We thank the reviewer for the careful assessment of our manuscript.

General comments:
This study analyses Earth System Model outputs in a creative way and pinpoints specific issues with the current generation of Earth System Models concerning the level of aggregation of model outputs and more importantly a pervasive structural issue. Koven et al. provide estimates of uncertainties arising from initial discrepancies and changes in productivity versus turnover in Earth System Model projections of the 'living' and 'dead' components of the carbon cycle using a linearisation approach. In my opinion, the study convincingly illustrates that a process, which they call 'false priming', currently occurs in all analysed models and masks intrinsic changes in turnover times for the 'dead' carbon pool.

We thank the reviewer for the overall positive assessment of the manuscript.

While I agree with their conclusions generally, I have a few general concerns:

(i) I feel that potentially underlying issues with models need to be highlighted and discussed in more detail. The result that changes in productivity are more important than changes in turnover is in my opinion an artefact of a lack of variability and realism in the representation of allocation and mortality processes in the current generation of models. The authors acknowledge this but I feel it needs further discussion.

I emphatically agree with the point the reviewer is making and believe that the manuscript more than merely acknowledges this; in fact the main conclusion of the live pool analysis is that these models are almost completely static with respect to turnover times. While there are no large-scale observations of this change, such a response is highly unlikely in reality and therefore represents a major shortcoming of these models. To me, this is the major result of the live pool analysis, and hence appears in our abstract. We can add further discussion to make more clear that we share this interpretation with the reviewer.

(ii) The assumption that woody carbon turnover is a reasonable approximation of 'life' carbon turnover for all ecosystems is questionable, especially given that they state a significant difference between forested and non-forested ecosystems.

We do not in fact make this assumption. First of all, our definition of live carbon turnover is not based on knowing wood carbon turnover; it is simply the ratio of non-respiratory output fluxes from the live C pools to the stocks of live C pools.

The point of the approximation in eqn 15 is that, since the bulk of live carbon, both as a global integral and in any given forested ecosystem (but excluding of course grassland ecosystems), is in wood (including both stems and coarse roots), we can, to first order, ignore the non-wood component and say that live carbon turnover is the product of the fractional allocation to wood and the turnover time of wood. This in principle allows us to ask, in the case where live carbon turnover is changing, whether the driver for such changes are changes in allocation to wood or due to changes in wood carbon turnover. In practice, this set of models do not in general report sufficient detail to make such a separation, so we do not do that here, and instead only qualitatively describe the known behavior of the one model that does show such behavior.

(iii) The information about specific details of scenarios and runs seems contradictory at times and could profit from some more detail. Especially, more detail is necessary concerning the CO2 feedbacks in the terrestrial and ocean biogeochemistry components, because the biogeochemically-coupled scenario is used for the calculation of the 'false priming' coefficient but does surely also include other feedbacks. I also wonder, if scenarios would have been available to determine the impact of land-use change, which I would expect to be crucial given that the terrestrial surface has been largely appropriated by humans.

We will go over the manuscript to ensure that sufficient detail has been provided. We do list here the possibility of terrestrial biophysical feedbacks, both locally-mediated and due to changes in the large-scale circulation, behind the false priming effect. There is the possibility of a small contribution to climate changes from the ocean due to ocean color changes from marine ecosystem responses to elevated CO2, which we may note. We agree that land use changes will in practice be large, and this analysis could in principle be applied to scenarios that include land-use (in CMIP5, this would be the historical+rcp4.5, esmFdbk2, and esmFixClim1 runs). We focus on the simpler runs here as a first step in understanding the model behavior and we will leave the combined land-use and atmospheric forcing analysis for future work.

(iv) I think it is worth mentioning how these models performed in benchmarking tests, i.e. evaluation against historical climate datasets.

Considerable work to benchmark these models has gone on and continues, and we do point out some relevant comparisons with data in the text. We may add more discussion of this point and point out specific model-data comparisons where relevant to our argument.

(v) The study assumes equilibrium of the carbon pools at the end of the pre-industrial period, but does not treat explicitly the cases which are actually not at equilibrium at this stage. Further discussion is needed in my opinion.

