Reply to Referee #1

Replies to Referee #1

General comments:

Exbrayat and co-authors present an important finding that initial conditions strongly control projections of soil C dynamics from the CMIP5 archive. Although well recognized in atmospheric sciences, I'm not sure similar insight is often noted in global biogeochemical dynamics. Beyond this important finding, however, the other analyses and discussion presented don't offer much new insight into refining our understanding or representation of soil C processes across scales, and display items (besides Fig. 1) are not that different from results already published by Todd-Brown and others (2013, 2014). More broadly, I'm concerned that some parts soil C community may be overly interested in constraining uncertainty in soil C projections, but not necessarily doing so for the right reasons. For example ideas discussed by Knutti & Sedláček (2013) relating to the physical climate system certainly apply to C cycle projections as well. In my estimation, discussing these considerations would improve the present manuscript.

We thank the reviewer for their suggestions. Todd-Brown et al. (2013, 2014) have indeed already highlighted the significant discrepancy in CMIP5 soil C stores. It is also true that we investigate this issue without suggesting alternative representations of soil C within these models – we *do* suggest constraining uncertainty in soil C without explicitly discussing "the right reasons" for this constraint. However, neither of these points devalues the central finding and message of this work. Here, we highlight that the lack of agreement in soil C between CMIP5 models can be explained by the response of each model to the equilibrium spin-up procedure, and further that this discrepancy is accurately predicted by each model's SOC residence time.

The effect of specifying what are apparently marginally different approaches to microbial decomposition is significant, both in terms of SOC stores (in pre industrial *and* future simulations) and fluxes. We argue here that future model projections are unlikely to be consistent as these models represent very different amount of carbon in their active cycle from the beginning of the transient climate change experiments. We finally express our concern that no model seems to be able to reproduce observational datasets of SOC pools. These are novel findings that we feel are of considerable importance to the climate science community, particularly for CMIP-6 experimental design.

We have better introduced our true aim of explaining why CMIP5 models vary so much for SOC by citing the work by Knutti and Sedláček that lists the likely reasons for model discrepancies such as the lack of process understanding or the lack of observations to constrain parameterizations ll. 44-52:

> Soil carbon pools of widely different sizes have the potential to react differently to future climate change. We therefore examine the likely reasons for the large differences between CMIP5 models in their simulation of *SOC*. As noted by Knutti and Sedláček (2013) in the context of climate models, multiple sources of disagreement between models may exist such as a lack of process understanding, or the reduced availability of relevant observational datasets to constrain models. Discriminating between these sources of uncertainty to understand why CMIP5 models differ so significantly in the amount of *SOC*, and subsequently in the total amount of C mobilized in the global cycle, would enable an improvement in model projections of the resilience of *SOC* pools and improve our confidence in the sign of the soil carbon feedback in the future.

Reply to Referee #1

Specific Comments:
I'm not sure why the authors report separate values from the same modeling centers (e.g., models E&F, G&H, I&J, K&L, N&O)? Given the similarity of results reported here (Table 3, Figs 1-4) and previous work (Todd-Brown et al. 2014), these don't really appear to independent observations. It also doesn't appear that soil biogeochemical or land models are different among these duplicated models. Thus, I would encourage the authors to consider repeating the analysis without unwarranted pseudo-replication.

We did point out that our conclusions were not affected by whether we averaged projections from the same modelling centres or not. However, as this will not change the outcome of our analyses, we agree that results would gain in clarity if we avoid pseudo-replication. We have done so in the revised manuscript and also corrected data in the Tables. This is now detailed in Section 2.2 ll. 111-114:

> As models did not start their historical simulations at the same time, we focus our analyses on the overlapping period of 1861-2100. We also averaged all simulations from the same model or institution in an attempt to account for model dependence (see Bishop and Abramowitz, 2013, for a discussion on the topic).

Results presented here are astoundingly brief with three more figures presented in the discussion. I'd consider revising the manuscript so that analyses that are not introduced until the discussion but described in methods and results.

