UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPROVING GLOBAL CARBON CYCLE MODELS WITH OBSERVATIONS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

OLEKSANDRA HARARUK
Norman, Oklahoma
2014

IMPROVING GLOBAL CARBON CYCLE MODELS WITH OBSERVATIONS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Yiqi Luo, Chair

_____
Dr. S. Lakshmivarahan

_____
Dr. Heather McCarthy

_____
Dr. Xiangming Xiao

_____
Dr. Daniel Ricciuto

To Andrey

Acknowledgements

I wish to thank my advisor, Dr. Yiqi Luo, for his valuable advice and guidance over the course of the Ph.D. program, his help with manuscript development, teaching me not to lose the sight of a big picture behind the small details, and creating opportunities to develop and improve my leadership skills. I thank Dr. S. Lakshmivarahan for building my knowledge about scientific computing and data assimilation methods from the ground up. I wish to thank Dr. Heather McCarthy, Dr. S. Lakshmivarahan, Dr. Xiangming Xiao, and Dr. Daniel Ricciuto from Oak Ridge National Laboratory for agreeing to serve on my committee.

I would like to thank Dr. Matthew J. Smith from Microsoft Research for sharing his expertise and computational resources, which helped me complete the simulations required for this dissertation hundred times faster than I would without them. I also thank Dr. Ensheng Weng for his help with understanding the codes for global land models, Dr. Shuli Niu for her valuable feedback to my research, Dr. Jianyang Xia, Dr. Xuhui Zhou, Dr. Lifen Jiang, Dr. Xia Xu, Zheng Shi, and other lab members for their help and creating a great working atmosphere. I would like to give special thanks to Adell Hopper for making sure I met departmental and Graduate College deadlines, which allowed me to focus on my research.

Finally and most importantly, I would like to thank my husband, Andrey Koval. His love, support, and faith in all my beginnings played a key role in completion of this dissertation.

Table of Contents

Acknowledgements iv

List of Tables viii

List of Figures ix

Abstract xv

Chapter 1 Introduction 1

Chapter 2 Improvement of global litter turnover rate predictions using a Bayesian MCMC approach 6

2.1 Introduction 8

2.2 Methods 10

2.2.1 Litter decay rate 10

2.2.2 Observed data 11

2.2.3 Parameter estimation 11

2.2.4 Leaf litter feedbacks to climate change 14

2.3 Results and Discussion 15

2.3.1 Estimated parameters 15

2.3.2 Calibrated model performance 18

2.3.3 Global leaf litter distribution 20

2.3.4 Litter feedbacks to climate change 24

2.4 Conclusion 27

Chapter 3 Evaluation and improvement of a global land model against soil carbon data using a Bayesian MCMC method 29

3.1 Introduction 31
3.2 Methods 35
3.2.1 CLM-CASA' and its C-only version for data assimilation 35
3.2.2 Data assimilation for parameter estimation 38
3.2.3 Parameters to be estimated 42
3.2.4 Database 44
3.2.5 Forward analysis of carbon dynamics with original and optimized parameters 46
3.3 Results 46
3.3.1 Evaluation and improvements of modeled SOC 46
3.3.2 Constrained parameters related to soil carbon dynamics 48
3.3.3 Parameter correlations 52
3.3.4 Improvement of soil carbon estimation with data assimilation 54
3.2.5 Carbon pool responses to environmental change 54
3.4 Discussion 57
3.4.1 Current status of soil carbon modeling 57
3.4.2 Improvement of soil carbon modeling 58
3.4.3 Uncertainty in future C projections 61
3.5 Conclusion 62
Supplemental Materials 64
Chapter 4 Modelling microbial processes in soil improves global carbon stocks predictions 66
4.2 Methods 70

4.2.1 Models 70
4.2.2 Data 74
4.2.3 Parameter estimation 75
4.2.4 Forward model runs and stability analysis 77
4.3 Results 79
4.3.1 Performance of the microbial model formulations after calibration 79
4.3.2 Posterior parameter distributions and correlations 82
4.3.3 SOC feedbacks to climate change 89
4.4 Discussion 93
4.4.1 Progress in simulating SOC dynamics 93
4.4.2 Uncertainties in future SOC dynamics 94
4.4.3 Future improvements 96
4.5 Conclusion 98
Supplemental Materials 99
Chapter 5 Conclusions and Implications 105
5.1 Conclusions 106
5.2 Implications for future work 106
References 108

List of Tables

Chapter 2

Table 2.1 Parameter characteristics. MLE is maximum likelihood estimate, and G-R is the result of Gelman-Rubin chain convergence diagnostics

Chapter 3

Table 3.1 CLM-CASA' parameter characteristics. MLE is maximum likelihood estimate, and G-R is the result of Gelman-Rubin chain convergence diagnostics.

Table 3.2 Correlations between the calibrated parameters. The darker the shade of the cell the stronger the correlation The strongest positive correlations were between passive C pool turnover and clay effect on partitioning from slow pool and passive pool; and between sand effect on C partitioning from soil microbial C pool to passive pool and clay effect on partitioning from slow pool to passive pool. The strongest negative correlations were between sand effect on C partitioning from soil microbial C pool to passive pool and sand effect on C partitioning from soil microbial C pool to slow pool.

Chapter 4

Table 4.1 Parameter correlations in the 2-pool microbial model

Table 4.2 Parameter correlations in the 4-pool microbial model

Table S4.1 Model parameter description, prior ranges, posterior ranges, and their convergence indices. MLE is maximum likelihood estimate, and G-R is Gelman-Rubin diagnostics

List of Figures

Chapter 2

Figure 2.1 Posterior parameter distributions for two models with different assumptions of litter quality limitation of its turnover rate: $\boldsymbol{f_1(Q) = \exp(-a \times L_s)}$, and $\boldsymbol{f_2(Q) = LN^{-b}}$. Baseline leaf litter turnover rate was higher under the $\boldsymbol{f_2(Q)}$ assumption, and leaf litter turnover was less sensitive to temperature under $\boldsymbol{f_2(Q)}$ assumption than $\boldsymbol{f_1(Q)}$ assumption.

Figure 2.2 Comparison of the leaf litter turnover rates produced by CENTURY (a), calibrated CENTURY (b), and the model with different assumption about litter quality limitation on its turnover rate (c). Calibration improved performance of CENTURY, however, changing the assumption about litter quality limitation of its turnover rate improved model performance.

Figure 2.3 Global distribution of aboveground litter (excluding woody debris) before (a, c) and after (b, d) calibration of the turnover rates, and their comparisons to the observations from Vogt et al. (1986) (error bars represent standard deviations calculated for the observations per given latitude). Litter pool sizes are determined not only by litter turnover rates, but also by the input rate, therefore we compared modeled litter input rates to the observations (e) reported in Vogt et al. (1986).

Figure 2.4 Comparison of modeled turnover rates to the observed turnover rates from Harmon et al. (2009) before (a) and after (b) calibration. The explained variability of the

observations was correlated with annual precipitation (e), site-level RMSE's were correlated with mean annual temperatures (f), and there were no significant relationships between lignin:N ratio and model residuals (c) or between model residuals and litter lignin content (d).

