**Responses to Comments from Reviewers**

(Manuscript number: NCOMMS-17-20128-T)

**Dear Editor**:

Thank you very much for giving us the opportunity to revise this manuscript by addressing all the reviewers’ concerns. We also appreciate the two reviewers’ insightful comments and suggestions. Below, we address the reviewers’ comments and questions point-by-point. We have made changes in the text accordingly and highlighted them in red. The major points are listed as follows:

**Broad readership**: Dr. Wieder pointed out that our manuscript “*casts too narrowly for readers of the journal, and could benefit by taking a step back to articulate a broader narrative that will give the work larger impact (and readership)*”. We agree with Dr. Wieder and think this suggestion is also applicable for Nature Communications. We have taken a step back by starting the manuscript with a broad narrative, the climate-carbon feedback, instead of directly diving into uncertainties in soil carbon modeling. Please see lines 38-47. Furthermore, we also discussed the results with a broad audience in mind. Please see lines 197-206. We believe that the larger picture described in the revised manuscript is adequate for attracting broader audience. Please see our detailed response to comment 2.

**Format of the manuscript:** Reviewer #2 suggested “*the current study deserves to be reported in a format longer than a letter.*” Dr. Wieder was also concerned with many jargons in the results and readers may have to go to supplementary information for these jargons, “*that’s impossible to understanding without looking at SI tables, SI figure 1, and having a good knowledge of these models to begin with*.”. Fortunately, Nature Communications publish longer format, article. Therefore, we moved all the methods section in the supplementary information into the main methods. We furthermore moved the model diagrams and the three tables in the supplementary information to the main text and added explanation in the main text along with the model parameters. Please see the method section, Table 1, 2 and 3, and Figure 1 for relevant information.

**Model structures and environmental modifiers:** Dr. Wieder pointed out that “*Based on Fig. S1 and Table S1, it’s not clear the authors are accurately representing the model forms of CLM4 (and CLM4.5?). For example, CLM4 has 3 litter and 4 SOM pools, but this study only looks at the turnover times, transfer coefficients for 3 soil C pools? Moreover, CLM4 doesn’t represent fluxes between fast and passive pools, I fear this is a different model structure.* ” We have clarified the confusion of the model structures in the revised manuscript. First, our so called ‘CLM 4.0’ is indeed not the soil carbon model structure in CLM 4.0. It is the soil carbon cascade embedded in the Community Land Model version 4.5 (CLM 4.5bgc) without activating the vertically-resolved parameterization. We meant to say that it is a conventional Century-type model. We, therefore changed the name from CLM 4.0 to conventional Century-type model. Second, the parameterization of soil carbon model in CLM 4.5 is CLM 4.5bgc, which has three soil carbon pools in each of the 10 soil layers. The one Dr. Wieder mentioned was CLM 4.5cn. Please see Koven et al., (2013) for detailed information on the two model structures. We apologize for this confusion. We should have been more specific in methods. We have updated the names. Please see methods section (lines 313-373) and Figure 1 in the revised manuscript.

Dr. Wieder also suggested us considering the “*uncertainties in the environmental modifiers for the first-order models*”. And “*This seems like a major factor controlling the response of first order models to environmental perturbations that’s not being considered for a subset of the models here.*” As suggested and due to that “*mimics doesn't consider moisture, N or O2 scalars*”, we explored the parameter related to temperature scalar (i.e. Q10.^ ((Tsoil-25)/10)) in the two first-order models in order to more accurately compare the uncertainties among models. Please see detailed response to comment 9.

**Assimilated datasets:** Both reviewers mentioned the possibility of using a separate, more accurate database (Northern Circumpolar Soil Carbon Database: NCSCD) for the high latitude soil instead of using the corresponding part in Harmonized World Soil Database (HWSD). We took the suggestion and replaced the soil carbon in the permafrost regions in HWSD with NCSCD, and re-ran our programs. As expected, the NCSCD indeed changed the posterior distribution of key parameters such as decay rates of slow and passive soil carbon. Both values decreased in order to better match the large soil carbon in the NCSCD database. Please compare Figure S4 and S5 in the former version with Figure S4, S5 in the revised version. As a result, the models now under-estimate soil carbon in the high latitude instead of over-estimating soil carbon in the previous results. Using the new dataset also increased Vs (regression coefficient for calculating maximum reaction rate) and decreased Ksc (regression coefficient for calculating half saturation constant for chemically recalcitrant soil carbon) and Ksa (regression coefficient for calculating half saturation constant for available soil carbon), the key parameters in MIMICS model. However, the major finding that data constrain parameters and projections, and model structures and initial conditions amplify the predicted uncertainty, does not change. Please see Figure 2, 3 and 4 in the revised manuscript. In addition, the changes in parameters by assimilating the new dataset also prove the effectiveness of our data assimilation algorithm.

**Algorithm:** Reviewer #2 made a great suggestion about testing effectiveness of our algorithm. “*I am also unsure whether 50,000 accepted parameter sets are enough to reach convergence considering the lack of data constraints. An assessment of the MCMC's convergence is missing to defend this apparently low number of accepted parameter sets.*”According to this comment, we have tested the MCMC convergence using Gelman-Rubin (G-R) diagnostic method (Gelman and Rubin, 1992) in the revised manuscript. Description of this method is in lines 537-548. The G-R results for each model parameter have been added to the last column of the three tables (Table 1, 2 and 3). Based on the G-R statistics, the MCMC reaches convergence for all the model parameters. The G-R algorithm and all the parameter samples have been uploaded to my GitHub (need a link).