We have revised the manuscript accordingly. All figures are introduced in the results section and the discussion has been revised according to further comments by both reviewers.

Figure 2 doesn't seem to present any valuable information, since the authors calculated SOC inputs (eq. 3), and by definition there is no change in initial SOC pools. Thus, the 1:1 relationship presented only confirms that the CMIP5 models were spun up correctly, such that SOCin = Rh.

The most obvious factor that may have explained the large range in pre-industrial SOC pools is simply that SOC pools were not in equilibrium. We investigated and discovered that SOC pools were indeed in equilibrium and use Figure 2 to support our argument. This figure has meaning to those working in this field and so we believe it is worth keeping, but we now explain it in more detail ll. 151-159:

> We next investigate the likely reasons for the existence of this pre-industrial CMIP5 range in total SOC. The first obvious step is to check whether models are at equilibrium prior to and climate change experiments. Models may not agree on total SOC simply because some of them, and especially those at the extremes of the CMIP5 spectrum, are still drifting toward their own steady-state and therefore do not comply with our experiment protocol. In Figure 2 we show the relationship between pre-industrial SOC_{in} and R_h. This relationship is highly significant ($R^2 = 1$; $p < 0.001$) and strongly suggests that all models were equilibrated under pre-industrial boundary conditions. This removes the possibility that models were not in equlibrium and means that the 6-fold CMIP5 range is likely linked with the internal terrestrial processes represented in these models.

We reinforce, however, that equilibrium was extremely unlikely to be true in the real world at the nominal time of pre-industrial simulations due to the long time required for soil genesis and as a significant proportion of the landscape had been disturbed (a point noted by Referee 2) ll. 208-215:

> Our results raise a critical problem linked to model initialization by spin-up. 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 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).

Similarly, the finding presented in Fig. 4, that initial global SOC pools are directly related to their residence time (calculated here as SOC/Rh, or the inverse of their decay rate) is also not that surprising. Moreover, this result is not markedly different from the reduced complexity model already presented by Todd-Brown and others (2013, 2014) that explains a most of the variation between CMIP5 models.

SOC dynamics are represented with first-order kinetics in all CMIP5 models. Differences between CMIP5 models only reside in how many pools they consider, and the formulation of the environmental factors controlling decomposition and residence time. We are aware of the work with reduced complexity models by Todd-Brown and others, but our manuscript addresses different ideas. They successfully used a reduced complexity model to explain the current distribution of SOC in CMIP5 models, and highlighted the controlling factors of the change in SOC in RCP8.5 projections. We found that differences in SOC pools exist at the onset of the historical experiments, and demonstrate here that it has to do with residence time. It is consistent with findings from the ISI-MIP project (Friend et al., 2014) and we have cited this work in our revisions ll. 201-205:

> Model-specific parameter k and environmental response functions f_T and f_W drive SOC pools to the size required by the residence time they simulate to compensate for SOC_{in}. This observation corroborates the predominance of residence time in the uncertainty of ecosystem response to climate change (Friend et al., 2014) and Figure 4 shows that it is independent from the number of pools considered in each model.

and observations of the relationship between the number of pools and residence time are related in the results section ll. 171-174:

> Further, residence times are not affected by the number of SOC pools represented. Models with the longest residence time have alternatively 9 (model E) or 2 pools (models H and I), while models with the shortest residence time have 8 (model A), 6 (models C and J) or 4 pools (model F).

The logic supporting the recommendation that simulating initial / or present day soil C pools may improve confidence in future projections seems tenuous at best (p. 3489, l. 23-27 & Conclusion). I agree, this would reduce variation in model projections, but it provides no constraint on the process level representation in models. Moreover, soil C pools may be significantly underestimated in the HWSD, especially at high latitudes.

We agree that constraining models to equilibrate within acceptable ranges of SOC is not the simple answer to this complex problem and other aspects of the modelling have to be improved as well. It is, however, one of the very few observational constraints we have for modelling this system and as such is critically important if model representations are to reflect real world processes.

