Dear Authors

I have read the revised version of your manuscript and your answer to reviewers' comments. Although your answers address well most comments, I still have a major concern about this manuscript.

Dear Editor,

We have revised the attached paper. In this revision we recognized that our core message was not being clearly communicated and we have made necessary edits. We reiterate that our paper is about the process a model goes through to achieve the "equilibrated state" via spin-up and the implications of this early (and usually undocumented) procedure. We demonstrate that the spin-up procedure to the "equilibrated state" causes the 6-fold range reported by Todd-Brown et al. (2013) and this in turn is caused by the parameterization of turnover rates (this paper).

To our knowledge, this has not been reported elsewhere and it is challenging to draw conclusions from models that are not consistent in the total amount of carbon they represent. We therefore provide guidelines to improve the consistency between SOC models, a required step to clarify whether soil carbon will respond to climate change as a sink or a source of atmospheric carbon.

We hope this is now clear in the revised paper and list detailed replies to your comments in the following.

I strongly agree with reviewer 1 in that apart from one or two points, there is no new major contributions in the analyses you present. A lot of what is presented here has been already discussed in Todd-Brown et al. 2013 (Biogeosciences 10: 1717), 2014 (Biogeosciences 11: 2341), Wieder et al. 2014 (Glob Biog Cycles 28: 211), and Nishina et al. 2014 (Earth Sys Dynam 5: 197). For instance, a sentence in the abstract of Todd-Brown et al. 2014 reads, 'Most of the model-to-model variation in SOC change was explained by initial SOC stocks combined with the relative changes in soil inputs and decomposition rates'. Your analysis is less supportive of the assertion that inputs are a major source of variability among models, but apart from this the analysis presented reinstates a lot of what has been said before.

We respectfully disagree with this point, although accept the responsibility of better communicating the novelty of our results. Todd-Brown et al. explain that SOC stocks contribute to model-to-model variations in SOC change. Firstly, we better acknowledge this in the revised paper ll. 50-53:

> This work is founded in the recognition that the *SOC* varies among the CMIP5 models *for the present day* over a 6-fold range (Todd-Brown et al., 2013) and this range contributes to model-to-model variations in SOC change in the future (Todd-Brown et al., 2014).

Critically, however, we explain why these differences in equilibrated conditions exist, and better introduce this aim in the revised introduction ll. 53-57:

> We explore *why* this 6-fold range exists and ultimately show that individual model responses to the spin-up procedure, particularly the dominant role of turnover time relative to SOC input, are the key reason for this range. Explaining why the amount of carbon mobilised in the active cycle varies greatly between models is critical but has been largely ignored in the literature to date.

We emphasise that this largely a technical aspect of Earth system modelling that needs to be solved to make models more consistent, and hence improve the confidence in their projections ll. 55-67:

> Explaining why the amount of carbon mobilised in the active cycle varies greatly between models is critical but has been largely ignored in the literature to date. As noted by Knutti and Sedláček (2013), there may be multiple sources of disagreement between models such as a lack of process understanding, or the reduced availability of relevant observational datasets to constrain models. Technical aspects of climate modelling, such as how different state variables are initialized or spun up to an equilibrated state prior to an experiment being conducted, and how equilibration is defined in this context, can also lead to major differences between model simulations. Discriminating between these sources of uncertainty to understand why CMIP5 models differ so significantly in the amount of *SOC* in the present day, and subsequently in the total amount of C mobilized in the global cycle under a future climate, enables an improvement in model projections. Increasing the consistency between models is required to improve our confidence in the sign of the soil carbon feedback in the future.

We reiterate this in the conclusion ll. 303-308:

> We therefore identify the spin-up procedure, and especially the response of microbial decomposition during this very long model integration, as a key source of uncertainty in the simulation of *SOC* in CMIP5 models. Critically, this involves the interaction of a technical and a process-linked uncertainty in CMIP5 models' experimental framework. The technical methods used for spin-up are model specific and not commonly reported. Interlinked with the technical uncertainty is the parameterization of processes within the spin-up period.

