to understand more clearly the model deficit, the mechanisms, and uncertainties. We hope this paves a way forward for future model development and validations, and eventually better future projections.

p788, l2ff: Not for all models! Do you mean model ensemble? This itself is not any interesting, unless you say something why some models do differently from others. If the model-to-model differences can be linked with the model settings or parameterization, this paper will be of great significance. At least some attempts should be made.

Yes we meant model ensemble, and we have reflect that in our revised text. Thank you for your suggestion, we have now performed a lot of new analyses and showed spatial maps for each individual models (Figure S2, S3, S6-9) in our revised

manuscript. We have also use the results of three models of sensitivity experiments to show model-to-model mechanisms, two of them included as new figures, and one attached here. In addition, we discussed on individual models and cited findings from previous literature (Anav et al., 2013; Arora et al., 2013) throughout the text. We hope the additional material in our revised manuscript is interesting to you and potential readers.

Section 3.2: Suggest deletion of Fig. 3. Looks nice but unnecessarily complicated. Since figure 4 essentially has the same information I recommend deleting. Some the text from this subsection can be moved to the next subsection.

We agree that Fig. 3 is unnecessarily complex. In our revised manuscript, we deleted Fig. 3 and merged the essential contents into Section 3.1.

p790, l19: What if you excluded MRI-ESM1 or MIROC-ESM model? This is one of example where statistics are more often misleading than convincing.

Indeed, with only ten models, the statistics should be treated with extra caution. In our revised manuscript, we added the following statement in section 3.5:

"Note that this result is based on the 10 models we analyzed; it is subject to large uncertainty and may change substantially with inclusion or exclusion of certain model(s)."

p790, l23ff: Without doing further analysis/discussions of the drivers of these changes in models, I doubt that this manuscript is a useful contribution. At least bring some discussions about the temperature, rainfall, and land cover change between 1961 and 2080.

Following your suggestion, we have included a new section (3.4) in our Results, analyzed GFDL-ESM2M, IPSL-CM5A-LR and MIROC-ESM results from sensitivity experiments. We have also included discussions on climate and land cover change between historical and future periods (Figure S6-S10, and related text in Discussions). We also cited Brovkin et al. (2013) for land cover change.

p791, l5: This "possibly" can be avoided if you just bring in temp, pcp and luc for analysis

We have now presented the soil temperature, soil moisture and crop area change between historical and future periods in our analysis. We deleted this sentence in our new conclusions section, because it is not part of our major conclusions. The interannual variability in the time series is another interesting issue, but it is not the focus of our study.

p791, l25: May be add "crop cycle" too here

As we discussed in our new discussion section, for CMIP5 models, land cover change is unlikely to increase the seasonal amplitude of CO2. Simulating the effect of agricultural intensification on crop cycle is an aspect that we would like to see improved in future versions of the models.

p792, l14ff: Is this para needed?

Thank for your comment. We have deleted this unnecessary paragraph.

Figure 1: should the units for panel b) be ppm/mo, unless you converted ppm to CO2 burden. Please check

We have indeed converted ppm to CO2 burden, as we originally described in method section. We have now also added relevant description in Figure 1's legend to prevent confusing the readers (Thank you!).

Reviewer #2:

This paper examines emission-driven simulations of terrestrial and atmospheric CMIP5 models to assess future trends in seasonal amplitude of atmospheric CO2. The topic is interesting and timely but the paper's contribution to understanding terrestrial biosphere processes and modeled responses to environmental change is marginal, mainly be- cause the description of results does not penetrate deeply into the findings and largely treats the models in aggregate.

Thank you for your helpful general statement. In our revised manuscript, we have done a number of new analyses. As a result, we largely expanded the original result section, and added a new discussions section. We hope you find the revised paper improving the aspects lacking in the original manuscript.

Aggregate patterns may be dominated by only a few models, particularly since the seasonality varies considerably between models (up to a factor of 3), and aggregate patterns may be different from individual model responses. It would be useful therefore to describe in more detail the results for individual models, also looking at the factorial experiments for those models where it is available.