As we show, the size of pools depends mostly upon the residence time simulated by each model. Of course, residence times could be adjusted to increase or decrease the amount of SOC required by each model to reach equilibrium. Therefore, we need to make sure that these pools are sustained by an input representative of carbon uptake. We make this point clearer in our discussion ll. 208-223:

> Our results raise a critical problem linked to model initialization by spin-up. 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 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. However, this would require multiple realisations of computationally expensive models, or the use of emulators. Therefore, as simulated SOC does not vary much during historical experiments, we suggest that one could use available estimates and confidence interval of modern SOC stocks such as those provided globally by HWSD and other (Shangguan et al., 2014), and possibly incorporating regional data that may better represent high latitude stocks and initial conditions for permafrost (e.g. Northern Circumpolar Soil Carbon Database; Hugelius et al., 2013). Of course, while changing parameter values corresponding to SOC residence time is relatively straightforward, it would be important to ensure that these pools are sustained by an input representative of carbon uptake.

The choice of using HWSD was motivated by several factors. First, it has been used in previous studies focusing on CMIP5 models (see Todd-Brown et al., 2013). Second, to our knowledge it was the only global dataset available at the time this study was conducted. We are aware of a more recent dataset by Shangguan et al. (2014), and the improved NCSCD (Hugelius et al., 2013) for high latitudes only, but the range in simulation of SOC by CMIP5 models exists regardless of which dataset we use. Of course, HWSD may under- or overestimate soil pools, and that is why we consider a confidence interval rather than the average value. As should be evident above, we have included more on the different datasets available in the revised manuscript.

Thus, following recommendations to initialize models to the HWSD dataset may omit critical permafrost C dynamics and climate feedbacks in this C rich region, but such considerations are never discussed in the manuscript. Separately, couldn't models achieve appropriate present day soil C stocks, but have wildly different environmental scalars (fT and fW, eq. 2) that would provide alternative sensitivities to environmental change in future scenarios (also see Friend et al. 2014).

Reply to Referee #1

Could variation in soil C inputs drive divergent projections in soil C storage- especially in future scenarios? These parameters can also be constrained w/ observations (e.g. Rayner et al. 2011), but a thoughtful discussion along these lines is absent from the current manuscript.

We agree that ultimately we need to ensure that models achieve the appropriate pool sizes for the right reasons, both regionally and globally. The first step in this process – essentially what we are advocating – is to make sure that observational data sources are actually used at all. We are confident that we are in agreement that observational datasets are valuable and developers should work towards better using this information when parameterizing their models. We have modified Section 4 to make this point clearer ll. 228-234:

> As decomposition processes are represented following first-order kinetics, simulating more realistic initial *SOC* stocks in response to adequate uptake fluxes would likely lead models to represent more correct modern stocks. Nevertheless, as each model relies on its own formulation of the response functions f_T and f_W, the ensemble would still exhibit different sensitivities of SOC stocks to climate change. However, by removing a degree of freedom associated with initial conditions, we believe that these observational datasets are a valuable tool for improving the confidence we can have in projections of SOC fluxes and feedbacks on future climate change.

Technical corrections:

Was N active in the BCC-CSM1.1 simulations used here? (p. 3485, l. 12)

We have enquired with Dr. Zhang from the Beijing Climate Center about the model version used in these experiments and the N cycle was not active in the current simulations. We have corrected table and figures accordingly.

This sentence seems awkward "We also averaged all realizations of the same model to retain one estimate per structure and account for model dependence". (p. 3486, l. 2-4) What are the structures referring to? Were ensembles from the same model (Table 1) averaged to give a single value for each model (Table 3, Figs 1-4)?

We initially averaged all realizations of a same model. However, following the reviewer's comments we have further averaged models from the same centre ll. 111-114:

> As models did not start their historical simulations at the same time, we focus our analyses on the overlapping period of 1861-2100. We also averaged all simulations from the same model or institution in an attempt to account for model dependence (see Bishop and Abramowitz, 2013, for a discussion on the topic).