**Data availability**: Dr. Wieder had an excellent suggestion on making our data public. We completely agree with him. It is actually our tradition to make data and algorithms available to public after publications (Shi et al., 2015, 2016a, 2016b, Luo et al., 2017). We usually archive small data files to the public repository, such as dryad, and upload large data files and programs (in Gb, even Tb) to our own server (http://ecolab.ou.edu/download/). Unfortunately, we are in a middle of moving from our current institute to another one. Our lab webpage is temporarily unavailable. That’s why we did not provide data availability statement in the original version. Now I have archived all the algorithms and parameter values in the GitHub (link needed) and will upload all the relevant data to our server once we set it up in this new institute. Or we are open to suggestion of archive large data files (> 1 Tb).

Thank you very much for considering our revised manuscript. We hope you will find our revision satisfactory.

Sincerely,

Zheng Shi

Department of Microbiology and Plant Biology

University of Oklahoma, OK, USA

Tel: +1 (405) 325-6519

Email: zheng\_shi\_ecology@outlook.com

**Response to Reviewer #1**   
  
[Comment 1] Shi and co-authors present a compelling study looking at uncertainties in parameters, structures, and initial conditions between three different soil carbon models. The work is done very well, and should be highly relevant for readers of Nature Climate Change- especially given the journal’s history publishing important papers on this topic.

[**Response**] We really appreciate the positive and constructive evaluation by Dr. Wieder.

[Comment 2] As presented, however, I feel the work is cast too narrowly for readers of the journal, and could benefit by taking a step back to articulate a broader narrative that will give the work larger impact (and readership). My suggestions are intended to encourage the authors to think about how to achieve this goal, without being overly prescriptive.

[**Response**] According to this great comment, we have tried to attract a broad readership in the revised manuscript in two ways. First, we started the manuscript with a broad narrative in carbon-climate feedback instead of uncertainties in soil carbon projection (Lines 38-47).

“Human activities such as fossil fuel combustion and land use change, are the dominant drivers of the fast increase in atmospheric CO2 concentration. The increase in atmospheric CO2 concentration has been altering the climate system through additional radiative forcing. Carbon-climate feedback is a major mechanism for regulating climate change. For example, terrestrial ecosystems can uptake about 1/3 fossil fuel CO2 emissions and thus have the potential to slow down climate warming. Soils contain the largest carbon (C) stock in terrestrial ecosystems, twice as large as the content of the atmospheric C pool. Therefore, a slight loss in the soil C stock to climate change may cause substantially positive feedback to the atmospheric CO2, which could further warm the climate system. It is therefore essential to determine the sign and strength of such soil C-climate feedback.”

Second, we devoted three paragraphs to the discussion of the signs and strength of the carbon-climate feedback (Lines 196-232).

“Substantial uncertainties in soil C-climate feedback were predicted as a result of different model structures, parameter values and initial conditions. The conventional Century-type model predicted consistently positive soil C-climate feedback with small uncertainty, which indicates effectiveness of data-driven projections. The consistently positive feedback suggests that the model structure determines the trajectory of soil C response to climate change for this family of models. These projections are consistent with Hararuk et al., who showed decreasing soil C under RCP 8.5 in a similar model to the one employed here.

We expected a similar, positive soil C-climate feedback in the vertically-resolved model to that in the conventional model. Even though the vertically-resolved model is parameterized with mixing of soil C across soil layers due to bioturbation, cryoturbation, and diffusion, the soil C decomposition in each layer has the same representation as the conventional model. Inclusion of the vertical dimension may not alter the fundamental behaviors of the model in terms of both steady state estimation and long-term projection, as the soil C dynamics are still jointly determined by soil C influx and decay rate. In contrast to this hypothesis, adding soil layers to the conventional model allowed for both positive and negative feedbacks, due possibly to the smaller equilibrium soil C predicted by the model. The finding reveals that adding vertical resolution to the C decomposition parameterization can generate diverse responses of soil C to climate change.

The microbial model predicted a much wider range of future soil C change and diverse trajectories with both negative and positive feedbacks, indicating that projections from MIMICS are extremely sensitive to model parameters. This was somewhat expected due to the nonlinearity of the C uptake processes by microbes. The large spread in projections of the microbial model also suggests that reducing projection uncertainty requires more observations than are available at the present time in order to better constrain model parameters. Additional datasets are needed to tease apart multiple processes and further reduce the uncertainty. These results caution against inference from more sophisticated models in the presence of limited data constraints. On the other hand, MIMICS generated the best spatial comparison to soil C observations, which demonstrates model flexibility and encourages further exploration of explicit microbial dynamics in soil C model parameterization. Our results imply that the scientific community should include microbial models in future ensemble model predictions and comparisons to increase projection confidence.”  
  
[Comment 3] I also have more particular concerns about data availability and the details of the model structures and parameter estimation being considered that are outlined below, but the biggest revision should be to cast a compelling story that will be interesting to a broader readership.

[**Response**] Please see our specific responses to each comments below.

1) Presentation and readability   
[Comment 4] Several of the sentences do not read clearly (e.g. lines 22, 24, 28, 45, 69 to name a few), and should be edited for grammar and fluency on revision. This is obviously correctable, but distracting and surprising coming from the authors.

[**Response**] We have carefully check the grammar and fluency in this revised version.  
  
