


ARTICLE

DOI: 10.1038/s41467-018-04526-9

OPEN

# Model structures amplify uncertainty in predicted soil carbon responses to climate change

Zheng Shi <sup>1,2</sup>, Sean Crowell<sup>2</sup>, Yiqi Luo<sup>3,4</sup> & Berrien Moore III<sup>2</sup>

Large model uncertainty in projected future soil carbon (C) dynamics has been well documented. However, our understanding of the sources of this uncertainty is limited. Here we quantify the uncertainties arising from model parameters, structures and their interactions, and how those uncertainties propagate through different models to projections of future soil carbon stocks. Both the vertically resolved model and the microbial explicit model project much greater uncertainties to climate change than the conventional soil C model, with both positive and negative C-climate feedbacks, whereas the conventional model consistently predicts positive soil C-climate feedback. Our findings suggest that diverse model structures are necessary to increase confidence in soil C projection. However, the larger uncertainty in the complex models also suggests that we need to strike a balance between model complexity and the need to include diverse model structures in order to forecast soil C dynamics with high confidence and low uncertainty.

<sup>1</sup>Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, 210037 Nanjing, China. <sup>2</sup>School of Meteorology, University of Oklahoma, Norman, OK 73019, USA. <sup>3</sup>Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ 86011, USA. <sup>4</sup>Department for Earth System Science, Tsinghua University, 10084 Beijing, China. Correspondence and requests for materials should be addressed to Z.S. (email: [zheng\\_shi\\_ecology@outlook.com](mailto:zheng_shi_ecology@outlook.com)) or to S.C. (email: [scrowell@ou.edu](mailto:scrowell@ou.edu))

Human activities such as fossil fuel combustion and land use change are the dominant drivers of the fast increase in atmospheric CO<sub>2</sub> concentration<sup>1,2</sup>. The increase in atmospheric CO<sub>2</sub> concentration has been altering the climate system through additional radiative forcing<sup>1</sup>. Carbon-climate feedback is a major mechanism for regulating climate change. For example, terrestrial ecosystems can uptake about 1/3 fossil-fuel CO<sub>2</sub> emissions<sup>3</sup> 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<sup>4</sup>. Therefore, a slight loss in the soil C stock to climate change may cause substantially positive feedback to the atmospheric CO<sub>2</sub>, which could further warm the climate system. It is therefore essential to determine the sign and strength of such soil C-climate feedback.

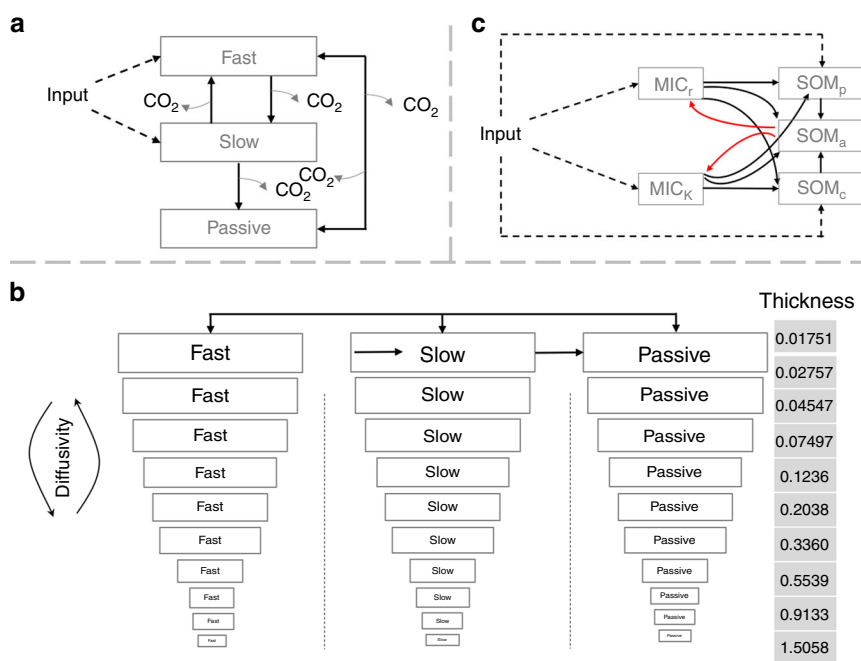
Global land C models are critical tools for quantifying the response of soil C to climate change<sup>5,6</sup>. Large uncertainty in the predictions of soil C has been well documented among these models. Several model inter-comparisons (e.g., CMIP<sup>7</sup> and MsTMIP<sup>8</sup>) have demonstrated that global land C models vary considerably in their estimates of the global soil C stock for the contemporary period<sup>5,9</sup> and for the future greenhouse gas emission scenarios<sup>6</sup>. For example, Todd-Brown et al.<sup>5,6</sup> reported large differences in estimated contemporary global soil C stocks, ranging from 510 to 3040 Pg C projected by 11 Earth system models, and in projected change over 21st century ranging from a loss of 72 Pg C to a gain of 253 Pg C under the worst-case greenhouse gas emission scenario; Tian et al.<sup>9</sup> reported substantial differences in estimated contemporary global soil C stocks ranging from 425 Pg C to 2111 Pg C by ten terrestrial biosphere models in the Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP).

The global land C models differ in model structure, parameter values, and initial conditions, each of which may contribute substantially to the overall uncertainty across models. Past studies have shown that different model structures can generate different soil C projections<sup>10–12</sup>; initial conditions positively correlate with

projected soil C content<sup>5,13</sup>; and classical parameterization causes large uncertainty in projected changes in soil C<sup>14,15</sup>. Model structure may determine the range of model projection and meantime the choice of parameter values for a given model structure defines the quantitative accuracy relative to observations. Therefore, parameterization is likely to interact with model structure and even initial conditions to impact the model projections. These considerations imply that diverse soil C decomposition models are needed to increase projection confidence<sup>16</sup>. Alternative structures to the conventional Century-type models include microbial models that simulate decomposition processes with explicit microbial traits as well as models that simulate interactions between soil layers at different depths<sup>16–19</sup>. Exploring uncertainty generated by these model structures and parameterization is critical for global land C modeling, but little effort has been dedicated to addressing it, especially at a global scale due mostly to computational cost.

In this work, we utilized a Markov Chain Monte Carlo (MCMC) technique to sample the parameter space for three different modeling frameworks in order to produce a calibrated ensemble of parameter values weighted by agreement with soil C observations. We chose three representative global soil C decomposition models with different structures (Fig. 1), among which were conventional Century-type model, a vertically resolved soil C model with explicit soil depth embedded in CLM 4.5<sup>11</sup> (CLM 4.5bgc) and a microbial model (the Microbial-Mineral Carbon Stabilization: MIMICS<sup>20</sup>). A set of projections driven by the Representative Concentration Pathway 8.5 (RCP 8.5) was carried out for all the three models with their posterior parameter ensembles to generate a distribution of predicted soil C, from which statistics such as the mean and uncertainty were estimated.