Figure 2.5 Global leaf litter pool (a), and its response to changing climate with (c) and without (b) the assumption of changing litter quality with increasing CO_2.

Chapter 3

Figure 3.1 CLM-CASA' model structure and global parameter averages. Carbon enters the system through photosynthesis and is partitioned among three live pools. From the live pools carbon is transferred to the five litter pools, and from the litter pools it is transferred to the three soil pools. Values in the boxes are pool residence times; values outside the boxes are partitioning coefficients (values in blue are initial model values, values in red are results of parameter optimization). After data assimilation soil passive C residence time increased by 36 years, whereas slow C residence time decreased by 2.1 years; compared to initial values, more C was transferred to soil C pools from soil litter and less C is transferred to soil C from surface litter.

Figure 3.2 IGBP-DIS soil carbon distribution. Soil carbon varies from 0 kg/m² in deserts to 60 kg/m² in the northern regions.

Figure 3.3 Spatial correspondence of CLM-CASA' produced SOC to the IGBP-DIS reported SOC before (a,b) and after (b,c) data assimilation; and standard deviations in the modeled soil C after data assimilation (d). The points in panel (b) represent the grid cell values. Model with default parameters explained 27% of variation in the observed soil C, whereas model with calibrated parameters explained 41% of variability in the observed soil C. The regions with highest uncertainty were located in the northern latitudes and in the tropics.

Figure 3.4 Frequency distributions of 20 calibrated parameters. The most constrained parameters were temperature sensitivity of heterotrophic respiration ($\mathbf{Q_{10}}$), clay effect on C partitioning from slow to passive pools ($\boldsymbol{t_7}$), and sand effect on C partitioning from soil microbial to passive pool.

Figure 3.5 Change in CLM-CASA's soil (a) and litter (b) carbon under RCP 8.5 climate change scenario. Blue lines are projections of the model with original parameters, red lines – model with maximum likelihood parameters, and gray lines are projections of the models with parameter samples from the posterior distributions (sample size = 2000), representing the uncertainties of projections. Soils released 45 Pg C less in the model with calibrated parameters than the model with original parameters.

Figure 3.6 . Point-by-point comparison of the modeled soil heterotrophic respiration to the observed data derived from Bond-Lamberty and Thomson (2012a). Parameter

calibration did not change the explained variability in the observations, it reduced the RMSE by 10%.

Figure S3.1 SOC stocks under different combinations of C transfer coefficients from plant to soil pools (or C partitioning to soil) and soil C residence times. For the purpose of illustration, NPP in this figure is fixed at 0.1 $\mathbf{kg\ m^{-2}\ year^{-1}}$ and Soil C equals the product of NPP, C transfer coefficient and residence time. The color bar represents soil C pools in $\mathbf{g\ m^{-2}}$. In this figure we illustrate the equifinality of SOC content in a gridcell: same SOC value can be obtained with different combinations of C transfer coefficients from plant to soil and SOC residence times under fixed (observed) NPP.

Figure S3.2 Comparison of traditional and speeded up spin up of soil C.

Chapter 4

Figure 4.1 Schematic representation of the 2-pool (a) and 4-pool (b) microbial models

Figure 4.2 Annual soil C influx used to drive the soil sub-models (a); performance of calibrated microbial and conventional models (b); distribution of changes in the residuals' magnitudes after switching from conventional to microbial model formulation in the environmental space (c): circle diameters represent the relative magnitudes of change, and colors indicate the direction of change; differences in the natural logarithms of calibrated soil residence times between microbial and

conventional models (d): circle diameters represent the relative magnitudes of change, and colors indicate the direction of change

Figure 4.3 Model residuals for microbial biomass plotted in the environmental space: (a) 2-pool microbial model; (b) 4-pool microbial model. Circle diameters represent the relative magnitudes of the residuals.

Figure 4.4 Posterior probability density functions of the microbial models' parameters

Figure 4.5 We calculated Q10's for the microbial models as ratios of SOC turnover rates at temperatures T+10°C to SOC turnover rates at temperatures T. Spatial distribution of the temperature sensitivities of SOC turnover rates and their uncertainties expressed as normalized standard deviations (SD) in the 2-pool microbial model (a, c) and 4-pool microbial model (b, d). Constrained Q10 for the conventional model was 1.86 with SD =4% of the mean, and was constant.

Figure 4.6 Global frequency distributions of total SOC pools produced by a calibrated CENTURY-type model [**a**, from Hararuk et al., (2014)], 2-pool and 4-pool microbial models (**c** and **d** respectively), and the cumulative SOC changes under RCP8.5 scenario (**b-d**, red lines are maximum likelihood cumulative changes, and gray lines are sample runs, representing uncertainty)

Figure 4.7 Spatial distribution of oscillations in the 2-pool (a), and 4-pool (b, c) models, and time required to damp the oscillations (d-f). 4-pool model had higher oscillation period to convergence time ratio than two pool model, which indicated that oscillations will be damped early in their evolution.

Figure S4.1 Comparison of CLM-CASA' soil C input with a 7-year average of MODIS data (Zhao & Running, 2010)

Figure S4.2 Differences in the magnitudes of the residuals between calibrated microbial models and conventional models in the environmental space. The circle diameters represent relative magnitudes of the residuals.

Figure S4.3 Distribution of soil C residence times in the environmental space simulated by a calibrated conventional model (a), and microbial models (b).

Figure S4.4 Change of temperature sensitivity of soil C residence time with respect to microbial carbon use efficiency and substrate limitation.

Figure S4.5 Relative change in SOC input in response to the RCP8.5 scenario simulated by CESM.

Abstract

In the future warming world terrestrial ecosystems may mitigate increasing temperatures by sequestering CO_2 from the atmosphere, or they can intensify future global change, amplifying the rate of CO_2 production in response to warming. Ecosystems' response to climate change depends on controls over carbon (C) influx and storage, with the latter determined by ecosystem pools' turnover rates. Global C cycle models perform well in predicting C influx rates, the gross and net primary productivity (GPP and NPP), however, their simulation of carbon storage requires improvement. This dissertation is focused on improving the models' performance in simulating carbon storage and turnover rates.

In the first chapter I describe the importance of understanding the controls over ecosystem carbon storage; give an overview of current global carbon cycle model performance in C storage simulation; and describe benefits of data assimilation for model improvement. In the second chapter I focus on improving the modeled turnover rates of the surface leaf litter. I first illustrate the poor prediction of surface leaf litter turnover rates by a commonly used first-order decay model, then use a global observed dataset of litter turnover rates and a Bayesian Markov Chain Monte Carlo (MCMC) approach to calibrate the model. After calibration the model explained 43% of spatial variability in the observed litter turnover rates, which was better than the initial 15%. After calibration the nature of the structural lignin limitation of litter turnover rates became unrealistic, therefore I altered litter quality limitation function to be dependent on litter lignin-to-nitrogen ratio. The change in the litter quality limitation assumption led to further increase in the explained variability in the observations to 61%, and the

estimated degree of lignin-to-nitrogen limitation of litter turnover rate was comparable to the values reported in literature. Lastly, model calibration resulted in reduction of temperature sensitivity of the litter turnover rates from Q10=2 to Q10=1.45.