In addition to presenting the ensemble model results, we now show curves (new Figure 3 and 5) and spatial plots (Supplemental figures) of individual models, and discussed the model behaviors at numerous places. We find that for the CO2 and NBP amplitude increase, our original conclusions based on model ensembles still hold for most of the individual models. Even though the magnitude of seasonal cycle can vary by a factor of 3, the phase is similar for most of the models. Following your suggestion, we also analyzed results from two sensitivity experiments for the GFDL and IPSL models, and presented in the new section 3.4 discussing the mechanisms for NBP amplitude increase. Hope you find the new results informative.

The statements on P792 about the inability to draw meaningful conclusions from 4 models and the inability to separate carbon and climate effects from factorial experiments are false. Even though the factorial experiments use different emissions than RCP8.5, they are still useful for examining mechanisms in the models. Several impactful papers have made these comparisons with 4 or fewer models.

We agree with your opinion. In our case, only three of the ten modeling groups have provided NBP results for both of the factorial experiments (esmFdbk2 and esmFixClim2), and CanESM2 cannot reproduce the correct phase of NBP seasonal cycle (discussed in Anav et al. 2013). In addition to our new section 3.4 which displays GFDLand IPSL results from sensitivity experiments, here we also attach the NBP spatial patterns from MIROC-ESM (Figure R3), which has strong climate-carbon feedback (Arora et al. 2013). Comparing to the widespread response of this model in Figure S2 and S3, this figure presents no obvious increase in net carbon uptake during peak growing season for Tibetan Plateau and Russia, and no obvious increase in carbon release during dormant season for most of the Northern high latitudes (albeit under a weaker carbon emission increase condition). This result indicate climate change do play a big role in the global CO2 and NBP amplitude increase of this model. Unfortunately, we do not have the results of the model to explicitly show its response under constant CO2 and changing climate.