We also think there is a point of confusion in our paper that essentially comes down to the semantics of initialisation and have extensively edited it to resolve this. All previous papers on CMIP5 have started with a pre-industrial state in the second half of the 19th century – commonly called the "initial state" because this is the time that CMIP5 models begin to report data. However, this is not the true "initial state" – rather it is the "equilibrated state" achieved following a multi-century spin-up process from the initial state to an equilibrated state. We define "initial state", "equilibrated state" explicitly in a new paragraph to avoid confusion as this terms are often used interchangeably in the literature 69-92:

> To avoid misconceptions, we define and differentiate between two states that are commonly called "initial" states in land modelling. Our definition of "initial state", which is not known or reported in CMIP5 models, is the state at the beginning of a climate model integration. This "initial state" may come from a previous simulation, from off-line simulations, from observations or via expert judgment. In the case of *SOC*, it may be initialized as a "cold start" or in a state equilibrated with an atmosphere that reflects the period prior to the beginning of a simulation. This model state is then commonly integrated forward in time until those model states that are considered important are in equilibrium with the atmospheric model over some period of time and to a degree that is defined by the modeller (but not reported). This generates what we define as an "equilibrated state". In CMIP5, simulations are then reported from the beginning of the historical period (say

> 1850), initialized with this "equilibrated state" and integrated forward in time to the present day under observed forcings, and then into the future using a representative concentration pathway (Taylor et al., 2012). The values of a climate model's state variables at 1850 are commonly thought of as the "initial state" but they are not; it is the model-specific equilibrated state under pre-industrial forcing and this reflects the skill of the climate model to represent global and regional temperatures, rainfall and so forth. We therefore call this the "equilibrated state" and note that this differs from the "initial state" due to the Earth system model's simulated climate, the definition of "equilibrium" over time and space and *crucially* how the state variables are parameterized. Here we show that a great deal of the 6-fold range in *SOC* in the CMIP5 models at the "equilibrated state" assumed representative of 1850 (and consequently in the present day reported by Todd-Brown et al. (2013)) is a consequence of the procedures used to evolve the model from the "initial state" to the "equilibrated state". These procedures may influence how *SOC* changes through to 2100 (Todd-Brown et al., 2014) due to the current state-of-the-art representation of SOC decomposition.

We prove that the 6-fold range in equilibrated state is the result of model specific responses to spin-up from an unknown initial state. We also highlight the dominant role of turnover time over SOC input via Figure 4. We believe that explaining *why* the amount of carbon mobilised in the active cycle varies 6-fold from model-to-model is critical and worth reporting, especially since observational datasets exist that could help solve this problem. We note this point in the discussion ll. 291-294:

> However, by removing a degree of freedom associated with spin-up procedures, we believe that these observational datasets are a valuable tool for increasing the consistency between models and making them more comparable. It would improve the confidence we can have in projections of SOC fluxes and feedbacks on future climate change.

and in the conclusion ll. 316-320:

> In conclusion, we recommend that future intercomparisons should constrain model parameters so that each model achieves an equilibrated state similar to observations as the outcome of the spin-up procedure. This would remove a degree of freedom associated with the process linking initialization to equilibration via a poorly constrained spin-up procedure when comparing differences in projected changes.

Reporting equilibrated conditions has been overlooked in the papers the Editor cites, and more generally in most of ecosystem modelling studies. We then provide guidelines to make model more consistent in the amount of carbon they simulate, as removing this degree of freedom would enable a less biased comparison of the simulated effect of climate change on SOC and corresponding feedbacks. We detail this point in the discussion that has been edited ll. 261-280:

> According to our analysis of the CMIP5 models, a simple solution to reduce the uncertainty in simulated *SOC* stocks would be to modify model parameters, especially those related to *SOC* turn-over, to obtain a steady-state consistent from model to model with SOC values representative of pre-industrial conditions. Alternatively, because of the millennial time-scales of soil genesis, as well as land use changes, steady-state of global SOC stocks is not guaranteed to have existed at

> the end of the pre-industrial era. Therefore, one could choose to only consider model parameters that achieve modern stocks in accordance with observations in response to past changes (e.g. Exbrayat et al., 2014). However, this would require multiple realisations of computationally expensive models, or the use of emulators. Furthermore, it would be necessary to represent site history, and especially disturbances, with a high degree of confidence during simulations to avoid over-fitting parameters and this may not be realistic at global scale. Therefore, assuming an equilibrated pre-industrial state is a more readily available option that is supported by the lack of variations in simulated SOC during historical experiments. Thus, we suggest that one could use available estimates and confidence interval of modern SOC stocks to constrain the pre-industrial equilibrated state. These estimates include global datasets such as HWSD and other (Shangguan et al., 2014) but also regional data that may better represent high latitude stocks and permafrost (e.g. Northern Circumpolar Soil Carbon Database; Hugelius et al., 2013). Of course, while changing parameter values corresponding to SOC turnover time is relatively straightforward, it would be important to ensure that these pools are sustained by an input representative of carbon uptake.