This is a good point; we had assumed the models had followed the protocol and equilibrated prior to the transient runs. In any case, the effect of non-equlibrium conditions will be a gradual equilibration response that is common in all forcings; thus comparing the response to the different forcings will nonetheless yield information about the nature of the response even in the case where non-equilibrium conditions occur.

(vi) The manuscript could be made a bit more concise by cutting out some equations (see specific comments about the equations). Finally, the figures and figure captions need tidying and some typos and grammatical errors need to be corrected. The figure axis are generally very small and it is almost impossible to read the r2 and regression equations.

We can make some graphical and grammatical edits to enhance the readability of the manuscript.

Specific comments:
P5757 : I think the title should strictly-speaking read 'Controls on terrestrial carbon feedbacks by productivity vs. turnover in five CMIP5 Earth System Models', because the authors analysed only five of the models that participated in CMIP5.

We analyzed all of the models that contributed sufficient information to the CMIP5 archive to allow us to conduct the analysis, so we feel that the existing title is accurate.

P5759 L26 : This is the most studied effect on plants and the primary effect in models, but I would not rule out that other effects such as changing water-use-efficiency, which are not captured by these models according to Keenan et al. (2013), or changes on mortality regimes could actually be more important. I feel that these issues could be at least mentioned briefly.

I think the point is that changing WUE is one result of changing photosynthesis, so the existing text is accurate. In any case our point here is simply to mention that the effects occur, so we can mention that WUE is one result.

P5758 L20 : I agree that this study gives an indication that *false priming* masks the intrinsic changes in turnover times. However, I think it needs further highlighting that the linearisation approach is only an approximation based on the available model outputs. I think there is further need to follow this up due to potential issue with the approach.

Agreed. Other approaches are certainly possible that could give a more accurate approaximation.

P5759 L27 : Although the increase in productivity is widely observed, it is questionable if it will be sustained (i.e. Norby et al., 2010 or Hungate et al., 2003). Furthermore, the transient increase in vegetation carbon could also result from changes in water-use-efficiency, allocation or turnover. Or at least these effects could be partially responsible for the observed increase.

Certainly there are many questions about whether CO2 fertilization actually occurs as strongly as models predict and what the specific effects of this behavior are, which has been covered extensively in the literature, and we already cite some of these papers here.

P5760 L11 : I would replace 'warming' with 'temperature'. Although the global, long-term trend is a warming, local and short term changes can be cooling and they also impact productivity (especially chilling and frost damage).

Agreed, the point here is not to assert any specific trend.

P5761 L21 : This sentence was not clear to me initially. Although it is correct, I suggest the authors rephrase and provide a bit more detail here.

P5762 E 2, 3 and 4 : The exact choice of the turnover calculation needs better justification and in my opinion there is no need for all three equations. I suggest the authors only justify why they use NPP rather than GPP as a flux and why they chose to lump all mortality terms together and represent them as a turnover time. To make the manuscript more concise only one equation - presumably equation (4) - would be necessary.

As discussed in the response to reviewer 1, we choose NPP rather than GPP so that we can use the productivity vs turnover dichotomy also map onto fast versus slow carbon processes. Because the bulk of autotrophic respiration passes quickly through an ecosystem, we feel that it makes sense to treat it as a net productivity that includes all of the fast processes, while the growth and mortality are slower-timescale processes and thus are more appropriately grouped together.

We will keep the equations as-is, since we feel that keeping them helps to clarify the method of the paper.

P5763 L7 : How long would it take for the carbon stock to reach equilibrium, if the productivity and turnover terms would be held constant?

This is a very good question that cannot be answered by the current experimental protocol; to know this, one would have to have some sort of pulse-response function (e.g. following Thompson and Randerson, "*Impulse response functions of terrestrial carbon cycle models: method and application*", GCB, 1999) for each of the models. I will note this in the text.

P5763 E7 : I suggest that the authors also replace the *fl->d* by *Cl / l* to make the connection between the two pools more obvious.
We want to label all fluxes as f_x and all pools as C_x for consistency.