We hypothesized that uncertainty in projected soil C change by varying parameter values can be substantial: specifically CLM 4.5 and MIMICS would have larger projection uncertainty than the conventional Century-type model, especially the microbial model which is nonlinear and has a larger parameter space than the other two models; the two non-microbial models would predict



**Fig. 1** Soil carbon decomposition models with distinct structures. **a** Conventional Century-type model, **b** vertically resolved model in CLM 4.5bgc, and **c** a microbial explicit model, Microbial-Mineral Carbon Stabilization (MIMICS). MIC microbial biomass, SOM soil organic matter. Please refer to methods and Tables 1 and 2 and Supplementary Table 1 for parameterization. Unit of thickness of each soil layer is in meter

**Table 1 Descriptions of parameters in the conventional model**

Short name	Description	Unit	LL	UL	Default	Mean $\pm$ SD	G-R
$t_1$	Coefficient to calculate $f_{21}$ (intercept)	-	0	1	0.85	$0.12 \pm 0.064$	1.00
$t_2$	Coefficient to calculate $f_{21}$ (slope)	-	0	1	0.68	$0.048 \pm 0.04$	1.00
$f_{31}$	Fraction of C in fast soil C transferring to passive soil C	-	0	0.01	0.005	$0.006 \pm 0.0025$	1.00
$f_{12}$	Fraction of C in slow soil C transferring to fast soil C	-	0.1	0.6	0.4185	$0.47 \pm 0.084$	1.00
$f_{32}$	Fraction of C in slow soil C transferring to passive soil C	-	0	0.05	0.0315	$0.042 \pm 0.0055$	1.00
$f_{13}$	Fraction of C in passive soil C transferring to fast soil C	-	0.3	0.7	0.45	$0.50 \pm 0.11$	1.00
$k_1$	Turnover rate of C from fast soil C	$\text{g C g C}^{-1} \text{yr}^{-1}$	1	15	7.3	$7.97 \pm 3.72$	1.00
$k_2$	Turnover rate of C from slow soil C	$\text{g C g C}^{-1} \text{yr}^{-1}$	0.1	0.5	0.2	$0.28 \pm 0.11$	1.00
$k_3$	Turnover rate of C from passive soil C	$\text{g C g C}^{-1} \text{yr}^{-1}$	0.001	0.01	0.0045	$0.0013 \pm 0.00020$	1.00
$Q_{10}$	Temperature sensitivity	-	1	3	2	$1.28 \pm 0.054$	1.00

LL lower limit, UL upper limit  
Parameter names, ranges, units, default values, posterior mean, and standard deviation (SD), and G-R statistics in the conventional Century-type model

**Table 2 Descriptions of parameters in the vertically resolved model**

Short name	Description	Unit	LL	UL	Default	Mean $\pm$ SD	G-R
$D_1$	Diffusivity in non-permafrost regions	$\text{m}^2 \text{yr}^{-1}$	$0.3 \times 10^{-4}$	$16 \times 10^{-4}$	$1 \times 10^{-4}$	$(1.4 \pm 0.81) \times 10^{-4}$	1.02
$D_2$	Diffusivity in permafrost regions	$\text{m}^2 \text{yr}^{-1}$	$0.3 \times 10^{-4}$	$16 \times 10^{-4}$	$4 \times 10^{-4}$	$(9.3 \pm 4.0) \times 10^{-4}$	1.00
$z_t$	e-folding depth for depth scalar	m	0	1	0.5	$0.41 \pm 0.05$	1.00
$t_1$	Coefficient to calculate $f_{21}$ (intercept)	-	0	1	0.85	$0.28 \pm 0.14$	1.00
$t_2$	Coefficient to calculate $f_{21}$ (slope)	-	0	1	0.68	$0.089 \pm 0.069$	1.00
$f_{31}$	Fraction of C in fast soil C transferring to passive soil C	-	0	0.01	0.005	$0.0053 \pm 0.0026$	1.00
$f_{12}$	Fraction of C in slow soil C transferring to fast soil C	-	0.1	0.6	0.4185	$0.37 \pm 0.13$	1.00
$f_{32}$	Fraction of C in slow soil C transferring to passive soil C	-	0	0.05	0.0315	$0.031 \pm 0.011$	1.00
$f_{13}$	Fraction of C in passive soil C transferring to fast soil C	-	0.3	0.7	0.45	$0.50 \pm 0.11$	1.00
$k_1$	Turnover rate of C from fast soil C	$\text{g C g C}^{-1} \text{yr}^{-1}$	1	15	7.3	$7.52 \pm 3.76$	1.00
$k_2$	Turnover rate of C from slow soil C	$\text{g C g C}^{-1} \text{yr}^{-1}$	0.1	0.5	0.2	$0.22 \pm 0.086$	1.01
$k_3$	Turnover rate of C from passive soil C	$\text{g C g C}^{-1} \text{yr}^{-1}$	0.001	0.01	0.0045	$0.0048 \pm 0.0018$	1.01
$Q_{10}$	Temperature sensitivity	-	1	3	2	$1.03 \pm 0.024$	1.00

LL lower limit, UL upper limit  
Parameter names, ranges, units, default values, posterior mean, and standard deviation (SD), and G-R statistics in the vertically resolved model

decrease in soil C in response to climate change, but the microbial model could predict either negative or positive responses depending on parameter values; the projected soil C content would be correlated with initial conditions in the two non-microbial models but not in the microbial model. We found that CLM 4.5 and MIMICS with data-driven parameter values projected much greater uncertainties in soil C responses to climate change than the conventional Century-type soil C model, with both positive and negative C-climate feedbacks.