[Comment 5] There is a good deal of abbreviated jargon in the main text of the paper (e.g. paragraphs starting on line 92 & 201) that’s impossible to understanding without looking at SI tables, SI figure 1, and having a good knowledge of these models to begin with. As presented I fear the results will be unintelligible to the more general audience the journal is trying to reach.   
[**Response**] Thank you for the thoughtful comments. Explanations to these abbreviated jargons have been added in the text. Please see lines 101-123. Furthermore, we have moved the methods in the supplementary information, the model diagram (Figure 1 in the revised manuscript) and three tables explaining the model parameters to the main text (Table 1, 2 and 3 in the revised manuscript).

[Comment 6] Display items are not terribly compelling. In particular, Figs. 1 & 4 are maybe interesting for those interested in parameter estimation and sensitivities- especially related to soil C models, but they offer relatively little of interest for the more general readership the journal is trying to attract.

[**Response**] We have used a violin plot instead of box plot in Figure 2 in the revised version. The violin plot is able to show both parameter probability distribution and statistics such as mean or median. We think that this figure contains important information in terms of parameter values and distributions which audience may be interested in. In terms of Figure 4 (now Figure 5 in the revised version), we now used principal component analysis as suggested by the second reviewer (see comment 48) to show the relationships between model parameters and changes in projected soil carbon. Besides sensitivity of soil carbon changes to parameters, the other objective of this figure was to show the signs of such sensitivity may switch for the same parameters in different models. For example, soil carbon changes positively associated with turnover rate of slow soil carbon (k2) but negative with turnover rate of passive soil carbon (k3) in the conventional model; however, it is the other way around in the vertically-resolved model (Figure 5).   
  
2) Model structures and parameter estimation  
[Comment 7] Based on Fig. S1 and Table S1, it’s not clear the authors are accurately representing the model forms of CLM4 (and CLM4.5?). For example, CLM4 has 3 litter and 4 SOM pools, but this study only looks at the turnover times, transfer coefficients for 3 soil C pools? Moreover, CLM4 doesn’t represent fluxes between fast and passive pools, I fear this is a different model structure.

[**Response**] We apologize for the confusion. In CLM 4.5, there are two soil C decomposition modules, CLM 4.5cn and CLM 4.5bgc (Koven et al., 2013). Both are depth-resolved. As pointed out by Dr. Wieder, CLM 4.5cn has 3 litter and 4 SOM pools, and CLM 4.5bgc has 3 litter and 3 SOM pools. Here we used the parameterization for CLM 4.5bgc. Furthermore, in the CLM 4.5bgc, one can deactivate the soil C depth modules. Our so call CLM 4.0 is therefore CLM 4.5bgc without depth. We name it conventional Century-type model in the revised version to avoid confusion. Relevant information has been updated in the methods section (Lines 313-373).  
  
[Comment 8] More generally, the conventional models also simulate litter pools, but these parameters and stocks don’t appear to be considered, whereas it looks like they are for MIMICS? Does this approach misrepresent the simplicity of the microbial implicit models? 

[**Response**] We apologize for the confusion. As we focus on investigating the uncertainties by model parameters and structures of soil carbon decomposition, we did not include litter dynamics. Instead, we obtained the soil carbon input from the original model run (CLM 4.5 bgc) to drive all three models. Please see the methods section (Lines 439-444). The litter pools were not simulated in any of the three models. The arrows from Input to LITm and LITs in MIMICS may have caused the confusion. We have removed the two arrows accordingly. Please see figure 1c in the revised version. We aim to keep the input the same for all models.

[Comment 9] Similarly, why aren’t uncertainties in the environmental modifiers considered for the first-order models, whereas they are estimated for MIMICS (Vslope, Vint, Kslope, Kint)? This seems like a major factor controlling the response of first order models to environmental perturbations that’s not being considered for a subset of the models here. I realize mimics doesn't consider moisture, N or O2 scalars- but should the temperature scalar for CLM4 and 4.5 at least be considered here to more accurately compare the uncertainties among models?

[**Response**] Thank you for the great comments. We did not consider the uncertainties in the environmental modifiers due to the fact mentioned by Dr. Wieder, the lack of moisture, N and depth scalar in MIMCIS. However, we agree with Dr. Wieder that temperature scalar can introduce significant amount of uncertainty into model projections. We therefore introduced the Q10 function (Q10.^ ((Tsoil-25)/10)) into the two non-microbial models in the revised manuscript. It turned out that the global soil C data had great information for the Q10 parameter (Figure 2, Figure S4 and S5 in the revised manuscript). The estimation of Q10’s are smaller than the default value (2), but comparable to previous research. “The means of Q10’s are 1.25 and 1.06 in the conventional and vertically-resolved model, respectively, which are less than the default value (2), but close to empirical values.”

[Comment 10] Carbon use efficiency (or MGE) is a huge nob in these models, but it’s not considered at all here. I can understand why- but the decision deserves justification in the text.   
[**Response**] Thank you for the great comment. The main reason for not considering MGE in the study is that we use microbial biomass database (Xu et al., 2013) as the other constraint to the MIMICS model and therefore, MGE is not involved in the calculation of soil carbon. We have added justification for not treating MGE as a parameter in the study. Please see lines 406-409.