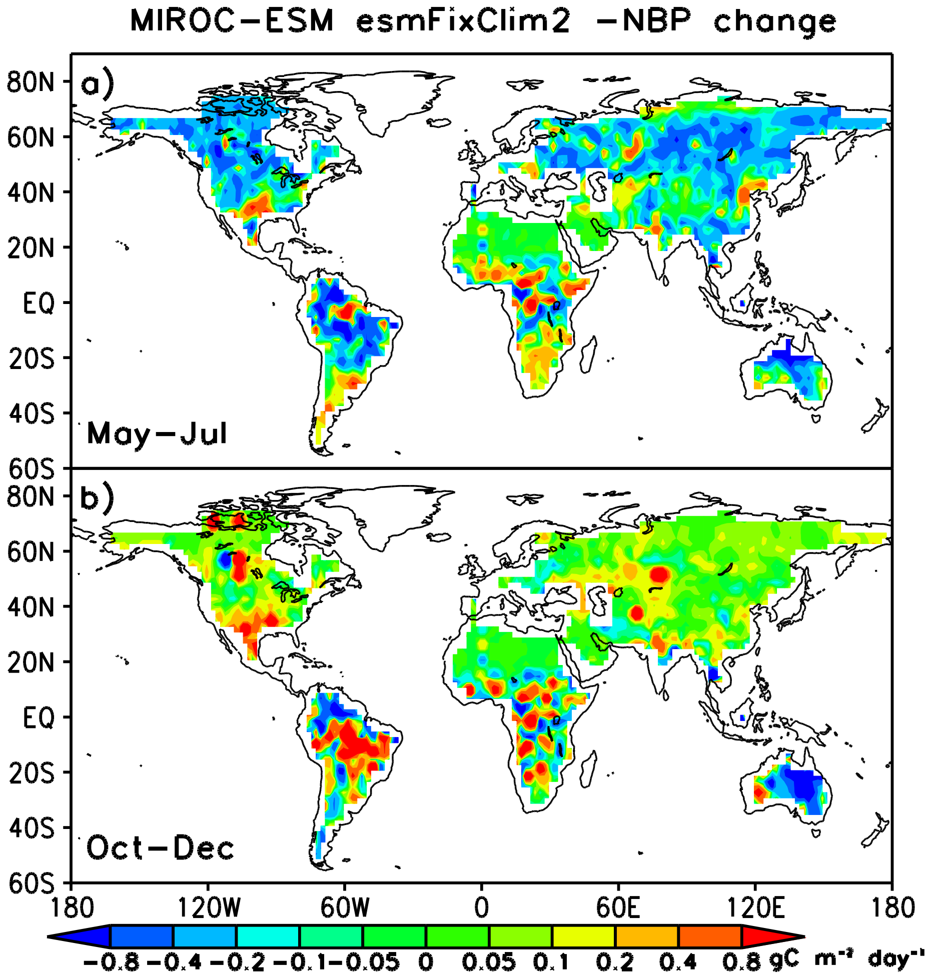

Figure R3. Spatial patterns of MIROC-ESM −NBP changes from the esmFixClim2 experiment.

The authors should look more specifically at Northern Hemisphere fluxes and near-surface CO2, where the trends in seasonal amplitude have been observed. If the global increase in CO2 amplitude is 70% to 2100, what does that mean for temperate and boreal fluxes and for Barrow CO2?

Thank you for the suggestion! We think these are very interesting questions, and we added new comparison at MLO and Pt. Barrow for model-observation CO_2. Our findings and comparison with the work by Graven et al. (2013) now comprises a big part of the new discussion section.

In addition to the global CO2 and NBP relative amplitude change presented in Figure 2, we also repeated the calculation for 25-90N, which shows a higher increase (81±46\% and 77±43\% for CO_2 and NBP, respectively).

It should be explained why the HadGEM model is omitted, particularly since it has been shown to have strong carbon-climate feedbacks, and its simulations showed the largest change in CO_2 amplitude in Graven et al 2013. Including this model would likely reduce the correlation in Figure 6 since it has a relatively small sink but relatively large change in seasonal NBP amplitude.

We are very curious about the results from the HadGEM model as well. We believe the analyses in Graven et al. (2013) are from CMIP5 Experiment 3.2, which is driven by CO_2 concentration instead of emission. For the emission-driven runs examined in our study, unfortunately both CO_2 and NBP are not available for the HadGEM models. In an earlier stage, we have contacted colleagues from the U.K. Met Office, and Dr. Liddicoat kindly explained why the HadGEM team chose not to provide NBP for ESM simulations. The diagnosed carbon flux from HadGEM landuse component is not coupled with the atmosphere module (using the land use flux from Houghton et al. 2008 instead). Therefore, even if we calculate NBP from NPP, Rh and the Land use flux ("purely diagnostic, and different from what the atmosphere sees"\textemdash personal communication with Dr. Liddicoat, Feb 6, 2014) they provided, it is not the same as net atmosphere-land carbon flux in the other models. After careful consideration, we decided it is important to keep consistency in our comparisons and omit the HadGEM model.

For Figure 6, by presenting the cross-model correlation, we were trying to show CO_2 fertilization is possibly a major mechanisms for global CO_2 and NBP amplitude increase in many of the 10 models. With only ten models, the statistics should be treated with extra caution. In our revised manuscript, we added the following statement in section 3.5:
"Note that this result is based on the 10 models we analyzed; it is subject to large uncertainty and may change substantially with inclusion or exclusion of certain model(s)."

It is interesting that the simulations appear to predict decreased seasonality in North- ern tropics and subtropics, similar to the result in Graven et al 2013, but the description of this figure is not clear and it seems that it may not account for regional or interhemi- spheric differences in NBP phasing.

Sorry for the confusion in our original Figure 5 (now Figure 4) and its description. This figure shows that the tropics contribute negatively to the global total -NBP seasonal amplitude (October maxima minus June minima) increase, which is not to say the seasonality in the tropics has decreased. This is because tropics has a different seasonal cycle phase, which is not reflected in our figure. Thanks again for your comment, we have now largely expanded the original section (now section 3.3) and added two figures (Figure 5 and Figure S1). We discussed the difference of the zonal figures, and presented results for individual models and for regional aggregates. We hope you find the revised version informative.

The comparison with previous work is lacking in general and particularly in the Discus- sion where only three references are cited. A disproportionate part of the Introduction is given to early work rather than the current state of knowledge. The writing needs to be improved overall. The text is also imprecise in several places, for example referring to the amplitude but not specifying the amplitude of what, or referring to the amplitude but meaning the amplitude change.