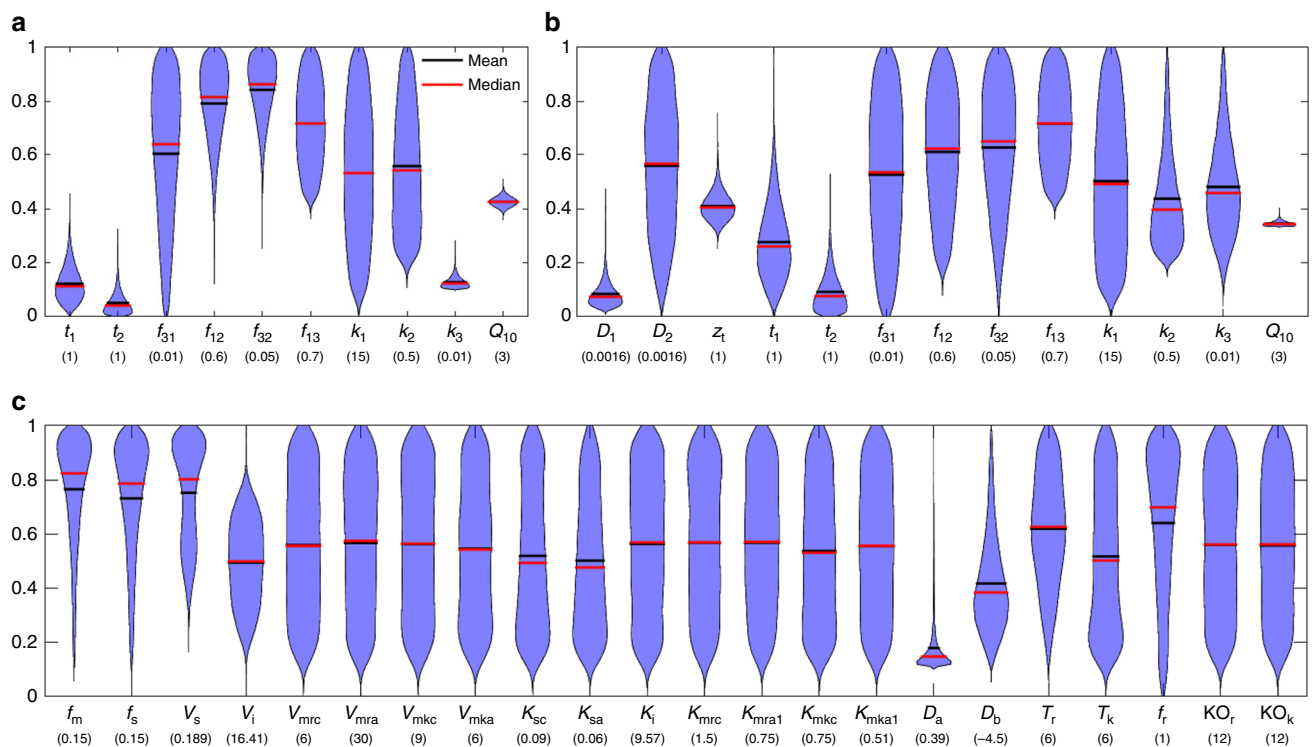
## Results

**Posterior distribution of model parameters.** Data-driven parameter ensembles were derived from assimilating re-gridded global soil C data in the Harmonized World Soil Database (HWSD<sup>10</sup>) and the Northern Circumpolar Soil Carbon Database (NCSCD<sup>21</sup>) (Supplementary Figs. 1, 2 and see Methods). We assumed that the soil C content in the HWSD and NCSCD was at steady state. Their probability distributions are log-normal (Supplementary Fig. 3). Applying Bayes' theorem, we used the MCMC technique to generate posterior ensemble parameter values within parameter boundaries (Tables 1 and 2, Supplementary Table 1). The probability inversion was effective in terms of constraining the targeted parameters in the three soil C decomposition models (Fig. 2, Supplementary Figs. 4, 5, 6). Specifically, coefficients (i.e.,  $t_1$  and  $t_2$ ) for calculating  $f_{21}$  (fraction of C in fast soil C transferring to slow

soil C), decay rate of passive soil C ( $k_3$ ), and temperature sensitivity ( $Q_{10}$ ) were well constrained in the conventional Century-type model (Fig. 2a and Supplementary Fig. 4); observed soil C vertical profiles further helped constrain the decay rate of slow soil C ( $k_2$ ) in the vertically resolved model, particularly for the vertical diffusivity parameters ( $D_1$  and  $D_2$ ) and e-folding depth ( $z_t$ ) (Fig. 2b, Supplementary Fig. 5). Interestingly, the posterior mean of  $D_2$  was found to be larger than  $D_1$ , as diffusivity in permafrost soil was found to be faster than non-permafrost soil, which is mainly due to higher cryoturbation<sup>11</sup>. The  $Q_{10}$  mean is 1.25 in the conventional model and 1.06 in the vertically resolved model, both of which are less than the default value (2), but close to empirical values<sup>22</sup>. The transfer coefficients ( $f_{ij}$ ) were not well constrained in either of the two models.

In MIMICS, parameters related to uptake rate ( $V_s$  and  $V_i$ ) and desorption rate of physically protected soil C ( $D_a$  and  $D_b$ ), and proportion of litter input and the two microbial C pools ( $f_m$ ,  $f_s$ , and  $f_r$ ) were well constrained (Fig. 2c and Supplementary Fig. 6). None of the modifiers (e.g.,  $V_{\text{mrc}}$ ,  $V_{\text{mkc}}$ ,  $K_{\text{mrc}}$ , and  $K_{\text{mkc}}$ ) for calculating uptake rate and half saturation constant were well constrained.

**Steady states in soil carbon stock.** The spatial patterns of the estimated soil C content by the three models were comparable to the soil C database (Fig. 3a, c, e). The conventional model



**Fig. 2** Violin plot of the accepted model parameter values in the three decomposition models. **a** The conventional model; **b** the vertically resolved model; **c** the microbial model (MIMICS). The narrower distributions of the parameters suggest that they are better constrained. For posterior distribution of each parameters, please refer to Supplementary Figs. 4, 5, and 6. For meanings of parameters, refer to Tables 1, 2, and Supplementary Table 1. Note that the values on the y-axis are normalized to the range of [0, 1] by the corresponding values in the brackets under each parameter. In the other words, the real values can be derived by multiplying the values in the brackets

significantly underestimated soil C in the high latitude (Fig. 3a) with large errors (Fig. 3b). However, the vertically resolved model generated relative small biases and errors in the high latitude (Fig. 3c, d) due to its vertical dynamics and explicit parameterization for permafrost soil. This finding highlights the advantage of adding soil layers to simulating high latitude soil C. The smaller soil C biases in the high latitude in MIMICS is due partly to the fact that observed soil C down to 1 m depth instead of 3 m was compared with the modeled soil C (Fig. 3e, f).