“Carbon use efficiency or microbial growth efficience (MGE) is a key parameter in microbial models23, 25. However, we did not consider it as a parameter in our study due to that we used microbial biomass data39 as an input to the MIMICS model. As a result, MGE is not involved in calculating soil C in MIMICS. ”

[Comment 11] Lastly, I typically sum litter and soil C stocks when calculating global averages from these models (especially for MIMICS). The decision presents challenges when comparing to HWSD observations, as ‘litter’ stocks are ambiguous (or not included)?

[**Response**] As we mentioned in the response to the comment 8, litter carbon and soil carbon were separated in our calculation. We did not simulate litter carbon, but took carbon input from litter to soil directly from original model run. Relevant information has been added in the revised manuscript (Lines xxx-xxx).

3) Data availability  
[Comment 12] This work benefits from the open access of information provided by a wealth of scientists including model code and data (globally gridded inputs, stocks, and forcings over the 20th and 21st centuries). The authors appropriately credit this work, but make no effort to make the new code, simulations, or results that went into this study available for others to use. I’m unsure on the journal’s expectations on this topic, but regardless it strikes me as poor practice not to further contribute to the scientific community.   
At a minimum, reporting the mean parameter values (and their SD) for all of the parameters used to generate Fig. 2 would be valuable in supplementary tables. I realize this information can be gleaned from SI Figs 4-6, but it’s not presented very conveniently for people who actually want to use of the insight generated by this study.

[**Response**] Thank you for the great suggestion. We have a tradition of publishing all our codes and data after our paper is accepted for publication (Shi et al., 2015, 2016a, 2016b; Luo et al., 2017). We usually archive the small data files to the public repository, such as dryad, and upload the large data files and programs (in Gb, even Tb) to our own server (http://ecolab.ou.edu/download). Unfortunately, we are in a middle of moving from our current institute to another one. Our lab webpage is temporarily unavailable. That’s why we did not provide data availability statement in the previous version. Now I have archived all the data assimilation algorithms and parameter values in the GitHub (link needed) and will upload all the relevant datasets to our server once we set it up in this new institute.

4) Minor points and technical concerns  
[Comment 13] The first paragraph reads more like an abstract for the text, whereas it seems like the journal has other expectations, I worry the strong soils focus in this text (while appropriate for the work presented) may not be broad enough for readers of Nature Climate Change. This can be changed on revision.  
[**Response**] we have updated the abstract according to the requirement of Nature Communications.

[Comment 14] In my mind the idea of confidence and diverse model structures (line 38) borrows heavily from Bradford et al. 2016, which should be cited here (as in line 64).

[**Response**] Exactly. The reference has been cited.

[Comment 15] Line 28, 74, & 96 It’s not clear what plausible ‘physical and biological boundaries’ are, or how their plausibility was determined in this study?

[**Response**] We apologize for the confusion. We simply mean that the ranges of the model parameters are within published ranges. We have changed them to avoid confusion.  
  
[Comment 16] The HWSD has low soil C densities at high latitudes, where many prefer to use results from the NCSCD (Koven et al. 2013; Todd-Brown et al. 2013). I wonder why this wasn’t considered in the present study? Given the emphasis on confidence and uncertainty in the paper, I wonder how much confidence we actually have in the observations the models are being tuned to match?

[**Response**] We really appreciate this great comment. Reviewer #2 also mentioned this NCSCD dataset. We therefore replaced the high latitude data in the HWSD with NCSCD data, and ran all the data assimilation algorithms again. We indeed see interesting differences. First, turnover rates of slow and passive soil carbon decrease by assimilating the new dataset in the two non-microbial models. Please compare Figure S4 and S5 in the former version with Figure S4, S5 in the revised version. The result is expected due to that decreased turnover rates can increase the steady states of soil carbon, and also proves the effectiveness of our data assimilation algorithms. Using the new dataset also increased Vs (regression coefficient for calculating maximum reaction rate) and decreased Ksc (regression coefficient for calculating half saturation constant for chemically recalcitrant soil carbon) and Ksa (regression coefficient for calculating half saturation constant for available soil carbon), the key parameters in MIMICS model. Please compare Figure S6 in the former version with Figure S6 in the revised version. Please also see Figure S1, S2 and S3 for grid level or total soil C content.  
  
[Comment 18] Line 110. I wonder if both microbial functional groups are both present (in most grid cells) with particular parameter combinations. This could serve as an independent constraint on the plausibility of particular parameter combinations (as the models assumes both functional groups should be present in all soils)?

[**Response**] Both functional groups are present in all grid cells. We used a global parameter fr (proportion of r-selection microbial biomass) to calculate respective biomass carbon in each functional group. We only used this one global parameter due to lack of knowledge of spatial relationships between functional fraction and climate, vegetation and/or edaphic properties. We discussed this as a potential limitation of this study. Please see lines 280-287.

“Another possible limitation relates to the two functional microbial groups in MIMICS. We did not estimate proportion of the microbial functional groups in each grid cell. Instead, we applied a global parameter (fr: proportion of r-selection microbial biomass) to calculate the r-selection microbial biomass, and the remainder is k-selection microbial biomass. We made this assumption due to the uncertainty in the spatial pattern of the relationship between the two microbial functional groups and climate, edaphic properties and/or vegetation characteristics. Future research may focus on developing relationships between this parameter and climate, soil and vegetation.”  
  
[Comment 19] It looks like the authors use observationally derived microbial biomass data as a constraint for MIMICS, although it’s not really clear how this was done? I also wonder how much faith one should have in these data, especially when extrapolated to global scales and a depth of 1 m!? Besides greater detail on how these were used, some discussion (even if only in the SI) would about the limitations of this secondary constraint seems warranted?