In the third chapter I improve the simulation of soil organic carbon (SOC) storage in CLM-CASA'. Long-term land carbon-cycle feedback to climate change is largely determined by dynamics of SOC. However, most evaluation studies conducted so far indicate that global land models predict SOC poorly. I evaluated SOC predictions by CLM-CASA', investigated underlying causes of mismatches between model predictions and observations, and calibrated model parameters using Bayesian MCMC technique to improve the prediction of SOC. I compared modeled SOC to observed soil C pools provided by IGBP-DIS globally gridded data product and found that CLM-CASA' on average underestimated SOC pools by 65% (r^2=0.28). I applied data assimilation to CLM-CASA' to estimate SOC residence times, C partitioning coefficients among the pools, as well as temperature sensitivity of C decomposition. The model with calibrated parameters explained 41% of the global variability in the observed SOC, which was substantial improvement from the initial 27%. The projections differed between models with original and calibrated parameters: over 95 years the amount of C released from soils reduced by 48 Pg C, and the amount of C released from litter reduced by 6.5 Pg C. Thus, assimilating observed soil carbon data into the model improved fitness between modeled and observed SOC, and reduced the amount of C released under changing climate.

Despite calibration, CLM-CASA' still explained only 41% of variability in the observed SOC, and that led me to explore alternative assumptions about SOC dynamics.

CENTURY-type models (including CLM-CASA') represent microbial activity via a fraction of the substrate pool, modified by an environmental limitation function. Alternatively, microbial models simulate heterotrophic respiration as a function of microbial biomass, environmental, and substrate limitation. In the fourth chapter I calibrated two microbial model formulations (a two- and a four-pool model) to global total soil organic carbon and microbial biomass pools, and compared the models' performance to that of CLM-CASA'. Once calibrated, both microbial models explained 51% of variability in the observed soil carbon, which was 10% more than the amount explained by the calibrated CLM-CASA'. SOC in the microbial models was more sensitive to climate change than SOC in the CENTURY-type model: maximum likelihood magnitude of SOC decrease after 95 years of climate change was almost 5-fold higher in the microbial models than in CLM-CASA'. The uncertainties of SOC feedbacks to 95 years of climate change were also larger in the microbial models than in CLM-CASA', which was due to non-linear and oscillatory dynamics in the microbial models.

These studies showed that current models did not perform well in simulating carbon dynamics in the dead organic matter pools, however their performance could be improved after calibration against the global observed datasets. In addition, using a Bayesian MCMC technique for model calibration allowed to generate parameter uncertainties, which could be propagated in the model to generate data-informed uncertainties for the modeled pools and their feedbacks to global climate change. It is essential to continue the efforts of calibrating and validating various model formulations

using more globally observed datasets to identify models best representing reality and increase the confidence in the model projections.

Keywords: carbon cycle, data assimilation, global change, terrestrial ecosystems

Chapter 1

Introduction

Carbon dioxide (CO_2) is a greenhouse gas that is directly related to global temperature anomalies (Cox *et al.*, 2000). Since the beginning of industrial revolution anthropogenic CO_2 emissions have been rising at the rate proportional to countries' gross domestic product corrected for carbon intensity of energy (Raupach *et al.*, 2007). In the recent years global financial crisis has accelerated the rates of anthropogenic CO_2 emissions (Peters *et al.*, 2012), and these rates are not projected to diminish in the near future (Köne & Büke, 2010). Among the anthropogenic emissions CO_2 is responsible for the largest fraction of radiative forcing, and may increase mean global temperatures by 2.8-5.5°C before 2100 (Stocker *et al.*, 2013).

Both atmospheric CO_2 and temperatures affect terrestrial carbon (C) cycle. Elevated CO_2 stimulates ecosystem carbon storage (Luo *et al.*, 2006) by increasing photosynthetic C fixation rate (Luo & Mooney, 1995), and therefore mitigating the rate of increase in atmospheric CO_2 concentrations. Elevated temperatures increase the organic matter decomposition rates (Kätterer *et al.*, 1998), stimulating the natural CO_2 emissions. The opposite directions of ecosystem C storage response to elevated CO_2 and increasing temperatures illustrate that ecosystems may mitigate or worsen anthropogenically-driven climate change.

Current coupled Earth system models (ESMs) do not agree on whether terrestrial ecosystems will uptake or release CO_2 under climate change conditions, which causes large uncertainties in future temperatures and atmospheric CO_2 predictions (Friedlingstein *et al.*, 2006). One of the main causes of these uncertainties is non-uniform effect of temperature on organic matter decomposition rates across the ESMs (Friedlingstein *et al.*, 2006, Jones *et al.*, 2005, Jones *et al.*, 2003). For instance,

variability in parameters controlling temperature effect on organic matter turnover rates among 11 ESMs causes 6-fold variation in the soil carbon storage (Todd-Brown *et al.*, 2013b). Such spread in predictions of the largest terrestrial C pool highlights the lack of our understanding of the soil C dynamics and calls for improvements in the simulation of organic matter decomposition.

Data assimilation (DA) techniques allow estimating a set of model parameters that minimizes the error between modeled and observed data, thus, improving a model. Using the new – optimum – set of parameters in carbon cycle predictions increases models' reliability. Moreover, probabilistic DA approaches generate probability distributions for the optimum parameters, which can be used to produce the data-informed uncertainties in the model predictions – a rare feature in the global carbon cycle simulations. Lastly, data assimilation can help identify a flawed model formulation, which can be indicated by parameters that approach unrealistic values.

In the past two decades the carbon cycle research transitioned from a data-poor to a data-rich state. Globally observed data sets for ecosystem carbon input rates (Zhao & Running, 2010), soil carbon storage (Batjes, 2009, Batjes, 2014, Group, 2000), soil respiration rates (Bond-Lamberty & Thomson, 2012b), litter decomposition rates (Zhang *et al.*, 2008), and many more became available. Despite the data abundance, few studies focused on evaluation of the global carbon cycle models against global data [e.g. (Abramowitz *et al.*, 2008, Kucharik *et al.*, 2000, Randerson *et al.*, 2009a, Todd-Brown *et al.*, 2013b)] and even fewer studies took advantage of the global carbon data to improve the carbon cycle models [e.g. (Rayner *et al.*, 2005, Smith *et al.*, 2013, Ziehn *et al.*, 2011)].

The work in this dissertation addresses the uncertainties in organic matter decomposition simulated by global carbon cycle models. The studies in the dissertation are focused on improving the models' representation of the organic matter storage in litter and soils, as well as the controls over their turnover rates. In Chapter 2 I evaluate the performance of a commonly-used first-order decay model (Aber et al., 1990, Harmon et al., 2009, Olson, 1963) in simulating leaf litter turnover rates. Using probabilistic inversion I estimate model parameters, explore different assumptions about substrate quality limitation of leaf litter decomposition, and validate the best-performing model using global and regional observations. Finally, I illustrate how data assimilation changed the leaf litter feedbacks to a climate change scenario, and generate uncertainties for these feedbacks.

In Chapter 3 I evaluate the performance of the Community Land Model with Carnegie-Ames-Stanford Approach biogeochemistry sub-model (Oleson et al., 2004, Oleson et al., 2008, Parton et al., 1993), and improve it by assimilating a global observed soil organic carbon data. I evaluate the posterior parameter distributions by comparing them with the observations and checking whether parameters approach unrealistic values; evaluate the feedbacks of soil C to elevated temperatures and atmospheric CO_2 concentrations; and generate the uncertainties for the soil C feedbacks.