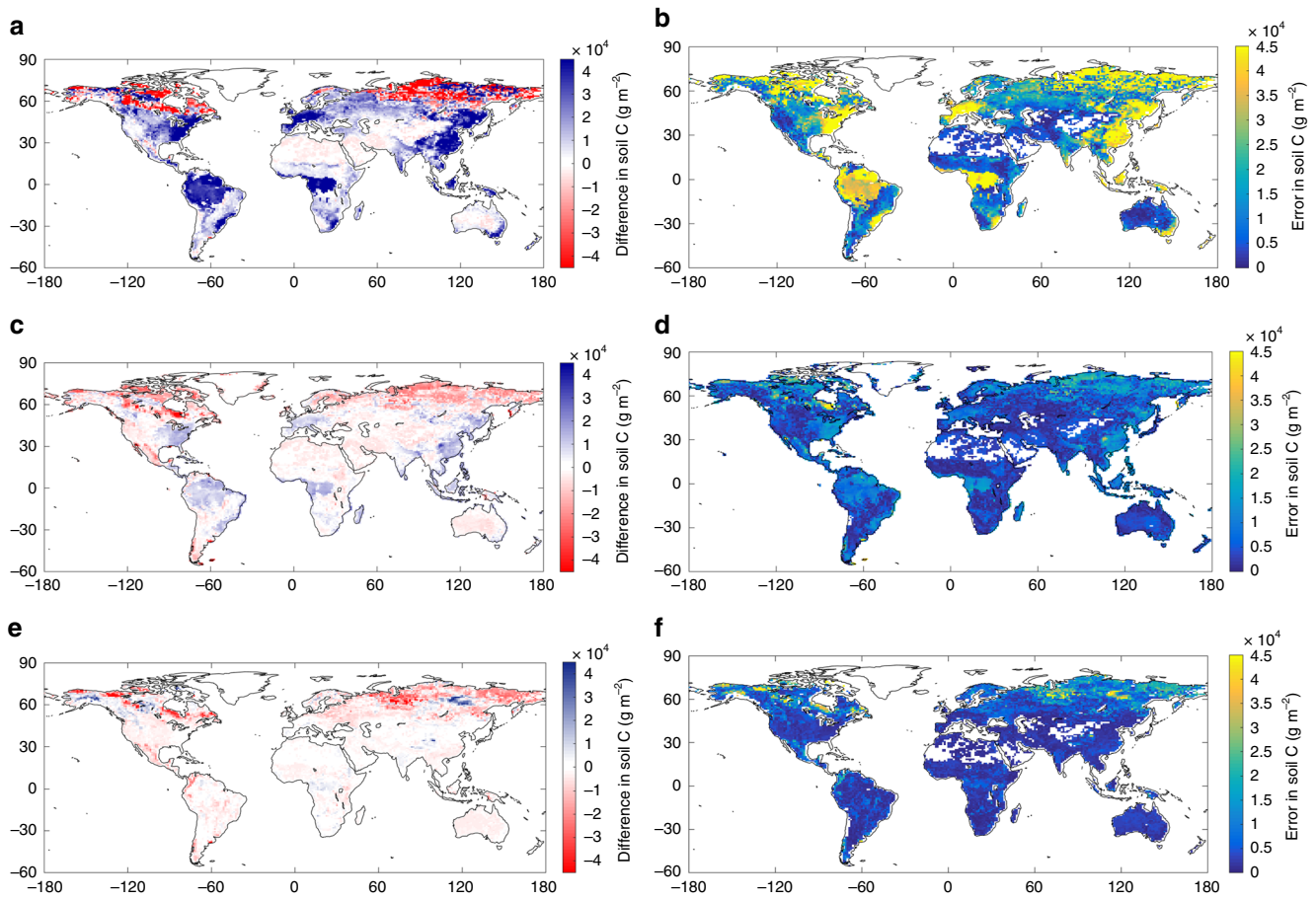
Positive biases compared to the observational data were widespread in the conventional and vertically resolved models in low latitude (Fig. 3a, c). In particular, the conventional Century-type model strongly overestimated soil C in tropical and near-coastal areas with large errors (Fig. 3a, b). The similar spatial biases between the two non-microbial models suggest the similarity in their structures. However, compared with the conventional model, the vertically resolved model had smaller errors, showing improvement by explicitly adding soil depth. In contrast to the two non-microbial models, the microbial model estimation showed small biases in the low latitude (Fig. 3e) with small errors (Fig. 3f). Overall, MIMICS estimated the best spatial fit to the observational data due possibly to having more parameters and explicit microbial dynamics<sup>20,23,24</sup>. As a result, CLM 4.5 and MIMICS generated smaller contemporary global total soil C relative to the observations, while the conventional Century-type model generated greater contemporary global total soil C in comparison to the observations (Supplementary Fig. 7).

**Uncertainties in soil carbon projections.** To illustrate the impact of parameter uncertainty on long-term soil C projection, we performed forward runs over 21st century with 1000 sets of

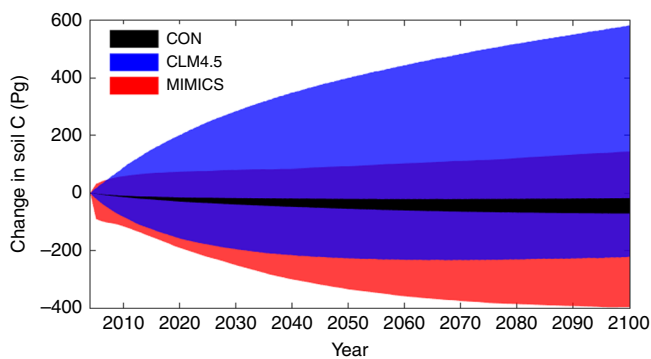
parameter values drawn from the posterior distribution for each model (Methods). The three models projected substantially different changes and trajectories in global total soil C over 21st century (Fig. 4). The conventional model projected consistent soil C loss with the least uncertainty (95% confidence interval:  $-71$  to  $-17$  Pg). Adding vertical resolution or microbial dynamics to the conventional model increased the projection uncertainty (95% confidence interval:  $-222$  to  $583$  Pg C and  $-397$  to  $144$  Pg C, respectively) as well as the sign of the soil C-climate feedback depending on parameters. The uncertainties in the vertically resolved model or MIMICS are more than 10 times larger than that in the conventional model. This interesting result shows that using more parameters and more explicit dynamics may lead to a larger prediction uncertainty due to feedbacks in the model dynamics, rather than less.

#### Sensitivity to initial conditions and model parameters.

Uncertainties in projected soil C among models have been linked to the model initial conditions in previous research<sup>5,13</sup>. Our results show that the initial conditions ( $S_i$ ) tightly correlated with the projected soil C in the two non-microbial models at global (Fig. 5; Supplementary Fig. 8a, c) and grid scale (Supplementary Fig. 9a, c), but did not correlate well with the changes in soil C in all the three models (Fig. 5; Supplementary Fig. 8b, d, f; Supplementary Fig. 9b, d, f) except for some low-latitude areas in the conventional model (Supplementary Fig. 9b). The microbial model's initial conditions were not correlated well with projected soil C at the global scale (Fig. 5 and Supplementary Fig. 8e) or grid scale (Supplementary Fig. 9e), except for significant correlations with predicted soil C at high latitudes (Supplementary Fig. 9e). The findings suggest that in general, uncertainty in the



**Fig. 3** Difference between estimated mean soil carbon and the soil carbon database. **a, b** are the difference and root mean squared error in the conventional model, respectively; **c, d** are the difference and root mean squared error in the vertically resolved model, respectively; **e, f** are the difference and root mean squared error in MIMICS, respectively



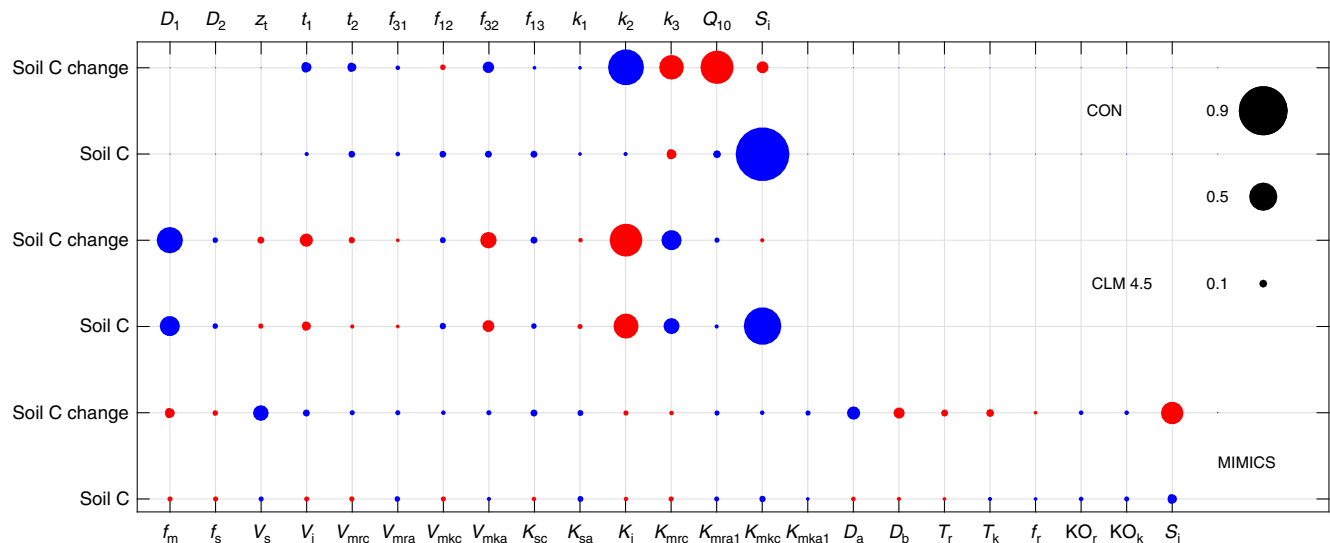
**Fig. 4** Changes in global total soil carbon under RCP 8.5. 1000 parameter sets were randomly sampled from the posterior distribution of parameters to generate the temporal trajectories in each model. The shaded area represents the 95% confidence interval generated by 1000 parameter sets. The time step is shown in year, from 2005 to 2100. CON in the legend is conventional Century-type model