[**Response**] The microbial biomass C was used as the steady state in the MIMICS model due to the fast turnover of microbes, typically less than one year. A global parameter, fr (fraction of r-selection microbial biomass) was multiplied by the microbial biomass database to calculate the r-selection microbial biomass and the remainder is the k-selection microbial biomass. The two functional group biomass was directly used in the equations (Lines 411-432) to calculate steady states of SOMa, SOMp and SOMc in the MIMICS model. Relevant information about how microbial biomass was used has been added in the methods (Lines 466-469). The limitation of the observed microbial biomass was also discussed. Please see lines 270-278.

“We caution the reader on several assumptions in this study when interpreting our results. First, we assumed the observed soil C dataset is at steady state due to its heterogeneity in time; second, we also assume the steady state in observed microbial biomass data used in the microbial model (MIMICS) due to the fast turnover rate of microbial processes. The steady state assumption is convenient when the datasets are highly heterogeneous in time and space, but may introduce uncertainty in projection. Furthermore, these static estimates of soil C and microbial biomass C provide limited insights into the fate of soil C pools and potential microbial activity under climate change. However, it is a common practice so far to make best use of these datasets”

[Comment 20] Line 120 & Fig S7. Where are the observations from in this figure, Wieder et al. (2014a,b) report ~ 1260 PgC globally from the HWSD that’s reportedly being used here?  
[**Response**] We calculated 1430 PgC globally from the regridded database (CLM resolution: 1.25 × 0.94°), not from the original HWSD resolution (30 arc sec). Our estimation is close to Tian et al., (2015) estimation (1400 PgC with resolution 0.5 × 0.5°). Our new estimation is 2210 PgC in the HWSD with NCSCD replacing the permafrost soil (Fig. S7 in the revised version).

[Comment 21] I’m surprised that both versions of CLM show carbon losses under RCP 8.5 with these optimized parameters, as previously first order models show soil C accumulation (Todd Brown et al. 2014). This is in line with some of the author’s previous findings, but it seems worth discussing this change in sign following calibration.

[**Response**] We devoted a whole section (three paragraphs) to discussing the sign and strength of the soil C-climate feedback. Note that CLM 4.5 now predicted either carbon gain or loss depending on parameter values. Please see lines 196-232 for details.

“Substantial uncertainties in soil C-climate feedback were predicted as a result of different model structures, parameter values and initial conditions. The conventional Century-type model predicted consistently positive soil C-climate feedback with small uncertainty, which indicates effectiveness of data-driven projections. The consistently positive feedback suggests that the model structure determines the trajectory of soil C response to climate change for this family of models. These projections are consistent with Hararuk et al., who showed decreasing soil C under RCP 8.5 in a similar model to the one employed here.

We expected a similar, positive soil C-climate feedback in the vertically-resolved model to that in the conventional model. Even though the vertically-resolved model is parameterized with mixing of soil C across soil layers due to bioturbation, cryoturbation, and diffusion, the soil C decomposition in each layer has the same representation as the conventional model. Inclusion of the vertical dimension may not alter the fundamental behaviors of the model in terms of both steady state estimation and long-term projection, as the soil C dynamics are still jointly determined by soil C influx and decay rate. In contrast to this hypothesis, adding soil layers to the conventional model allowed for both positive and negative feedbacks, due possibly to the smaller equilibrium soil C predicted by the model. The finding reveals that adding vertical resolution to the C decomposition parameterization can generate diverse responses of soil C to climate change.

The microbial model predicted a much wider range of future soil C change and diverse trajectories with both negative and positive feedbacks, indicating that projections from MIMICS are extremely sensitive to model parameters. This was somewhat expected due to the nonlinearity of the C uptake processes by microbes. The large spread in projections of the microbial model also suggests that reducing projection uncertainty requires more observations than are available at the present time in order to better constrain model parameters. Additional datasets are needed to tease apart multiple processes and further reduce the uncertainty. These results caution against inference from more sophisticated models in the presence of limited data constraints. On the other hand, MIMICS generated the best spatial comparison to soil C observations, which demonstrates model flexibility and encourages further exploration of explicit microbial dynamics in soil C model parameterization. Our results imply that the scientific community should include microbial models in future ensemble model predictions and comparisons to increase projection confidence.”  
  
[Comment 22] Lines 150-160. CLM45 has been validated with site specific 14C profile data (Koven et al. 2013) and other work uses litter decomposition studies to help parameterize and evaluate models (Bonan et al. 2013; Wieder et al. 2015). I realize such site-level calibrations are outside the scope of this manuscript, but they seem like independent controls than may be able to further constrain parametric uncertainties in the models and could be important to discuss, especially given the conclusion that such models require more observations (line 233), which were not used in the present analysis.

[**Response**] This point has been discussed in the revised version (Lines 234-267). We devoted a whole sections (three paragraphs) to the discussion of using additional datasets to further constrain parameters and predictions. Indeed, 14C soil profile has been used to constrain transfer coefficients and turnover rate at multiple sites (He et al., 2016), and suggests its potential on a global scale. We advocate using isotope data as complementary data source to better constrain model parameters. Meantime, we caution that additional processes (i.e., parameters) have to be integrated the existing model structures in order to assimilate 14C isotope data (Koven et al., 2013; He et al., 2016). More parameters may introduce additional uncertainties. In consistent with our implication, we point out that we need to strike a balance between model complexity and assimilated data. For more details, please refer to lines 234-267.