In Chapter 4 I explore a recently proposed soil C cycle model formulations (Allison *et al.*, 2010, German *et al.*, 2012) that simulate decomposition as a function of not only temperature, but also of microbial biomass. I calibrate the models against the observed total soil organic carbon and microbial biomass carbon; evaluate the

performance of the calibrated models; and compare the soil C climate change feedbacks and their uncertainties between microbial models and a calibrated model from Chapter 3.

Studies described in Chapter 2-4 improve our knowledge about global carbon cycle dynamics, illustrate the data-informed uncertainties in organic matter feedbacks to future climate change, and give insights into the type of observations needed to further improve the performance of the global C cycle models. It should be noted that Chapters 2-4 are developed for peer-review publication.

Chapter 2

Improvement of global litter turnover rate predictions using a Bayesian MCMC approach

Abstract

Global terrestrial carbon cycle has a strong influence on atmospheric CO_2 concentrations and temperatures. Litter turnover is a small, but important part of the global terrestrial carbon cycle as it is a critical stage in the soil organic matter formation and nutrient mineralization. Litter turnover rates have been observed on site, regional, and global levels, however little effort has been put into validating and calibrating litter decay models against the observations. This study was to evaluate predictions of leaf litter turnover rates by a commonly-used first order decay model with different assumptions about litter quality limitations on decomposition; investigate underlying causes of mismatches between model predictions and observations; and calibrate model parameters to improve its performance. Model with original parameters explained 15% of the variability in the observations and parameter calibration improved the explained variation to 44%. Assuming that litter decomposition was dependent on litter lignin-to-nitrogen ratio rather than litter structural lignin content improved the fraction of explained variability in observations to 62%. Litter C pool feedbacks to changing climate differed between original and best-fitting models: original model predicted a 15% decrease in the leaf litter pool after 95 years of climate change (2006-2100), whereas the best-fitting model predicted a 2% increase. Furthermore, assuming that litter quality decreased with increasing CO_2 concentrations resulted in original model predicting a 28% loss of leaf litter pool, and the best-fitting model predicting a 15% increase in litter pool. Thus, assimilating observed leaf litter turnover rates into a first-order decay model improved model fit and reversed the leaf litter pool feedbacks to the changing climate.

2.1 Introduction

Global carbon (C) cycle is tightly coupled with climate: climate regulates the ecosystem C storage capacity (Fung *et al.*, 2005, Xia *et al.*, 2013a), and carbon released from or sequestered by ecosystems has impact on climate (Falkowski *et al.*, 2000, Houghton *et al.*, 2001). Terrestrial ecosystems, in particular, have been shown to significantly affect temperature (Foley *et al.*, 2003), therefore it is important to accurately represent the feedbacks between terrestrial carbon cycle and climate. Improvements in prediction of these feedbacks will facilitate reliable assessments of the global change effects on the ecosystems as well as development of the mitigation strategies for these effects.

Global terrestrial C pool is estimated at around 2000 Pg C (Falkowski *et al.*, 2000), and although litter pool constitutes a small fraction of global terrestrial C pool [68-97 Pg C (Matthews, 1997)] litter decomposition is a critical stage in soil organic matter formation and nutrient mineralization (Austin & Ballaré, 2010). Multiple studies show that litter decomposition is controlled by climate (Gholz *et al.*, 2000, Hobbie, 1996, Hobbie *et al.*, 2000), initial litter lignin content or lignin to nitrogen (lignin:N) ratio (Melillo *et al.*, 1982, Shaw & Harte, 2001), and the origin of litter ["home-field advantage effect" (Ayres *et al.*, 2009, Gholz *et al.*, 2000)], however little effort has been put into calibrating those relationships against the observations to represent litter decomposition rates for various points on the globe.

With increase in the available ecological data, implementation of data-model fusion techniques for model improvement and uncertainty assessments have been gaining momentum (Luo *et al.*, 2009). Particularly, calibration of the litter

decomposition models was carried out on site and regional levels: Williams et al. (2005) used Ensemble Kalman Filter to calibrate litter decomposition (among other processes) against the observations in central Oregon; Keenan et al. (2012) and Xu et al. (2006) used Bayesian inversion to calibrate an ecosystem carbon cycle model for Harvard forest and Duke forest respectively; and Adair et al. (2008) calibrated several litter decomposition models with the observations from North and Central America.

Model calibration at the site level is useful for representing environmental effects on decomposition processes and their uncertainties for a particular set of environmental conditions, but it is unlikely that the obtained parameters will represent large-scale variability in the turnover rates. For the models to be suitable for use in large-scale simulations they have to be calibrated against regionally and globally distributed observations in order to capture the variability of decomposition across many locations across the globe [as in Adair et al. (2008)].

In this study we used leaf litter turnover rates observed across the globe to (1) calibrate first order decay model formulations commonly used to represent leaf litter decomposition; (2) evaluate the causes of mismatches between model estimates and observations; (3) evaluate the impacts of model calibration on predictive ability of litter pools; and (4) litter feedbacks to a climate change scenario along with the uncertainty of these feedbacks.

2.2 Methods

2.2.1 Litter decay rate

Litter mass loss is usually represented as an exponential decay process (Aber *et al.*, 1990, Harmon *et al.*, 2009, Olson, 1963):

$$X_t = X_o e^{-kt} \quad (2.1)$$

where X_t is the litter pool size at the time *t*, X_o is the initial litter pool size, and *k* is the decay rate. The decay rate is dependent on climate, and litter quality:

$$k = k_{base} \times f(T) \times f_n(Q) \quad (2.2)$$

where k_{base} is litter turnover rate under no climate or litter quality limitation, $f(T)$ is temperature limitation, and $f_n(Q)$ is litter quality limitation (n =1 or 2 depending on assumption about limitation). Temperature limitation is modeled as a Q10 function:

$$f(T) = Q_{10}^{(0.1\times(T-30))} \quad (2.3)$$

where Q_{10} is temperature sensitivity of heterotrophic respiration, and *T* is temperature. For litter quality limitation we used two assumptions: (1) litter quality limitation was determined by structural lignin carbon as in Parton et al. (1987); and (2) litter quality limitation was a function of lignin:N ratio in leaf litter as illustrated in Melillo et al. (1982), Stump and Binkley (1993), and Shaw and Harte (Shaw & Harte, 2001). Parton et al. (1987) modeled litter quality limitation as:

$$f_1(Q) = \exp(-a \times L_s) \quad (2.4)$$

where *a*=3, and L_s is fraction of lignin-C in organic matter and was calculated as:

$$L_s = \frac{0.65\times L}{0.45\times(1-F_L)} \quad (2.5)$$

where 0.65 was the approximated fraction of C in a lignin molecule, 0.45 was the C content in the surface leaf litter, F_L was the fraction of labile C in the surface leaf litter, and following Parton et al. (1987) was calculated as:

$$F_L = 0.85 - 0.018 \times LN \quad (2.6)$$

where LN was the lignin:N ratio of the surface leaf litter. We represented the second assumption in litter quality limitation as a power function of lignin:N ratio:

$$f_2(Q) = LN^{-b} \quad (2.7)$$

where *b* was an estimated parameter with initial value of 0, representing no litter quality limitation.