We have largely expanded the discussion section and included several insightful discussions from recent work including Wang et al. (2014), Anav et al. (2013), Arora et al. (2013) and Brovkin et al. (2013).

Starting from the abstract, we corrected the text in many places, trying to be precise about the information we want to convey. We hope you improved clarity in the revised version.

A major revision is needed to improve the presentation and to develop the scientific insights that can be gained from this analysis.

Many thanks for your constructive comments. In our revised manuscript, we have added six figures, a new discussion section and many other changes. Please see our final response for a list of major changes. After our work in the revision process, we are now even more confident in our original main conclusions, and are excited about our new findings. We hope you find the revised paper interesting and informative.

P781 L5 Amplitude figures are outdated

Thanks for pointing this out. We have updated the amplitude figures to 6.5 ppm (instead of 6 ppm) for Mauna Loa, and 17 ppm (instead of 15 ppm) for Point Barrow, using 2004-2013 average amplitude computed by the CCGCRV package.

P782 L5 Need reference for decreasing trend at Barrow

This sentence comes from our interpretation of Figure 1A in Graven et al. (2013). We now feel it is not necessary and removed it from the manuscript.

P782 L8 Graven 2013 gives 15% and 35%

We derived the figures from Figure 1A. We now change the description to:

"The latest analysis shows a 0.32% yr^{-1} increase in MLO amplitude and a 0.60% yr^{-1} increase in Point Barrow (Figure 1A, Graven et al., 2013). Over a 50-year period, this corresponds to an increase of 16% and 30% in MLO and Point Barrow CO_2 seasonal amplitude, respectively."

P782 L17 It is unclear what amplitude this paragraph is referring to, CO2 or NEP? Randerson et al 1997 and Gurney and Eckels 2011, at least, should be cited and described for NOAA data and CO2 inversion trends.

We have modified the text to clarify that the amplitude this paragraph refers to CO_2 amplitude. We have cited results from Randerson et al 1997 in this paragraph, and

Gurney and Eckels 2011 when we discuss relevant results in Section 3.1.

P783 L1 Add ", in comparison to Mauna Loa CO2 observations," after "for four TBMs"
P783 L8 Replace "might underestimate the amplitude" with "underestimate the CO2 amplitude change in the mid-troposphere at latitudes north of 45N".

We have modified the text as suggested, thank you!

P783 L22-25 There are grammatical errors in all of these questions.

We have corrected the questions, now it reads:

"How do CMIP5 models predict the amplitude and phase changes of CO_2 seasonal cycle in the future? Are the changes mostly driven by changes in production or respiration? Where do the models predict the largest amplitude changes will occur?"

P790 L15 Need reference

Early work by Hall 1975 (cited in our section 1) suggested the role of CO2 fertilization effect. Now with the addition of section 3.4 on mechanisms, we decide to start the section with our own findings, and modify the sentence to:

"Our analyses above suggest CO_2 fertilization effect is a major mechanism causing the amplitude increase in some models. If it is important in most models, we expect to see models with a larger change in mean carbon sink simulate a larger change in seasonal amplitude."

P792 L1 Either this work should be included in the analysis or the paragraph deleted

We have included the suggested new analyses and moved the description of the sensitivity experiments to our new section 3.4.

P792 L14 Delete paragraph.

We deleted this paragraph as suggested. In its place we added suggestions for future improvements.

Figure 2 – There are two lines of the same color

We have redesigned the color scheme for some of our plots. Specifically we have updated Figure 2 and 8 (old Figure 5) for clearer presentation.

Figure 3 - Not clear. Why are there more than two red and black lines in each panel? Why are the time periods, particularly the future period (2006-2099), so long?

Sorry for this unnecessarily complex figure. In our revised manuscript, we deleted

Fig. 3 and merged the essential contents into Section 3.1.

Figure 5 - Right panel is not explained. Does this account for different phasing of the NBP cycle in different regions?

We have reorganized and expanded our description, both in the figure legend and the main text. This figure does not account for different phasing, which is presented in our new Figure 5 and Figure S1.