initial conditions propagates through the simulation to the projection of future soil C, and this propagation is especially evident in the two non-microbial models.

Besides initial conditions, model parameters are also able to affect predicted soil C or C changes directly or indirectly through influencing initial conditions. In the conventional model, predicted soil C did not significantly correlate with any

model parameters, but the initial conditions; changes in soil C were positively associated with  $k_2$  (decay rate of slow soil C), but negatively with  $k_3$  (turnover rate of passive soil C) and  $Q_{10}$  (temperature sensitivity of soil C turnover) (Fig. 5). Besides positive correlation with initial conditions in the vertically resolved model, predicted soil C also positively associated with  $D_1$  (diffusivity in non-permafrost soils) and  $k_3$ , but negatively with  $k_2$ ; like predicted soil C, changes in soil C were positively associated with  $D_1$  and  $k_3$ , but negative with  $k_2$  (Fig. 5). Predicted soil C change weakly associated with  $V_s$  (regression coefficient for calculating maximum reaction rate) and  $D_a$  (coefficient for calculating desorption rate from physically protected soil C to available soil C); projected soil C content had no significantly linear relationships with any of the model parameters (Fig. 5).

Consistent with previous research, turnover rates often control soil C changes in the conventional model parameterizations<sup>12,25</sup>. In this study,  $k_2$  is the key parameter for soil C dynamics in the two non-microbial models. However, the relationships between  $k_2$  and soil C changes appear contradicted in the two models, positive in the conventional model but negative in the vertically resolved model. The possible reasoning is that conventional model mainly predicted C loss but vertically resolve model mainly predicted C gain; specifically, in the conventional model, more soil C would be transformed to the passive soil C with larger  $k_2$  to minimize soil C loss; in the vertically resolved model, larger  $k_2$  would cause more C loss to counteract soil C gain. In contrast, neither predicted soil C nor soil C changes were strongly



**Fig. 5** The relationships between soil C, model parameters, and initial conditions ( $S_i$ ) and model parameters under RCP 8.5 in the three models. CON is the conventional Century-type model. Blue circles represent positive correlations and red circles represent negative correlations. The size of a circle is determined by the correlation coefficient between model parameters or initial conditions and soil C or C changes. The parameter labels on the top of the figure are for CON and CLM 4.5; the parameter labels at the bottom are for MIMICS. For the meanings of the parameters, please refer to Tables 1 and 2 and Supplementary Table 1

correlated with any single parameter in the microbial model, but weakly associated with parameters related to decomposition of physically and chemically protected soil C. This finding indicates the complexity of process dynamic and interactions in the microbial model.

## Discussion

Substantial uncertainties in soil C-climate feedback existed 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<sup>19</sup>. These projections are consistent with Hararuk et al.<sup>14</sup>, who showed decreasing soil C under RCP 8.5 in a similar model to the one employed here, and are also within the predicted range in CMIP5 models<sup>5</sup>.

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 diffusivity of soil C across soil layers due to diffusion<sup>11</sup>, 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. However, the large spread in projection and lack of constraint on transfer coefficients suggest more relevant data requirement to reduce model uncertainty.

The microbial model also predicted wider range of future soil C change and diverse trajectories with both negative and positive

feedbacks. This was somewhat expected due to the non-linearity of the C uptake processes by microbes<sup>23,26,27</sup>. 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 especially needed to tease apart multiple processes and further reduce the uncertainty. Results from this study highlight that data constraints may limit the ability of data assimilation to reduce uncertainty in more complicated model structures.

The large uncertainty in the vertically resolved model and MIMICS might be engendered by either high degrees of model freedom or complex model structures, or both. Many unconstrained parameters due to data limitation, especially in MIMICS, led to the large predicted uncertainty. We therefore anticipate substantial uncertainty reduction in these models once more global data are available to inform the models in the future. On the other hand, the complex model structures may also contribute to the large uncertainty. In a previous study, Hararuk et al.<sup>14</sup> reported small uncertainty in soil C prediction by a conventional model with 20 free parameters. The number of free parameters is comparable to that in MIMICS (22) and is greater than that in the vertically resolved model (13). Therefore, it is likely that the model structures of CLM4.5 and MIMICS at least in part increased the projected uncertainty. In addition, slightly larger uncertainty and fewer model parameters in the vertically resolved model than in the MIMICS also support that the greater uncertainty is likely caused by model structure, if not solely. In contrast to larger uncertainty in CLM 4.5 and MIMICS, both models generated better spatial comparison to soil C observations, which demonstrates model flexibility and encourages further exploration of explicit soil depth and microbial dynamics in soil C model parameterization.

The lack of constraint on the transfer coefficients in the two non-microbial 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 modifiers of microbial maximum uptake rate and half saturation constant are of the most importance to reduce uncertainty. We acknowledge that besides transfer coefficients in the two non-microbial models and the modifiers in the microbial model, data related to other aspects of model processes may be also critical, but not revealed in our study due to model simplification. For example, data related to derivation of soil water scalar such as soil water content and water potential may be needed given that we simplified the calculation of water scalars by directly using the default parameters in the two non-microbial models and that there is under-representation of soil water impact in the microbial model<sup>18</sup>.

Isotopic data in C processes show great potential to constrain these processes<sup>11,28,29</sup>. Indeed, <sup>14</sup>C soil profiles have been used to constrain transfer coefficients and turnover rates at multiple sites<sup>28,30</sup>; 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<sup>31</sup>, sources of input<sup>32</sup>, or proportion of different soil C pools<sup>33</sup>. Other additional data, such as soil respiration<sup>34</sup> and soil C incubation datasets<sup>35,36</sup> 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. 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 with 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<sup>15</sup>.

We caution the reader on several assumptions and simplifications 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<sup>10,12,14</sup>.

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.

