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 thoughtful suggestions. The aim of this paper is to clearly flag that the lack of agreement between CMIP5 models of SOC, already highlighted by Todd-Brown et al., can be explained by the specific response of each of these to the spin-up procedure. We argue here that 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. This is very new in climate science and is not something the modelling community focussed on soil carbon would know. It is important that it is reported in a legitimate journal since it has profound implications to CMIP-6 model design in our view.

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 justified why we did not average projections from the same modelling centres p 3486 ll.2-8:

> We also averaged all realizations of the same model to retain one estimate per structure and account for model dependence (Bishop and Abramowitz, 2013) that may bias relationships presented hereafter toward the most represented model. We examined whether averaging models from the same institution led to different results but our conclusions were not affected by this choice. Therefore, we decided to treat these models as independent as they were labelled differently by their respective developers.

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 will do so in the revised manuscript.

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 will revise the manuscript accordingly.

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 for the large range in pre-industrial SOC pools is simply that perhaps models were not equilibrated. We investigated and discovered that it was not the case by using Figure 2 to support our argumentation. This figure has significant meaning to those working on this field we believe so we would like to keep it, but we will explain in more in a revised manuscript.

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 notions. 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 will cite this work in our revisions.

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.

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 will make this point clearer in our discussion.

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. We will develop a bit 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). 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 with this recommendation and we definitely need to ensure that models achieve the good pools for the good reasons, both regionally and globally. The key message, that may have been hidden here, is that observational datasets are valuable and developers should work towards better using this information when parameterizing their models.

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 will correct the 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 will further average models from the same centre.

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).

"Outliers" referred to models with either the least or the most SOC. We will rephrase this sentence to reflect this notion.

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.