2.2.2 Observed data

We used the global database of leaf litter turnover rates (*k*'s)compiled by Zhang et al. (2008). The database provided data on temperature, precipitation, litter lignin content, and nitrogen content, which allowed us to simulate *k*'s at each given site. We randomly separated 141 globally distributed data points into two groups: for model calibration (n=79), and for model validation (n=64).The fit statistics in the results section will be provided for the model performance on validation dataset.

2.2.3 Parameter estimation

We calibrated k_{base}, Q_{10}, *a*, and *b* using Bayesian probabilistic inversion. Mosegaard and Sambridge (2002) summarize Bayesian inversion as

$$p(\mathrm{c}|\mathrm{Z}) = v_c \times \mathrm{p}(\mathrm{Z}|\mathrm{c}) \times \mathrm{p}(\mathrm{c}) \quad (2.8)$$

where p(c|Z) is posterior probability density function of model parameters c; p(Z|c) is a likelihood function of parameters c; p(c) is prior probability density function of parameters c; and v_c is a normalization constant. We assumed that the prediction errors were normally distributed and uncorrelated, and calculated the likelihood function, p(Z|c), as

$$p(Z|c) = v_L \times \exp\left\{-\sum_{i=1}^{k} \frac{(Z_i - X_i)^2}{2\sigma_i^2}\right\} \quad (2.9)$$

where Z_i is *k* reported in Zhang et al. (2008) at *i*th site, X_i is simulated *k* for the *i*th site; $\sigma^2{}_i$ is the associated with *i*th observation; *k* is the total number of sites (n=79); and v_L is a constant. In their database Zhang et al. did not report the uncertainties associated with litter turnover rates, therefore we followed the approach used in Harmon and Challenor (1997) and Hararuk et al. (2014), and assumed a standard deviation of 30% for each observation, which we then used to calculate the variance.

We assigned minimum and maximum values to the parameters and used adaptive Metropolis (AM) algorithm (Haario *et al.*, 2001) to sample from the posterior parameter distributions. We generated a parameter chain by running AM algorithm in two steps: a proposing step and a moving step. In the proposing step a new parameter set c^{new} was generated from a previously accepted parameter set c^{k-1} through a proposal distribution $(c^{new}|c^{k-1})$. In the moving step a probability of acceptance $P(c^{k-1}|c^{new})$ was calculated as in (Marshall *et al.*, 2004):

$$P(c^{k-1}\big|c^{new}) = min\left\{1, \frac{p(Z|c^{new})p(c^{new})}{p(Z|c^{k-1})p(c^{k-1})}\right\} \quad (2.10)$$

The value of $P(c^{k-1}|c^{new})$ was then compared with a random number *U* from 0 to 1. Parameter set c^{new} was accepted if $P(c^{k-1}|c^{new}) \geq U$, otherwise c^k was set to c^{k-1}.

The AM algorithm required an initial parameter covariance matrix, which we generated from a test run of 40,000 simulations with uniform proposal distribution as in Xu et al. (2006):

$$c^{new} = c^{k-1} + r \times \frac{c^{max} - c^{min}}{D} \quad (2.11)$$

where c^{max} and c^{min} are upper and lower parameter limits *r* is a random number between -0.5 to 0.5, and D=5. From the test run results we calculated the covariance matrix C_0 and modified the proposal step to be

$$c^{new} = N(c^{k-1}, C_k) \quad (2.12)$$

$$C_k = \begin{cases} C_0 & k \leq k_0 \\ s_d Cov(c_0, \dots, c_{k-1}) & k > k_0 \end{cases} \quad (2.13)$$

where k_0 = 2000; $s_d = 2.38/\sqrt{3}$ (Gelman *et al.*, 1996).

We made five parallel runs (each run containing 200,000 simulations) starting at dispersed initial points in the parameter space. We discarded the first half of the simulations (as burn-in phase) and tested the second half for convergence to stationary distributions with Gelman-Rubin diagnostics (Gelman & Rubin, 1992).

We assigned the boundaries to model parameters (listed in Table 2.1) based on the literature and our assumptions. We varied temperature sensitivity, Q_{10}, between 1 (to assume that turnover rates were insensitive to temperature changes) and 3, which was slightly higher than empirical values (Gholz *et al.*, 2000, Smyth *et al.*, 2009, Zhou *et al.*, 2008). Baseline litter turnover rate, k_{base}, was varied between 0.5 years^{-1} (2 years) to 24 years^{-1} (≈ 2 weeks). The lower boundary for parameter *a* was set to test whether model formulation would yield unrealistic values, and upper boundary was reported to produce better model performance (Kirschbaum & Paul, 2002). The upper

boundary for parameter *b* was set slightly higher than the value reported in Melillo et al.(1982) (*b*=0.78), and the value calculated from Shaw and Harte (2001) (*b*=0.88); as for the parameter *a*, the lower boundary for the parameter *b* was set to test the model formulation for unrealistic dynamics.

Table 2.1 Parameter characteristics. MLE is maximum likelihood estimate, and G-R is the result of Gelman-Rubin chain convergence diagnostics

Parameter description	Symbol	Prior			MLE	G-R	Lower 95% bounds	Upper 95% bounds
		minimum	maximum	Value				
Model with $f_1(Q)$								
Baseline turnover rate, $years^{-1}$	k_{base}	0.50	24.00	7.00	0.93	1.00	0.87	1.03
Temperature sensitivity	Q_{10}	1.00	3.00	2.00	1.77	1.00	1.72	1.87
Litter quality limitation parameter	a	-5.00	5.00	3.00	-4.24	1.00	-4.88	-1.34
Model with $f_2(Q)$								
Baseline turnover rate, $years^{-1}$	k_{base}	0.50	24.00	7.00	2.37	1.00	2.18	2.73
Temperature sensitivity	Q_{10}	1.00	3.00	2.00	1.43	1.00	1.38	1.52
Litter quality limitation parameter	b	-1.00	1.00	0	0.36	1.00	0.34	0.40

2.2.4 Leaf litter feedbacks to climate change

We evaluated the uncertainties in surface litter feedbacks to climate change by running the best-performing calibrated model forward, driving it with a climate change scenario (increasing CO_2 and temperatures) and samples from the posterior parameter distributions. We used the Community Earth System Model (CESM) output for the Representative Concentration Pathway 8.5 (RCP8.5) experiment (specifically, the simulated temperature and C influx to leaves, which we assumed to be similar to leaf litter flux) to drive the leaf litter dynamics. The CESM model output was provided as a part of Coupled Model Intercomparison Project Phase 5 (CMIP5), and was available at http://pcmdi9.llnl.gov. Over 95 years CESM simulated a 3.5 K increase in mean global

temperature and atmospheric CO_2 increase to 1150 ppm by the year 2100 (Keppel-Aleks *et al.*, 2013). We first used global litter lignin content and CN ratios from the CESM model and the 2006-2010 temperature and C influx data to generate initial leaf litter pools using the semi-analytical model spin-up approach (Xia *et al.*, 2012):

$$X_{Litter} = \frac{NPP_L}{k} \quad (2.14)$$

where $X_{Litter}$was leaf litter pool, g/m², and NPP_L was C influx to leaves, g/m²/year. We then ran the litter dynamics model forward in time to the year 2100, generating litter feedbacks to the changing climate.