Lastly, we simplified the two non-microbial models in terms of their environmental modifiers, soil water scalar in particular, given data limitation and potential equifinality by the complex calculation of soil water scalar<sup>37</sup>. This simplification may underestimate the uncertainty in predictions by the two models; however, it is less likely for the conventional Century-type model to reach the similar magnitude of uncertainty in MIMICS even with the full representation of soil water scalar due to the large difference.

In summary, our results demonstrate the importance of model structure and parameterization in determining the predicted soil C response to climate change. CLM4.5 and MIMICS showed much greater uncertainty in projected soil C under RCP 8.5 and the conventional model consistently predicted strong positive C-climate feedback. The close correlations between initial conditions and projected soil C confirm that the projection of soil C is sensitive to initial conditions in the two non-microbial models, whereas the microbial model did not show any linear relationship between initial conditions and projected soil C.

To increase confidence in soil C projection, diverse model structures are necessary<sup>16</sup>, given that CLM 4.5 and MIMICS outperformed the conventional model in terms of estimation in the spatial distribution of soil C. However, the larger uncertainty in the projection of soil C by the two models also suggests that we need to strike a balance between model complexity and the need to include diverse model structures in order to forecast soil C dynamics with high confidence and low uncertainty. In addition, reducing the uncertainties of the two models require more observations than are available at the present time. Overall, whether vertically resolved models and microbial models are better representations of mechanisms for soil C dynamics remain debatable. However, they represent updated knowledge and important alternate model structures to enhance confidence in prediction. Our findings suggest that the scientific community should include alternative model structures in future ensemble model predictions and comparisons to increase projection confidence.

## Methods

**Models.** We performed parameter estimation on three soil C decomposition models with different but representative structures (Fig. 1). The conventional Century-type model (Fig. 1a) has the same structure as soil C cascade embedded in the Community Land Model version 4.5 (CLM 4.5) without activating the depth-resolved parameterization; to incorporate the effects of depth-resolved dynamics, we used the model in the CLM 4.5<sup>14,17</sup> (CLM 4.5bg; Fig. 1b); we used the MIMICS model<sup>20</sup> to explicitly simulate microbial soil decay processes (Fig. 1c). We briefly introduced the model parameterization in each model as follows.

The conventional model (Fig. 1a) represents soil C decomposition using three C pools and can be written in a matrix form as:

$$X'(t) = R + F \times \xi(t) \times K \times X(t). \quad (1)$$

$R = [R_1, R_2, R_3]$  is the litter input to the three soil carbon pools (labile, slow, and passive soil C).

$$F = \begin{bmatrix} -1 & f_{12} & f_{13} \\ f_{21} & -1 & 0 \\ f_{31} & f_{32} & -1 \end{bmatrix}$$

is a transfer matrix among soil C pools, where  $f_{21}$  is derived using the equation  $f_{21} = 1 - t - f_{31}$  where  $t = t_1 - t_2 \times (1 - \text{sand}\%)$ . The respiration coefficients can be derived by  $1 - f_{i,j}$  for each carbon pool.  $K = [K_1, K_2, K_3]$  is baseline turnover rate of soil C pools.  $\xi$  is the environmental modifier, and is a product of temperature scalar, soil moisture scalar, soil nitrogen scalar, and oxygen scalar.  $X = [X_1, X_2, X_3]$  is the soil C content in each of the three pools. To be consistent with the other two models, we only calculate temperature scalar using the equation  $Q_{10}^{(T_{\text{soil}} - 25)/10}$ , where  $Q_{10}$  is temperature sensitivity of decomposition and  $T_{\text{soil}}$  is soil temperature. The rest of the environmental scalars are average across the 10 soil layers from CLM 4.5. There are in total 10 global parameters in the conventional model. See Table 1 for detailed description of each parameter and its range and default value in the conventional model (i.e., CLM 4.5 without activating soil depth module).





The solution for MIMICS is

$$\text{SOM}_{p,ss} = (R_{l-p} + R_{mic-p})/D \quad (12)$$

$$\text{SOM}_{c,ss} = (-\text{coeff}_f + \sqrt{\text{coeff}_f^2 - 4 \times \text{coeff}_e \times \text{coeff}_g}) / (2 \times \text{coeff}_e) \quad (13)$$

$$\text{SOM}_{a,ss} = (-\text{coeff}_m + \sqrt{\text{coeff}_m^2 - 4 \times \text{coeff}_l \times \text{coeff}_n}) / (2 \times \text{coeff}_l), \quad (14)$$

where  $R_{l-p}$  and  $R_{mic-p}$  are the input to physically protected soil C from litter and microbial C, respectively;  $D$  is the turnover rate of  $\text{SOM}_p$ ,

$$\text{coeff}_e = \text{MIC}_r \times V_{\text{max},r} + \text{MIC}_k \times V_{\text{max},k} - R \quad (15)$$

$$\text{coeff}_f = \text{MIC}_r \times V_{\text{max},r} \times K_{o,k} \times K_{m,k} + K_{o,r} \times K_{m,r} \times \text{MIC}_k \times V_{\text{max},k} - (K_{o,r} \times K_{m,r} + K_{o,k} \times K_{m,k}) \times R \quad (16)$$

$$\text{coeff}_g = -(K_{o,r} \times K_{m,r} + K_{o,k} \times K_{m,k}) \times R \quad (17)$$

$$\text{coeff}_l = \text{MIC}_r \times V_{\text{max},r} + \text{MIC}_k \times V_{\text{max},k} - R \quad (18)$$

$$\text{coeff}_m = \text{MIC}_r \times V_{\text{max},r} \times K_{m,k} + K_{m,r} \times \text{MIC}_k \times V_{\text{max},k} - (K_{m,r} + K_{m,k}) \times R \quad (19)$$

$$\text{coeff}_n = -(K_{m,r} \times K_{m,k}) \times R, \quad (20)$$

where  $R$  is the total input to respective soil carbon. For more details about these parameters, refer to Supplementary Table 1 and Wieder et al.<sup>20</sup>.

The steady state of total microbial biomass (sum of the  $r$ -selection and  $k$ -selection microbial biomass) was from a published database in microbial biomass<sup>40</sup> ([https://daac.ornl.gov/SOILS/guides/Global\\_Microbial\\_Biomass\\_C\\_N\\_P.html](https://daac.ornl.gov/SOILS/guides/Global_Microbial_Biomass_C_N_P.html)). Some details are also provided in the data section below.

Note that all the litter inputs to soils in the three models are the same at grid level to avoid introducing bias into the results among models. The inputs were derived from running CLM 4.5 within 1850–2004. Mean annual input within 1850–2004 were used for steady-state calculation. Specifically, we used the Qian bias-corrected reanalysis dataset<sup>41</sup> 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.

**Datasets.** The observations we used were re-gridded top soil organic carbon (0–30 cm, upper panel) and subsoil organic carbon (30–100 cm, lower panel) from HWSD [Food Agriculture Organization, 2012]. The native resolution (30 arc sec) in HWSD was re-gridded to the CLM grid (i.e.,  $1.25 \times 0.94^\circ$ ) by Wieder et al.<sup>10</sup> (data source: <https://daac.ornl.gov/SOILS/guides/HWSD.html>).