We also add that increasing the consistency between models by removing this degree of freedom would improve the confidence we can put in ensemble projections (ll. 291-294, see above).

In summary, we take the analysis back to a major step *prior* to any earlier analysis. We fully accept this was not crystal clear and we have endeavoured to make it clearer in the revised paper.

However, I do acknowledge that analyzing model output from CMIP5 and checking possible reasons for disagreement among models could be a useful exercise that could inform model developers on the performance of their models. So, I will be open to consider a revised version of the manuscript that either 1) more clearly highlights the new findings and explore with more detail the problem of spin-up, or 2) provide elements for model developers to improve their models such as reporting the processed gridded data and/or give additional details in Table 2 that more clearly show differences in parameter values or model structures that lead to larger differences among models.

While option 2 would be interesting, it would require specialist knowledge of each model and accessing model output that is not reported in the CMIP5 archive, such as fluxes between the different ecosystem pools. We believe that Todd-Brown et al. (2013 and 2014) have provided thorough explanations for the evolution of SOC stocks in CMIP5 models and we would like to avoid being redundant.

Our paper focuses on the technical procedure that occurs before transient simulations of climate change are conducted. Therefore, we have chosen to further detail the problem of spin-up by first better defining differences between initial and equilibrated states and improving our explanations (see answer to previous comment). This is a critical technical point that should not be overlooked.

Regarding 1), I found simplistic your recommendation of improving parameterizations to achieve realistic preindustrial C stocks. I agree that in many places SOC is not in equilibrium during preindustrial times, but to represent this from a modeling point of view, it would be necessary to represent site history (successional/soil age, disturbance history, etc.) during the simulation period. An improvement in model parameters alone cannot achieve the desire result of obtaining C stocks that are not in equilibrium. However, representing site history at the global scale may be unrealistic. I think this is a topic that deserves more attention and you may take the opportunity with this manuscript to highlight this issue further.

We fully agree with this comment, particularly on the risk of over-fitting model parameters if disturbance processes are not realistically represented in the model. We have added this point to the discussion ll. 269-274:

> Furthermore, it would be necessary to represent site history, and especially disturbances, with a high degree of confidence during simulations to avoid over-fitting parameters and this may not be realistic at global scale. Therefore, assuming an equilibrated pre-industrial state is a more readily available option that is supported by the lack of variations in simulated SOC during historical experiments despite changing boundary conditions.

I do find interesting Figure 1, specially the bottom panel that shows future C stocks almost on the 1:1 line with pre-industrial C. This implies that the transient simulations, and in particular the functions fT and fW, do not modify future C stocks significantly. In the manuscript you briefly mention the effects of the different fT fW functions, but it would be very informative if you could provide additional details that help to explain why these functions did not modify C stocks in the transient simulations.

Once again, this paper focuses on the spin-up procedure, and especially on the fact that it is not adequately constrained. We simply illustrate in Figure 1 the implications of the range in equilibrated conditions by showing that it is conserved through time even under a strong climate change forcing. To explain why the functions did not modify C stocks is well beyond the scope of this paper but we do acknowledge that the dynamics behind the change in SOC simulated in each model have been investigated elsewhere ll. 223-227:

> This was not unexpected due to the slow response of SOC pools but it clearly shows that modern and future stocks are mostly defined by the equilibrated pool size while changes can be explained by a combination of changes in the input and output fluxes (see Todd-Brown et al., 2014 for a detailed account of these mechanisms).

I also found confusing your use of the term residence time throughout the manuscript. For example, in line 65 you refer to k from each pool as the baseline residence time, and on lines 78-79 the residence time is the inverse of SOC/Rh. Not only is this confusing, but also an inappropriate use of the term residence time. In the classical literature (Eriksson 1971, Ann Rev Ecol Sys 2:67, Bolin & Rodhe 1973, Tellus 25:58) the residence time is the time an atom spend in the reservoir, and only in single-pool reservoirs is the residence time equals to the ratio flux:stock. The correct term for your calculation of Rh/SOC is turnover time, and k should be called simply decomposition rate.

We have revised notations throughout the manuscript to avoid misconceptions.

In summary, the revised version of your manuscript provides only a small contribution compared to similar recent papers on this subject. I would be open to consider a revised version if you are able to better show the new contributions from this analysis and provide elements to move forward in improving SOC simulations in ESM.

We believe that we have addressed the Editor's concerns and hope that they will be satisfied by our edits.