P5764 L5 : In the methodology it is stated that land-use change was not considered as a forcing in the experiments, but here it is implied that is was not static either in some scenarios/experiments. I wonder how models implement internal variation in land-use change, when this is not prescribed as a forcing and how this internal variation affects the results. Please explain.

The text has been clarified to say that one would use harvest fluxes here if land-use were calculated, though it is not here.

P5764 L7 : Is the assumption that all fire related fluxes are generated from living pools reasonable given the models' fire modules? In reality CWD and make up a significant amount of the combustible biomass of an ecosystem.

We agree that this may be a source of error. However, the fire fluxes are small and we simply do not have the output to make a more accurate calculation, hence we choose to note the assumption explicitly and proceed.

P5764 L18 : This linearisation over such a long period (72 years) is possibly resulting in significant errors. I suggest the authors attempt to quantify the error as suggested by the other reviewer.
We will add further discussion of this point in the revised text.

P5766 L9 : Why is the smoothing done over 15 years? Would it change the results at all if the smoothing is done over 8, 9, 10 or 20 years?
It would not qualitatively change the results. We choose one number and report those results.

P5766 L20 : Again it is mentioned that land-use change and harvest were not considered at all. See my comment on P5764 L5.
That is correct; land use is not included in this experiment.

P5767 L3 : Indeed, I would expect lower decomposition rates at higher latitudes, but I would also expect lower productivity in higher latitudes. So is the change in turnover times only a result of the temperature-dependency of decomposition or also related to climatic effects on productivity. Please explain why the changes in productivity are not mentioned here.
We are not reporting changes to turnover times here, only the initial values of turnover times. Because we are describing turnover and productivity separately, we do not mention the productivity in a sentence about turnover.

P5767 L19 : The values in parenthesis are the range of r2 values, right? It is not perfectly clear to me.
We report both a range of r2 and regression coefficient (i.e. the slope a of y=ax+b) here.

P5768 L26 : Do the authors have an idea why HadGEM2-ES shows the opposing trends in turnover times?
We describe this in detail later in the paper, at this point we are first describing the behavior.

P5769 L12 : How much is the difference, when including leaves and roots? Using only the woody pool to estimate the carbon turnover, is likely to over-estimate the vegetation carbon turnover time, as foliage and fine roots have both generally shorter turnover times. While I think that the assumption is defendable for forests, I find it highly questionnable if this assumption is reasonable for non-forested ecosystems such as grasslands and savannahs, which are also investigated here. In fact, the authors even show that there is a significant difference in the turnover time of forested and non-forested ecosystems.

Note that we do not actually make this calculation in the paper, on account of there being insufficient model output to do so for most of these models. This approximation is just to make the point that turnover changes can essentially be driven by either allocation and/or mortality changes.

P5769 L15 : I assume that mortality of individual trees is either not at all incorporated due to the lack of individual representation or done very simplistically. This major caveat ought to be mentioned.

Agreed, we have added the following sentence: "For these models, both processes are highly parameterized: since none of these models include individual tree or cohort dynamics, mortality is typically treated as a constant background rate with possible disturbance-related additions, and allocation is treated either statically or as a simple functional relationship"

P5771 L18 : These statements are concerning global averages over the 72 year period, right?
No, these are not global integrals, but yes they refer to the 72-year period transient response.

P5777 L17 : It is as important to evaluate models against data and especially for turnover- related processes our understanding is limited and suitable dataset are rare. I feel this ought to be reflected upon here.
We agree.

Minor comments:
P5767 L7 : I do not think that 'likely biased low' is what the authors actually mean here. I suggest they rephrase to 'likely under-estimations'.
Changed

P5771 L26 : There is one 'seen' too much.
Changed
P5777 L10 : Missing a 'to', as in '...,it would be useful to be able...'
Changed
P5777 L10 : It should ready 'described' instead of 'describes'.
Changed
P5787 : It should read '...consider dynamic vegetation distributions...' and not '...consider dynamics vegetation dustribution...'.
Changed
P5788 : The word 'agreement' is missing an 'e'.
Changed