“The lack of constraint on the transfer coefficients in the two first-order decay models indicates that the two opposing mechanisms, transfer coefficient and turnover rate, require more informative data to be disentangled from one another. Substantial uncertainty in predicted soil C dynamics in MIMICS also suggests that additional data are necessary to further constrain parameters. In particular, our results suggest that observations related to the microbial maximum uptake rate and half saturation constant are of the most importance to reduce uncertainty.

Isotopic data in C processes show great potential to constrain these processes. Indeed, 14C soil profiles have been used to constrain transfer coefficients and turnover rates at multiple sites; isotopic labeling to trace C pathways is another powerful tool to provide additional constraints to relevant processes, such as distinguishing root respiration from total soil respiration, sources of input or proportion of different soil C pools. Other additional data, such as soil respiration and soil C incubation datasets are equally valuable constraints. We therefore advocate using isotopic data and other datasets as complementary sources to better constrain model parameters and hence projections.

However, there are limitations with these additional datasets for model-data integration, especially at global scale. First, additional processes (i.e., parameters) have to be incorporated into the existing model structures in order to leverage all the datasets. More parameters may introduce additional uncertainties and especially equifinality issue. Additionally, most of these data are collected at small spatial scales, and hence may not provide a good global parameter constraint. Linking data at the micro-scale to the intermediate and large scales presents significant challenges for primarily two reasons: 1. lack of effective upscaling scheme may introduce additional uncertainty to the data; 2. Data assimilation using global C models is difficult due to the fact that many parameters are global parameters with poorly understood regional variations due to climate, vegetation and edaphic properties. Leveraging these additional datasets as benchmarks instead of systematically assimilating them to constrain global model parameters may be the best use of these observations. Relaxing the global parameter assumption could be another option.”  
  
[Comment 23] Line 168. Why are SI Fig 10B and 11B not referenced?

[**Response**] Added in the revised version (Lines 156-168). Now they are Figure S8 and S9 in the revised version.  
  
[Comment 24] Line 177 & Fig. 3b - I'm unclear what the data-derived initial conditions are referring to, what parameters were used, or the particulars of how individual pools were initialized. Maybe this approach is common in Baysian analyses, but it seems completely foreign from the standard way C cycle models are run? Ultimately, there may be value in the approach, but a more careful explanation of the details seems warranted (in the SI or Methods).   
[**Response**] We have added more details on how to derive the initial conditions from datasets (Lines 521-529).

“We ran the three models with two different initial conditions: steady-state initial condition and data-derived initial condition. The steady state was estimated by each of the parameter sets following the equations above. The data-derived initial condition used the observational soil C as initial condition. Each individual pool size (e.g., slow soil C in the two non-microbial models and chemically recalcitrant C in the microbial model), was determined by scaling the observational data by the mean fraction of the pool. Mean soil C fractions were similarly computed for the steady-state initial condition. For example, in MIMCS,, where f\_SOMc is the fraction of the chemically recalcitrant C, SOMa is available soil C and SOMp is physically-protect C at steady states. ”

[Comment 25] Line 218. What is the negative feedback here? I’d assume that terrestrial C loss (as simulated by the two versions of CLM) would be a positive C cycle – climate feedback?

[**Response**] We apologize for the mistake. We meant positive C-climate feedback. It has been corrected in the revised version.  
  
[Comment 26] Line 544 & 594 more details on the spinup and transient simulations are needed here. Where did soil moisture and temperature data come from? What climate data were used to force the CLM4.5 runs (based on the years provided I’d assume Qian?) How do the authors move from this data atmosphere to the model atmosphere in the RCP8.5 simulations, as there are large differences in the temperature and productivity simulated by clm4.5 when switching between these forcing data it seems like a good deal of the transient result could result from inconsistencies between spinup and transient datasets?  
[**Response**] As Dr. Wieder pointed out that we used Qian (2006) bias-corrected reanalysis dataset to force the CLM 4.5 historical runs. Basically, for the model years 1850–1947, we cycle atmospheric forcing from the period 1948–1972, and use the corresponding atmospheric data for the years 1948–2004. The atmospheric data in RCP8.5 scenario forced CLM 4.5 for 2005-2100. We obtained inputs to soil carbon pools from running the scenario and then used the inputs to drive our soil carbon models.

We agree with Dr. Wieder that there is big difference in carbon fluxes between the spinup and RCP 8.5 scenario. We used this most extreme scenario to highlight the difference in soil carbon projections by different model structures. Coupling historical run with this extreme sceario is a common practice in showing the greatest difference in predictions (Hararuk et al., 2015; Wieder et al, 2013, 2015). Relevant information has been added in the methods (lines 442-444; lines 509-511).

[Comment 27] Line 604-607. I’m unclear how the data-derived initial conditions are being generated, or how model parameterizations were modified to achieve this steady state? The text here doesn't seem to apply to each of the model structures being investigated.

[**Response**] We have added more information on how the data-derived initial conditions are calculated. Basically we multiply the observational database by the ratio of each individual pool to the total soil carbon pool to derive the size of each individual pool; the ratio is computed when calculating the steady states from each set of model parameters. Please see lines 521-530 and our response to comment 24.  
  
[Comment 28] Fig 4. The text and other figures discuss CLM4, CLM45 and MIMICS. Should the figure be organized similarly?  
[**Response**] Corrected.