Elevated CO_2 has been reported to increase leaf litter lignin and decrease leaf litter nitrogen content (Liu *et al.*, 2005, Norby *et al.*, 2001) at a rate of 6.5% per ~300 ppm and 7.1% per ~300 ppm of increasing CO_2 respectively (Norby *et al.*, 2001). We assumed the change in litter chemistry was linear and monotonic, and applied the rates of change for lignin and nitrogen content to calculate new values for leaf litter quality limitation at each time step of a forward model run.

2.3 Results and Discussion

2.3.1 Estimated parameters

Most estimated parameters were well constrained within their prior ranges (Fig. 2.1). Litter quality limitation parameter *a* from eq. 2.4, however, was skewed against its minimum value, reversing the effect of lignin on litter decomposition: under optimum parameter values increase in litter lignin increased its turnover rate exponentially. Such effect of lignin on decomposition is unrealistic, therefore litter quality limitation function presented in eq. 2.4 does not reflect the observed patterns. Parameter *b* from

the eq. 2.7, on the other hand, was constrained within a realistic range with the maximum likelihood value of 0.36 and a 95% confidence interval (CI) of 0.34-0.40, yielding a negative relationship between litter quality and decomposition rate.

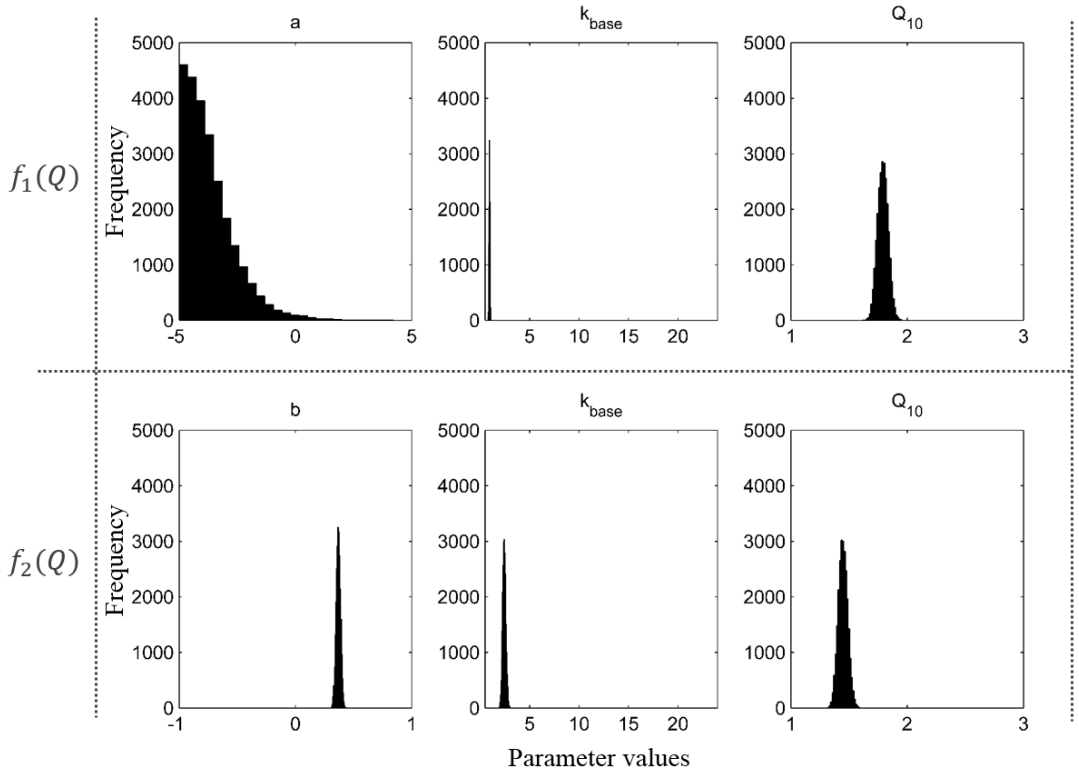


Figure 2.1 Posterior parameter distributions for two models with different assumptions of litter quality limitation of its turnover rate: $\boldsymbol{f_1(Q) = \exp(-a \times L_s)}$, and $\boldsymbol{f_2(Q) = LN^{-b}}$. Baseline leaf litter turnover rate was higher under the $\boldsymbol{f_2(Q)}$ assumption, and leaf litter turnover was less sensitive to temperature under $\boldsymbol{f_2(Q)}$ assumption than $\boldsymbol{f_1(Q)}$ assumption.

The maximum likelihood lignin:N effect on litter decomposition was weaker than the one reported in Melillo [(1982), *b*=0.78], and weaker than the ones calculated from the data in Shaw and Harte [(2001), *b*=0.88], Taylor et al. [(1989), *b*=0.41], and

Wieder et al. [(2009), *b*=0.54]. A possible explanation for such spread in litter quality effect on decomposition rate could be caused by differences in the microbial communities. Decomposition of the passive litter fraction is associated with higher fungi-to-bacteria ratio in the decomposer community implying that fungi decompose passive litter fraction better than bacteria (Beare *et al.*, 1992). Additionally, complex decomposer communities have been shown to increase turnover rate of the passive litter fraction (Coûteaux *et al.*, 1991, O'Neill & Norby, 1996), therefore sites with low values of *b* may have higher fungal biomass or more complex decomposer communities than the sites with lower values of *b*. The studies with observed values of *b*, however, did not have the data on microbial community composition, therefore the reason for such variety in the observed effects of lignin:N ratio on decomposition requires further investigation.

Baseline litter residence time, k_{base}, was lower in the model with litter quality limitation function $f_1(Q)$, than in the model with the function $f_2(Q)$ (Table 2.1), and both estimates were higher than the value reported in Adair et al. [(2008), k_{base}= 0.53]. One-pool litter decomposition model in Adair et al. (2008) did not include litter quality effect on decomposition, which was likely the reason for low k_{base} value as it was implicitly corrected for the litter quality effect. Similarly, there was no agreement in temperature sensitivities (Q_{10}) between the two models, however both estimates fell within the wide range of the values reported in the literature [from 1.17 to 2.7 (Gholz *et al.*, 2000, Smyth *et al.*, 2009, Wang *et al.*, 2012a, Zhou *et al.*, 2008)].

2.3.2 Calibrated model performance

Original model formulation for leaf litter turnover, *k*, explained 15% of variance in the observed *k*'s from Zhang et al. (2008) with the root-mean-square error (RMSE) of prediction equal to 0.39 (Fig. 2.2a). Calibration of the model with the original litter quality limitation function improved its performance, increasing the r^2 to 0.44, and decreasing RMSE by 41% (Fig. 2.2b). Changing the litter quality limitation function increased r^2 to 0.62, and reduced RMSE by 49% (Fig. 2.2c). Convergence of the parameter *a* to an unrealistic value and better fit statistics for the model with litter quality limitation $f_2(Q)$ than for the model with $f_1(Q)$ led us to the conclusion that litter lignin:N ratio was a better predictor of litter quality limitation of decomposition than litter structural lignin content.