What are the "outliers" referred to on P. 3488, l. 9? If this in reference to models outside the HWSD observations in Fig 1, the logic seems confusing since the next sentence (relating to Fig. 2) indicates that models have been spun-up appropriately (Rh = SOCin).

Reply to Referee #1

“Outliers” referred to models with either the least or the most SOC. We have replaced this word where relevant ll. 152-155:

> The first obvious step is to check whether models are at equilibrium prior to and climate change experiments. Models may not agree on total SOC simply because some of them, and especially those at the extremes of the CMIP5 spectrum, are still drifting toward their own steady-state and therefore do not comply with our experiment protocol.

References

Friend, A. D., Betts, R., Cadule, P., Ciais, P., Clerk, D., Dankers, R., Falloon, P., Gerten, D., Itoh, A., Kahana, R., Keribin, R. M., Kleidon, A., Lomas, M. R., Nishina, K., Ostberg, S., Pavlick, R., Peylin, P., Rademacher, T. T., Schaphoff, S., Vuichard, N., Wiltshire, A., and Woodward, F. I.: Anticipating terrestrial ecosystem response to future climate change and increase in atmospheric CO2, P. Natl. Acad. Sci. USA, 111, 3225–3227, 2014.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5, 3-13, doi:10.5194/essd-5-3-2013, 2013.

Shangguan, W., Y. Dai, Q. Duan, B. Liu, and H. Yuan (2014), A global soil data set for earth system modeling, J. Adv. Model. Earth Syst., 6, 249–263, doi:10.1002/2013MS000293.

Reply to Referee #2

General comments

"Response of microbial decomposition to spin-up explains CMIP5 soil carbon range until 2100" could be interesting to readers in Geoscientific Model Development. This paper clarified how to differ initial global SOC stocks among ESMs in CMIP5 experiment and the initial condition considerably governed future global soil stock behaviors even under the extreme climate change scenario in ESMs. Although the results and messages of this manuscript are very simple, I think that this study can still contribute to the further improvement in ESM due to seriously lack of constrains of initial global SOC as in this study.

I agree the overall comments given by Referee 1. Additionally, there are two major questions and some individual comments on here.

I thought the key finding in this study is that the soil decomposition processes is more dominant process to determine the initial global SOC stocks of current ESMs than C input onto soil from photosynthesis production. So, I recommended this finding should be emphasized more by additional analysis. For example, instead of just comparing between two linear regression analyses (Fig.3&4), can you analyze the relative importance of these two explanatory variables to total SOC?

We appreciate this comment and initially thought it was a very good suggestion, so we explored this in detail. When we looked closely at our results we noted that the relationship between SOCin and the pools is not significant ($R^2 = 0.01$, $p = 0.766$). We therefore suggest that any comparison between the relative importances of the two variables is likely based on non-significant relationships and could be very misleading. We therefore suggest that SOCin cannot be considered as an explanatory variable for SOC. In short, the amount of SOC_{in} cannot explain the size of the initial pools.

In fact, we are not sure the actual earth system getting the equilibrium in the global SOC stock even at industrial era. In addition, SOC accumulation and soil genesis need millennial time scale in situ. So, we can also choice the non-equilibrated state for global SOC stock in simulation. It means that we can get initial states of global SOC stock to be reaching the reference global SOC stock (HWSD) in spin-up procedure before getting the equilibrium (although this method is not used for C, N, O.). If GPP are well constrained by observations, this might seem not to be too worse option. Do you have any recommendation about whether getting the equilibrium of global SOC or not in spin-up procedure?

This is a very interesting suggestion with which we fully agree. Actually, due to the very long time needed by soil pools to equilibrate, we cannot be sure that they have been able to fully equilibrate in the pre-industrial times since, for example, the last ice age.