[Comment 29] Fig S1. What happened to the litter pools simulated by CLM4 and 4.5? More, I'm not sure the diagram for CLM4 is accurate (see Bonan et al. 2013; Oleson et al. 2013, Koven et al. 2013). By omitting these pools from the analysis, are the number of variables estimated for CLM4 and CLM4.5 artificially low compared to MIMICS?

[**Response**] Please see our responses to comment 8.  
  
[Comment 30] Fig S1. CLM4.5 does not simulate convection in its vertical representation of soil C pools, but advection and diffusion terms.

[**Response**] We changed it to ‘vertical mixing’ as used by Koven et al., 2013. Now it is Figure 1 in the revised manuscript.  
  
[Comment 31] Fig S8. For what model are these differences calculated, I’m assuming CLM45, but the caption should make this clear?  
[**Response**] We did not present a corresponding figure in the revised version due to minimal relevant information provided by the figure.

[Comment 32] Fig S9. Years should be provided on the x axis, not months.

[**Response**] We do not think it is necessary to present corresponding figure in the revised version due to overlap with Figure 4.   
  
[Comment 33] Figs S10 & 11, captions could be clarified and expanded on?

[**Response**] Updated. Now they are Figure S8 and S9 in the revised version.  
  
[Comment 34] Tables S1 & S2. Why are do t1 and t2 have the same description?

[Response] Updated. Now they are Table 1 and 2 in the revised version. The t1 is the slope to calculate f21 and t2 is the intercept.

**Response to Reviewer #2**   
  
[Comment 35] The uncertainty in the response of soil carbon processes to climate change is a key challenge to address if we want to increase the confidence we can put in ESM projections. Therefore, I believe that the approach used here is an important contribution as it shows that we are currently lacking observational datasets able to constrain more realistic models of microbial decomposition.   
[**Response**] Thank you for the positive comments.

[Comment 36] However, I feel that the current study deserves to be reported in a format longer than a letter. I have found the manuscript to be a quite technical contribution with large parts of the discussion focusing on some model-specific parameters, as shown in e.g. in Figures 1 & 4, which remains very hard to comprehend for someone not familiar with these particular models. As a result, the paper would be clearer if large parts of the extensive SI could be integrated in the main text.

[**Response**] The manuscript has been revised according to Nature Communications in the format of an article. We have moved all the methods in supplementary information to the main text. We also moved the three supplementary tables to the main text for audience’s convenience.

[Comment 37] For example, I am still not clear of the exact model-data fusion strategy that was used: was it one "global" MCMC per model (i.e. one set of global parameters being optimised)? or one MCMC per pixel per model? was the MCMC only used to make the model's steady-state match data from HWSD?

[**Response**] We apologize for the confusion. We did data assimilation with these global models in the framework of global parameters. Therefore, the parameters we optimized are all global parameters.

Yes. The MCMC was only used to match steady state with HWSD and NCSCD.

[Comment 38] I am also unsure whether 50,000 accepted parameter sets are enough to reach convergence considering the lack of data constraints. An assessment of the MCMC's convergence is missing to defend this apparently low number of accepted parameter sets.  
[**Response**] A Gelman-Rubin test in convergence has been conducted. Please see last columns of Table 1, 2, and 3 for results. They are all 1’s for all the parameters, which mean convergence. Please see lines xxx-xxx for description on Gelman-Rubin method.

“We used Gelman-Rubin (G-R) diagnostic method to determine convergence of MCMC simulations46. The idea of G-R test is that if the simulated Markov chain has reached convergence, the within-run variation within each chain should be roughly equal to the between-run variation among chains. Specifically, denoting each model parameter as ci, the parameter samples from K (K = 5) parallel M-H runs of length N (N = 10000),  the between (Bi) and within-run (Wi) variances are defined as





The G-R scale reduction statistics is given by



Once convergence is reached, *GRi*should approximately equal one.”

Below are some more specific comments that I hope will help the authors improve the manuscript.  
  
[Comment 39] l 45. a change in soil C pool dynamics

[**Response**] We changed it to ‘a slight loss in the soil C’. Please see line 44.

[Comment 40] l 46. references missing to CMIP (e.g. Eyring et al., 2016) and MsTMIP (Huntzinger et al., 2013)

[**Response**] updated. Please see line 51.

[Comment 41] l 58. "classical" parameterization

[**Response**] updated. Please see line 66.

[Comment 42] l 87. by "negative response" do you mean "decrease in soil C in response to climate change"

[**Response**] Updated. Please see line 96

[Comment 43] l 100. well well  
[Comment 44] l 116. fig 2 legend indicates 50,000 sets...

[**Response**] We randomly took 1000 parameter sets out of the 50000 sets.

[Comment 45] l 119-122. can you quantify biases with percentages or root mean squared error?

[**Response**] Thank you for the great comment. Root mean squared error has been added. Please see Figure 3b, d, and f.

[Comment 46] l 161. is there a way to tackle the lack of observational data by using some constraints

[**Response**] We devoted a whole section to discussing the advantages and challenges to use additional data constraints. Please see lines 236-269.

“The lack of constraint on the transfer coefficients in the two first-order decay models indicates that the two opposing mechanisms, transfer coefficient and turnover rate, require more informative data to be disentangled from one another. Substantial uncertainty in predicted soil C dynamics in MIMICS also suggests that additional data are necessary to further constrain parameters. In particular, our results suggest that observations related to the microbial maximum uptake rate and half saturation constant are of the most importance to reduce uncertainty.