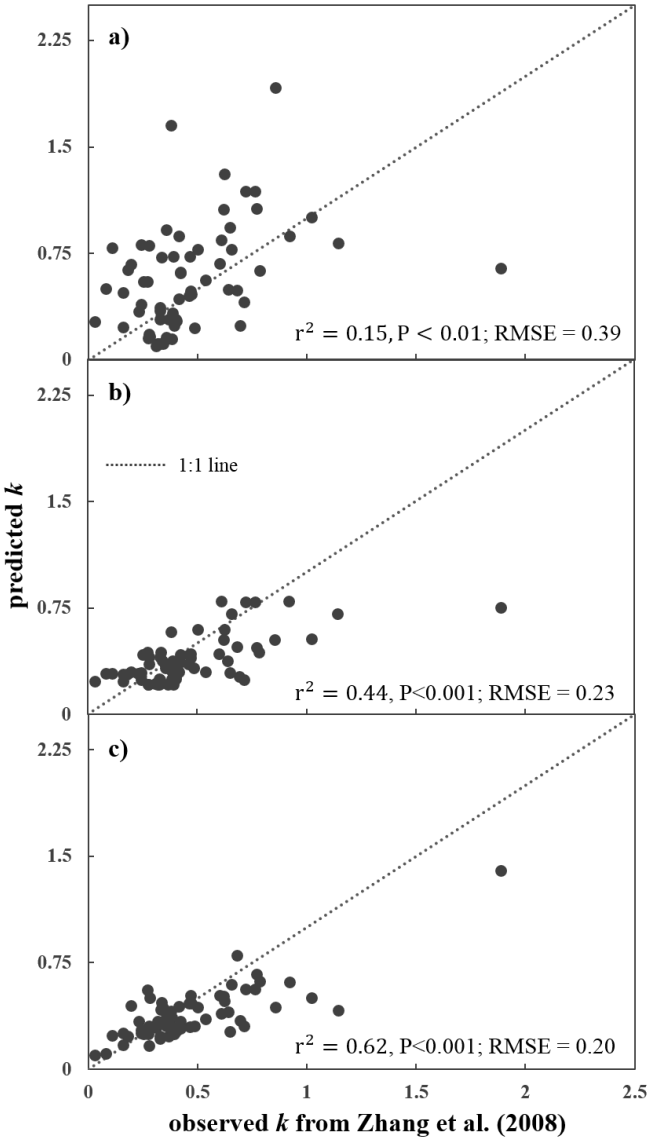


Figure 2.2 Comparison of the leaf litter turnover rates produced by CENTURY (a), calibrated CENTURY (b), and the model with different assumption about litter quality limitation on its turnover rate (c). Calibration improved performance of CENTURY, however, changing the assumption about litter quality limitation of its turnover rate improved model performance.

2.3.3 Global leaf litter distribution

We calculated the global distribution of the surface leaf litter pools as in eq. 2.14 using CESM leaf litter flux and the best fitting model for the leaf litter turnover rates (eqs. 2.2, 2.7). The best-fitting model predicted smaller litter pools than the original model in all regions except the tropics (Fig. 2.3), where the calibrated model predicted higher C storage compared to the original model. Comparison of our aboveground litter estimates to the ones provided in Vogt et al. (1986) revealed that calibration of the turnover rates did not improve model's predictive ability for litter pools in the low-temperature regions (Fig. 2.3c,d). Since C pools were determined by C influx rates and C pool turnover rates (Xia *et al.*, 2013a) the mismatches between the modeled litter pool estimates and the observations were caused either by errors in the litterfall or turnover rate predictions. Comparison of modeled and observed litterfall (Fig. 2.3e) revealed that most modeled estimates were within the range of the observed estimates with the largest mismatches (underpredictions) located in the tropical regions. Because there was general agreement between observed and modeled litter input estimates, the errors in modeled litter pools were caused by the errors in the litter turnover rates.