Due to the possible underestimation of permafrost soil C in HWSD, we replaced it with the NCSCD (<http://bolin.su.se/data/ncscd/netcdf.php>). The NCSCD was developed to quantify the Northern Circumpolar permafrost soil C stocks down to 3 m<sup>21</sup>. There are four soil layers in this database, 0–30, 0–100, 100–200, and 200–300 cm. We re-gridded the NCSCD at  $1^\circ$  resolution to the CLM 4.5 grid, the same resolution as the HWSD and microbial biomass database.

We use NCAR Command Language (NCL) with Earth System Modeling Framework (ESMF) software. This software generated weights for data regridding between different resolutions. The conserve method was selected to insure the total amount of regional soil C to be conserved after regridding.

The microbial biomass C was used as the steady state in the MIMICS model due to the fast turnover of microbes, which is typically less than 1 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. The remainder is the  $k$ -selection microbial biomass. The gridded soil microbial biomass C was in  $0.5^\circ$  resolution and available here [https://daac.ornl.gov/SOILS/guides/Global\\_Microbial\\_Biomass\\_C\\_N\\_P.html](https://daac.ornl.gov/SOILS/guides/Global_Microbial_Biomass_C_N_P.html). The database was re-gridded to the CLM grid. The soil depth is down to 1 m<sup>40</sup>.

Soil microbial biomass C data were re-gridded into  $0.94^\circ \times 1.25^\circ$  resolution from  $0.5^\circ \times 0.5^\circ$  resolution. We use NCL with ESMF software. This software generated weights for data regridding between different resolutions. The conserve method was selected to insure the total amount of regional soil C to be conserved after regridding.

**Model-data fusion.** We applied Bayes' theorem to estimate parameter values and associated uncertainties<sup>25,42</sup>.

$$p(\theta|Z) = \frac{p(Z|\theta) \times p(\theta)}{p(Z)}, \quad (21)$$

where  $p(\theta|Z)$  is the posterior distribution of the parameters  $\theta$  given the observations  $Z$ .  $p(Z|\theta)$  is the likelihood function for a parameter set calculated with the assumption that each parameter is independent from all other parameter and has log-normal distribution<sup>5,43</sup> (Supplementary Fig. 3) with a zero mean:

$$P(Z|\theta) \propto \exp\left\{-\sum \frac{[Z_i - \theta_i \times X]^2}{2\sigma_i^2}\right\}. \quad (22)$$

Here  $Z_i$  is the logarithm of  $i^{\text{th}}$  soil C observation in the observational database,  $X$  are the logarithms of the carbon pools from the model, and  $\theta$  is the mapping vector that maps the simulated carbon pools to observations.  $X$  is derived by assuming the current soil status is at steady state. We were conservative in assigning errors to the soil C with  $\sigma = 0.5 \times Z_i$ . For the conventional model, we assimilated the data by aggregating all the soil layers together which is 0–100 cm in non-permafrost regions and 0–300 cm permafrost regions; for CLM 4.5, we assimilate data for 0–100 cm in HWSD for non-permafrost soils, 0–100, 100–200, and 200–300 cm in NCSCD for permafrost soils, independently. In contrast to soils elsewhere, permafrost regions contain a huge amount of carbon stock in deeper soil, which validates the use of deeper soil carbon data. For MIMICS, however, we assimilated the soil C data down to 100 cm only due to the explicit 1-m depth parameterization<sup>20</sup>.

We assumed that the parameters are distributed uniformly within their prior ranges. Since most of the parameter range in MIMICS are unknown, we assumed the range of the distribution to be  $[\theta_o/3, 3\theta_o]$ , where  $\theta_o$  is the default value. Posterior probability distributions of parameters were obtained using a Metropolis–Hastings (M–H) algorithm, a MCMC technique<sup>44,45</sup>. The detailed description of M–H algorithm can be found in Xu et al.<sup>42</sup>.

In brief, the M–H algorithm consists of iterations of two steps: a proposing step and a moving step. In the proposing step, a new parameter set  $\theta^{\text{new}}$  is proposed based on the previously accepted parameter set  $\theta^{\text{old}}$  and a proposal distribution, which was uniform in our study:

$$\theta^{\text{new}} = \theta^{\text{old}} + r \times (\theta_{\text{max}} - \theta_{\text{min}})/D, \quad (23)$$

where  $\theta_{\text{max}}$  and  $\theta_{\text{min}}$  are the maximum and minimum values of parameters,  $r$  is a random variable between  $-0.5$  and  $0.5$ , and  $D$  is used to control the proposing step size and was set to 5 as is Xu et al.<sup>42</sup>. In each moving step,  $\theta^{\text{new}}$  was tested against the Metropolis criterion to examine if the new parameter set should be accepted or rejected. The first 2500 accepted samples were discarded (burn-in period) and the rest were used to generate posterior parameter distributions. In total, there are 50,000 accepted samples to construct the posterior distribution.

**Projection in soil carbon under RCP 8.5.** The soil C input and environmental modifiers (except the temperature scalar) used to drive the models were derived from running original CLM 4.5 under the worst-case greenhouse gas emission scenario, representative concentration pathway 8.5 (RCP 8.5). The atmospheric data were from Community Earth System Model output for the Representative Concentration Pathway 8.5 experiment, which were used to force CLM 4.5 for 2005–2100. We output the soil C input from litter and then used the inputs to drive our soil C models. The spatial and temporal changes in soil temperature, water content, and litter input to soil C within 2005–2100 were presented (Supplementary Figs. 10, 11).

We randomly sampled 1000 parameter sets out of the accepted posterior values. With each of the sampled parameter sets, we forced the vertically resolved model with C input and environmental modifiers obtained from CLM 4.5 model under RCP 8.5 from 2005–2100. For the conventional and microbial model, we derived total inputs and mean environmental modifiers of all the 10 soil layer to force the two models. For the two non-microbial models, we used a monthly time step. For the microbial model, daily time step was used due to its non-linear nature causing instability for longer time steps.

**Convergence of MCMC.** We used Gelman–Rubin (G–R) diagnostic method to determine convergence of MCMC simulations<sup>46</sup>. 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  $c_i$ , the parameter samples from  $K$  ( $K = 5$ ) parallel M–H runs of length  $N$  ( $N = 10,000$ ), the between ( $B_i$ ) and within-run ( $W_i$ ) variances are defined as:

$$B_i = \frac{N}{K-1} \sum_{k=1}^K (\bar{c}_i^k - \bar{c}_i)^2 \quad (24)$$

$$W_i = \frac{1}{K(N-1)} \sum_{k=1}^K \sum_{n=1}^N (\bar{c}_i^{n,k} - \bar{c}_i^k)^2. \quad (25)$$

The G–R scale reduction statistics is given by:

$$GR_i = \sqrt{\frac{W_i(N-1)/N + B_i/N}{W_i}} \quad (26)$$

Once convergence is reached,  $GR_i$  should approximately equal one.

**Data availability.** Data assimilation algorithms and parameter ensembles are available in the GitHub (<https://github.com/zshi0609/Global-DA-Project>). Large input datasets are available upon request to the correspondence authors.