Isotopic data in C processes show great potential to constrain these processes. Indeed, 14C soil profiles have been used to constrain transfer coefficients and turnover rates at multiple sites; isotopic labeling to trace C pathways is another powerful tool to provide additional constraints to relevant processes, such as distinguishing root respiration from total soil respiration, sources of input or proportion of different soil C pools. Other additional data, such as soil respiration and soil C incubation datasets are equally valuable constraints. We therefore advocate using isotopic data and other datasets as complementary sources to better constrain model parameters and hence projections.

However, there are limitations with these additional datasets for model-data integration, especially at global scale. First, additional processes (i.e., parameters) have to be incorporated into the existing model structures in order to leverage all the datasets. More parameters may introduce additional uncertainties and especially equifinality issue. Additionally, most of these data are collected at small spatial scales, and hence may not provide a good global parameter constraint. Linking data at the micro-scale to the intermediate and large scales presents significant challenges for primarily two reasons: 1. lack of effective upscaling scheme may introduce additional uncertainty to the data; 2. Data assimilation using global C models is difficult due to the fact that many parameters are global parameters with poorly understood regional variations due to climate, vegetation and edaphic properties. Leveraging these additional datasets as benchmarks instead of systematically assimilating them to constrain global model parameters may be the best use of these observations. Relaxing the global parameter assumption could be another option.”

[Comment 47] l 176. we know differences in initial conditions alter projections (Exbrayat et al., 2014) and that models are constantly drifting toward their steady-state under current condtions (Luo et al., 2017)... perhaps rephrase to "we want to quantify how much initial conditions..."

[**Response**] Updated according to this comment. Please see line 172

[Comment 48] l 203. did you use simple linear regressions, or did you account for parameter covariations? if so, using a PCA would be more suited with simple linear regressions

[**Response**] We used simple linear regressions. PCA was conducted as recommended. PAC not only shows the relationship between soil C change and parameters, but also illustrates correlations among model parameters themselves. Please see figure 5 in the revised version.

[Comment 49] l 201-212. this paragraph is very very specific to these models, please describe parameters to guide the reader

[**Response**] Thank you for the thoughtful comment. Explanations have been added. Please see lines 183-195.

[Comment 50] l 227-233. agreed, but can you be more specific on what kind of observations are required: stocks, fluxes?

[**Response**] We devoted a whole section to discussion of additional data needed to further constrain model parameter and then projections (lines 234-267). Please also see our responses to comments 22 and 46.

[Comment 51] Figure 1. please put parameter names on the figure

[**Response**] Updated according to the suggestion.

[Comment 52] Figure 2. would the biases at high latitudes be resolved by using NCSCD (Hugelius et al., 2013) instead of HWSD for these regions?

[**Response**] Thank you for the great comment. We therefore replaced the high latitude data in the HWSD with NCSCD data, and ran all the data assimilation algorithms again. We indeed see interesting differences. First, turnover rates of slow and passive soil carbon decrease by assimilating the new dataset in the two non-microbial models. Please compare Figure S4 and S5 in the former version with Figure S4, S5 in the revised version. The result is expected due to that decreased turnover rates can increase the steady states of soil carbon, and also proves the effectiveness of our data assimilation algorithms. Using the new dataset also increased Vs (regression coefficient for calculating maximum reaction rate) and decreased Ksc (regression coefficient for calculating half saturation constant for chemically recalcitrant soil carbon) and Ksa (regression coefficient for calculating half saturation constant for available soil carbon), the key parameters in MIMICS model. Please compare Figure S6 in the former version with Figure S6 in the revised version. Please also see Figure S1, S2 and S3 for grid level or total soil C content.  
  
[Comment 53] Figure 3. models are unlikely to match HWSD stocks, so I am not sure we learn much from Fig 3B and the corresponding text

[**Response**] The objective of Figure 3B (now Figure 4b in the revised version) was to see how changes in initial conditions alter the trajectory of soil carbon projections. We believe that this figure is important to prove this point. Even though the models show some biases, the modeled soil carbon is still comparable to the datasets (HWSD and NCSCD), especially in the vertically-resolved model and the MIMICS.

[Comment 54] Figure 4. is a well crafted summary figure that contains a lot of information. It however requires more discussion to identify and interpret why "certain" parameters are correlated with soil C stocks and soil C change.

[**Response**] According to comment 48, we changed this figure to biplot figure by performing PCA. We discuss why certain parameters are more correlated with soil C change. Please see lines 184-195.

“Besides exploring the relationships between initial conditions and model projections, we also investigated the sensitivity of projected soil C changes to the model parameters by conducting principal component analysis (see Methods). In the conventional model, changes in soil C content at the end of 21st Century were positively associated with k2 (decay rate of slow soil C), but negatively with k3 (turnover rate of passive soil C) and Q10 (temperature sensitivity of soil C turnover) (Fig. 5a). In contrast, changes in soil C were positively associated with D1 (diffusivity in non-permafrost soils) and k3, but negative with k2 in the vertically-resolved model (Fig. 5b). Soil C changes in MIMICS were positively associated with Vi (regression coefficient for calculating maximum reaction rate) and Ki (regression coefficient for calculating half saturation constant), but were not sensitive to other parameters (Fig. 5c). Consistent with previous research, turnover rates often control soil C changes in the conventional model parameterizations; uptake rates and half saturation constants are critical for controlling soil C changes in microbial models.”

**Cited literature in the response letter**