Also, HWSD represents the current state of the soil carbon pools and we could go further by selecting only model runs that are within this range for the representative period. We have done so with a reduced complexity model (Exbrayat et al., 2014). One issue is that it would potentially require multiple realisations of the computationally expensive Earth System model. However, this problem can be partially circumnavigated by using reduced complexity or statistical models to emulate the behaviour of the more complex model.

We have added several sentences about this issue in our revised manuscript:

> 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.

Individual comments

P3488L8-9 If you have any literatures using such an explanation, please cite here.

We do not have a reference to support this statement. We just suggest that checking whether some models are not at equilibrium is the first obvious step in this investigation of the reasons why simulated SOC pools vary that much between models. This is stated in the result section ll. 151-155:

> We next investigate the likely reasons for the existence of this pre-industrial CMIP5 range in total SOC. The first obvious step is to check whether models are at equilibrium prior to and climate change experiments. Models may not agree on total SOC simply because some of them, and especially those at the extremes of the CMIP5 spectrum, are still drifting toward their own steady-state and therefore do not comply with our experiment protocol.

P3488L29- P3489L2 I don't think these statements are meaningful. During the historical periods, it is likely that all models without N cycling scheme are parameterized under N limitation conditions. Therefore, the comparison between them doesn't give any information in this context.

We are not clear what is meant by this comment from the reviewer. The logic is that models with an active N cycle should have lower pools than the others because N actively reduces NPP and SOC_{in} in these models. However, here we find that it has more to do with these models having a faster SOC turnover rate ll. 165-167

> Similarly, the small equilibrated *SOC* pool size of models C and J seems unrelated to SOC_{in} despite these models including N limitations on plant productivity and SOC_{in}. In short, the amount of SOC_{in} cannot explain the size of the initial pools.

P3488L26- P3489L8 You should mention the differences in the variation between SOCin and decay constant among ESMs. Especially in SOCin, are there any comparable values in previous literatures?

Decay constants in ESMs may not be directly comparable because they simulate different climate that drive different land surface schemes to simulate soil moisture and soil temperature. However, we agree that SOC_{in} can be compared to literatures values of carbon uptake for example ll. 221-227:

> Of course, while changing parameter values corresponding to SOC residence time is relatively straightforward, it would be important to ensure that these pools are sustained by an input representative of carbon uptake. At equilibrium SOC_{in} equals net primary productivity (NPP) because plant pools do not vary in size. Here all models predict SOC_{in} within two standard deviations of the uncertainty range of modern, high confidence, NPP estimates (56.4 ± 8–9 Pg C yr^{-1}; Ito et al., 2011). Although not directly comparable with pre-industrial values, this global

> estimate indicates that models simulate acceptable values of global carbon uptake.

P3489L9-18 Are there any relationships between SOC residence time and (mean?) decay constant “k” in each ESM or between SOC residence time and the number of components in each ESM? This is the important information how to adjust decay constant of ESMs?

Our results do not suggest SOC residence time is affected by the number of pools represented in the model. This is, of course, a useful question and we have therefore added this piece of information in the revised manuscript ll.171-174:

> Further, residence times are not affected by the number of SOC pools represented. Models with the longest residence time have alternatively 9 (model E) or 2 pools (models H and I), while models with the shortest residence time have 8 (model A), 6 (models C and J) or 4 pools (model F).

Fig. 1&2 Please re-size the aspect ratio to be 1:1 (X axis: Y axis) of all figures.

Fig. 3&4 Please line up these two figures.

All figures There are too large significant digits in regression results.

Thank-you for these comments. We have taken the advice and improved the figures as suggested.

References

Exbrayat, J.-F., Pitman, A. J., and Abramowitz, G.: Disentangling residence time and temperature sensitivity of microbial decomposition in a global soil carbon model, Biogeosciences Discuss., 11, 4995-5021, doi:10.5194/bgd-11-4995-2014, 2014.

Ito, A.: A historical meta-analysis of global terrestrial net primary productivity: are estimates converging? Glob. Change Biol., 17, 3161–3175, doi:10.1111/j.1365-2486.2011.02450.x, 2011.