Received: 8 August 2017 Accepted: 4 May 2018

Published online: 04 June 2018

## References

- IPCC. *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. et al.) 1535 (Cambridge University Press, Cambridge, UK and New York, USA, 2013).
- Arneeth, A. et al. Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed. *Nat. Geosci.* **10**, 79–84 (2017).
- Le Quéré, C. et al. Global carbon budget 2016. *Earth Syst. Sci. Data* **8**, 605–649 (2016).
- Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
- Todd-Brown, K. E. O. et al. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **10**, 1717–1736 (2013).
- Todd-Brown, K. E. O. et al. Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* **11**, 2341–2356 (2014).
- Eyring, V. et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
- Huntzinger, D. N. et al. The north American carbon program multi-scale synthesis and terrestrial model intercomparison project—Part 1: overview and experimental design. *Geosci. Model Dev.* **6**, 2121–2133 (2013).
- Tian, H. et al. Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: current status and future directions. *Glob. Biogeochem. Cycles* **29**, 775–792 (2015).
- Wieder, W. R., Boehnert, J. & Bonan, G. B. Evaluating soil biogeochemistry parameterizations in Earth system models with observations. *Glob. Biogeochem. Cycles* **28**, 211–222 (2014).
- Koven, C. D. et al. The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences* **10**, 7109–7131 (2013).
- Hararuk, O., Smith, M. J. & Luo, Y. Microbial models with data-driven parameters predict stronger soil carbon responses to climate change. *Glob. Change Biol.* **21**, 2439–2453 (2015).
- Exbrayat, J. F., Pitman, A. J. & Abramowitz, G. Response of microbial decomposition to spin-up explains CMIP5 soil carbon range until 2100. *Geosci. Model Dev.* **7**, 2683–2692 (2014).
- Hararuk, O., Xia, J. & Luo, Y. Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov Chain Monte Carlo method. *J. Geophys. Res.* **119**, 2013JG002535 (2014).
- Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L. & Williams, M. The decadal state of the terrestrial carbon cycle: global retrievals of terrestrial carbon allocation, pools, and residence times. *Proc. Natl Acad. Sci. USA* **113**, 1285–1290 (2016).
- Bradford, M. A. et al. Managing uncertainty in soil carbon feedbacks to climate change. *Nat. Clim. Change* **6**, 751–758 (2016).
- Koven, C. D., Lawrence, D. M. & Riley, W. J. Permafrost carbon—climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics. *Proc. Natl Acad. Sci. USA* **112**, 3752–3757 (2015).
- Wieder, W. R. et al. Explicitly representing soil microbial processes in Earth system models. *Glob. Biogeochem. Cycles* **29**, 1782–1800 (2015).
- Luo, Y. et al. Toward more realistic projections of soil carbon dynamics by Earth system models. *Glob. Biogeochem. Cycles* **30**, 40–56 (2016).
- Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G. & Bonan, G. B. Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geosci. Model Dev.* **8**, 1789–1808 (2015).
- Hugelius, G. et al. A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region. *Earth Syst. Sci. Data* **5**, 393–402 (2013).
- Mahecha, M. D. et al. Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science* **329**, 838–840 (2010).
- Allison, S. D., Wallenstein, M. D. & Bradford, M. A. Soil-carbon response to warming dependent on microbial physiology. *Nat. Geosci.* **3**, 336–340 (2010).
- Li, J. W., Wang, G. S., Allison, S. D., Mayes, M. A. & Luo, Y. Q. Soil carbon sensitivity to temperature and carbon use efficiency compared across microbial-ecosystem models of varying complexity. *Biogeochemistry* **119**, 67–84 (2014).
- Weng, E. S. & Luo, Y. Q. Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. *Ecol. Appl.* **21**, 1490–1505 (2011).
- Wang, Y. P. et al. Oscillatory behavior of two nonlinear microbial models of soil carbon decomposition. *Biogeosciences* **11**, 1817–1831 (2014).
- Wang, Y. P. et al. Responses of two nonlinear microbial models to warming and increased carbon input. *Biogeosciences* **13**, 887–902 (2016).
- He, Y. et al. Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century. *Science* **353**, 1419–1424 (2016).
- Dwivedi, D. et al. Mineral properties, microbes, transport, and plant-input profiles control vertical distribution and age of soil carbon stocks. *Soil Biol. Biochem.* **107**, 244–259 (2017).
- Feng, W. et al. Methodological uncertainty in estimating carbon turnover times of soil fractions. *Soil Biol. Biochem.* **100**, 118–124 (2016).
- Andrews, J. A., Harrison, K. G., Matamala, R. & Schlesinger, W. H. Separation of root respiration from total soil respiration using Carbon-13 labeling during free-air carbon dioxide enrichment (FACE). This work was completed while the senior author was at the Dep. of Botany, Duke Univ., Durham, NC. *Soil Sci. Soc. Am. J.* **63**, 1429–1435 (1999).
- Singh, N. et al. Transformation and stabilization of pyrogenic organic matter in a temperate forest field experiment. *Glob. Change Biol.* **20**, 1629–1642 (2014).
- Torn, M. S. et al. A dual isotope approach to isolate soil carbon pools of different turnover times. *Biogeosciences* **10**, 8067–8081 (2013).
- Bond-Lamberty, B. & Thomson, A. Temperature-associated increases in the global soil respiration record. *Nature* **464**, 579–582 (2010).
- Liang, J. et al. Methods for estimating temperature sensitivity of soil organic matter based on incubation data: a comparative evaluation. *Soil Biol. Biochem.* **80**, 127–135 (2015).
- Xu, X. et al. Soil properties control decomposition of soil organic carbon: results from data-assimilation analysis. *Geoderma* **262**, 235–242 (2016).
- Bonan, G., Drewniak, B., & Huang, M. Technical description of version 4.5 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-503+STR, Boulder, CO (2013).
- Patankar, S. Numerical Heat Transfer and Fluid Flow (Taylor & Francis, Boca Raton, 1980).
- Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Change* **3**, 909–912 (2013).
- Xu, X., Thornton, P. E. & Post, W. M. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Glob. Ecol. Biogeogr.* **22**, 737–749 (2013).
- Qian, T., Dai, A., Trenberth, K. E. & Oleson, K. W. Simulation of global land surface conditions from 1948 to 2004. Part I: forcing data and evaluations. *J. Hydrometeorol.* **7**, 953–975 (2006).
- Xu, T., White, L., Hui, D. F., & Luo, Y. Q. Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. *Glob. Biogeochem. Cycles* **20**, GB2007 (2006).
- Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **47**, 151–163 (1996).
- Hastings, W. K. Monte-Carlo Sampling methods using markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
- Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).

## Acknowledgements

This research was supported by National Aeronautics and Space Administration (NASA) under OCO2 solicitation contract NNX15AJ37G. We would like to thank Chris Lu for the help in regridding soil carbon data. We appreciate the discussion with O. Hararuk, M. KJ, J. Jiang, P. Rayner, A. Schuh, and A. Chatterjee during development of the manuscript.

### Author contributions

Y.L., S.C., B.M., and Z.S. designed the study. Z.S. performed the analyses. All authors contributed to the writing and discussions.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-04526-9>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018