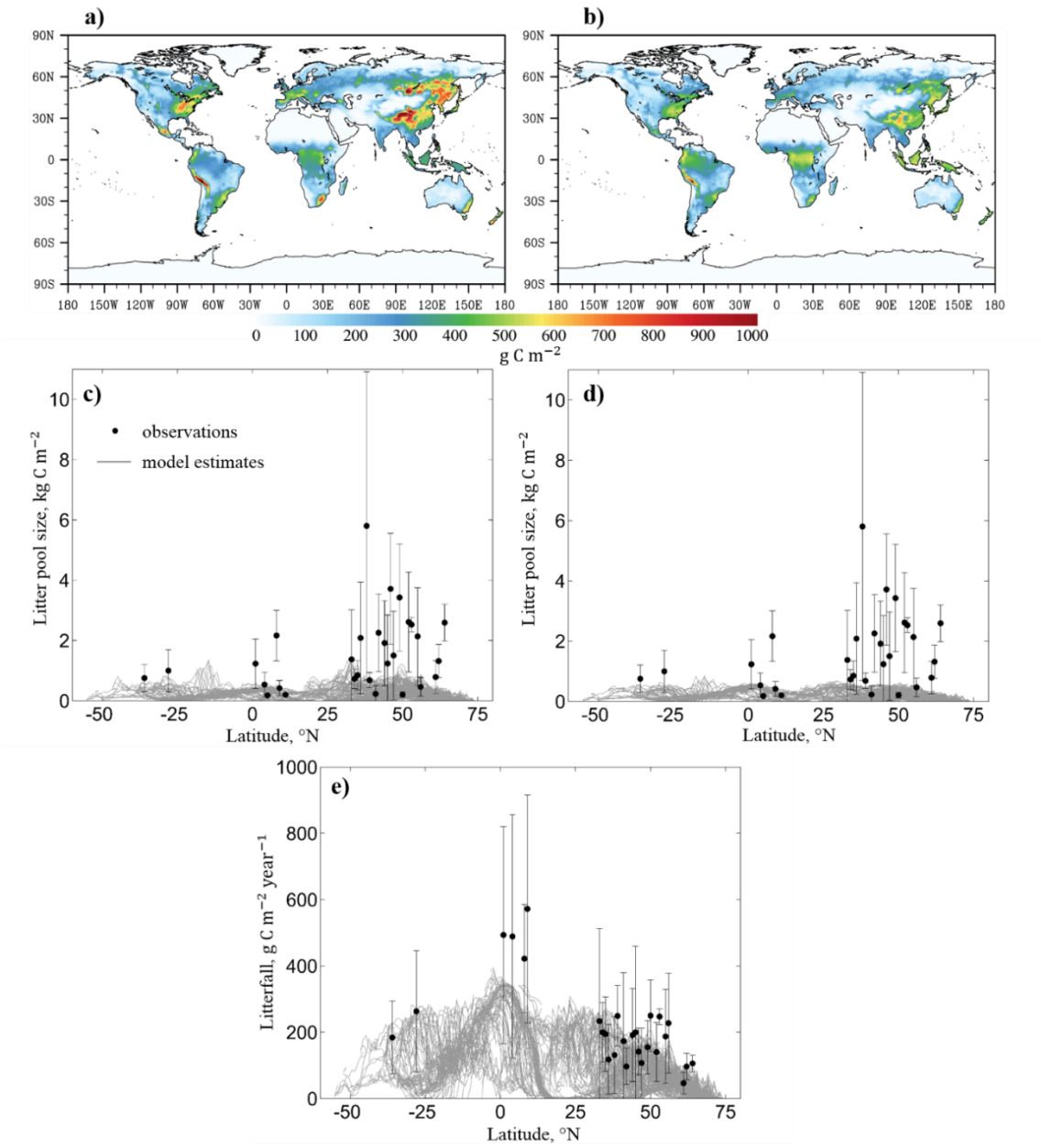


Figure 2.3 Global distribution of aboveground litter (excluding woody debris) before (a, c) and after (b, d) calibration of the turnover rates, and their comparisons to the observations from Vogt et al. (1986) (error bars represent standard deviations calculated for the observations per given latitude). Litter pool sizes are determined not only by litter turnover rates, but also by the input rate, therefore we compared modeled litter input rates to the observations (e) reported in Vogt et al. (1986).

Errors in the modeled turnover rate estimates could be caused by errors in the model input data, such as inaccurate temperature and global lignin:N ratio distributions assumed in the CESM, or the issues with the data that were used to calibrate the model. CESM assessment showed that model simulated land surface temperatures well (Lawrence *et al.*, 2011b), however there was no global observed lignin:N ratio data product to validate the CESM lignin:N distribution. Therefore lignin:N ratios remained a potentially large source of uncertainty for the global litter turnover rates prediction. To address a potential issue with the data used for model calibration we used an additional dataset for validation of our best-performing model. The dataset contained leaf litter turnover rates from the Long-term Intersite Decomposition Experiment Team (LIDET) (Harmon *et al.*, 2009) observed across multiple biomes and substrates. We calculated the leaf litter turnover rates from the temperatures and lignin:N ratios at the 27 LIDET sites distributed across North and Central America using the original and our best-fitting model, and compared them to the turnover rates from Harmon et al. (2009).

The calibrated model performed better than the original model (Fig. 2.4a,b), however it overpredicted low turnover rates, underpredicted high turnover rates, and overall had a much lower predictive ability compared to the data from Zhang et al. (2008) (Fig. 2.2c). Most of the turnover rates from the LIDET data were obtained from 10-year decomposition records, whereas the maximum length of decomposition records included in the Zhang et al. (2008) dataset was three years. Some studies argue that leaf litter turnover rates are best represented by two or three turnover rate components: fast, slow, and passive (Adair *et al.*, 2008, Harmon *et al.*, 2009) with the impact of slower components on the total leaf litter turnover rates dependent on litter quality (Adair *et al.*,

2008). Comparison of the LIDET turnover rates with our calibrated model estimates revealed no significant relationships neither between the model's residuals and lignin:N ratios (Fig. 2.4c) nor between the residuals and litter lignin content (Fig. 2.4d), therefore the model errors were not caused by the absence of explicitly modeled slow pools' turnover rates.

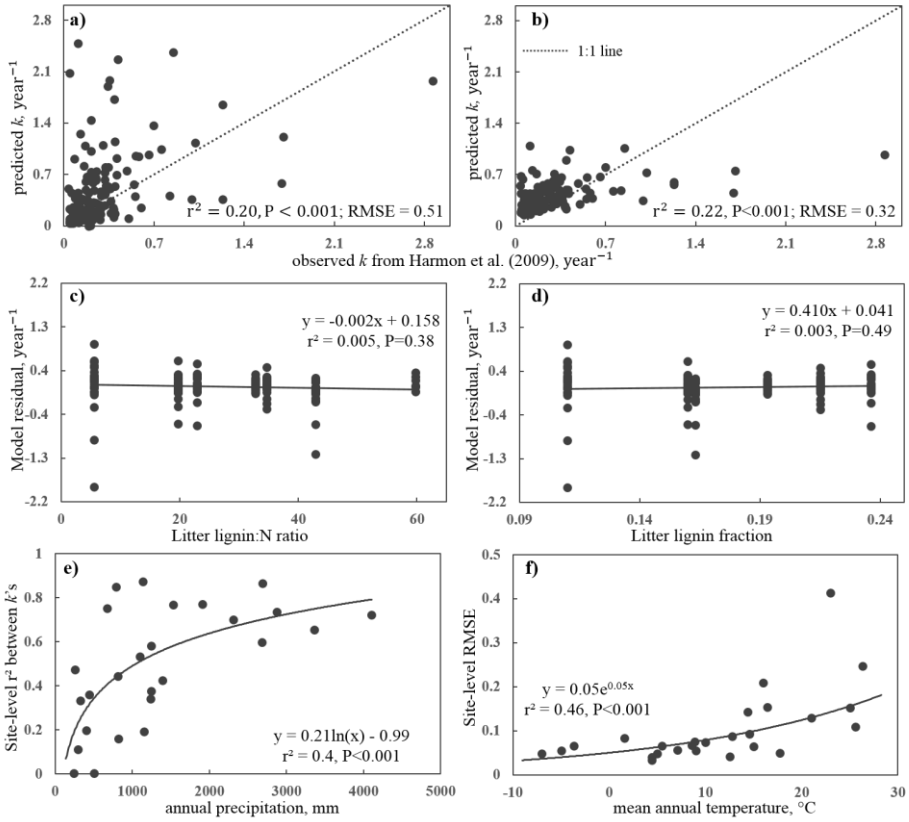


Figure 2.4 Comparison of modeled turnover rates to the observed turnover rates from Harmon et al. (2009) before (a) and after (b) calibration. The explained variability of the observations was correlated with annual precipitation (e), site-level RMSE's were correlated with mean annual temperatures (f), and there were no significant relationships between lignin:N ratio and model residuals (c) or between model residuals and litter lignin content (d).

At the site level fraction of the explained variance in the observed turnover rates was dependent on annual precipitation: model had higher predictive ability in humid climate (Fig. 2.4c). The predictability of the leaf litter turnover rates was also dependent on temperature: site-level RMSE's increased with increasing mean annual temperatures (Fig. 2.4d). Dependency of the turnover rate predictability on temperature and precipitation may point at the need to explicitly model microbial biomass dynamics in the litter pool. Although absent in our model formulation, precipitation affects organic matter decomposition, and its effect is represented better by microbial models than by the first order decay models in the arid regions (Lawrence *et al.*, 2009). Additionally, LIDET experiment presented turnover rates for litter transplants, therefore decrease in turnover rate predictability with increasing temperatures might be due to maladaptation of the microbial communities at LIDET sites to the foreign substrates (Gholz *et al.*, 2000).

2.3.4 Litter feedbacks to climate change

Global leaf litter pool size simulated by the best performing model was 26.3 Pg C with a 95% CI of 25.6-27.5 Pg C, which was lower than the original model estimate of 29.3 Pg C (Fig. 2.5a). The calibrated estimate was higher than the observed global leaf litter pool [13 Pg C (Matthews, 1997)], but lower than the previously reported model estimates [60 Pg C (Esser *et al.*, 1982) and 51 Pg C (Potter *et al.*, 1993)].

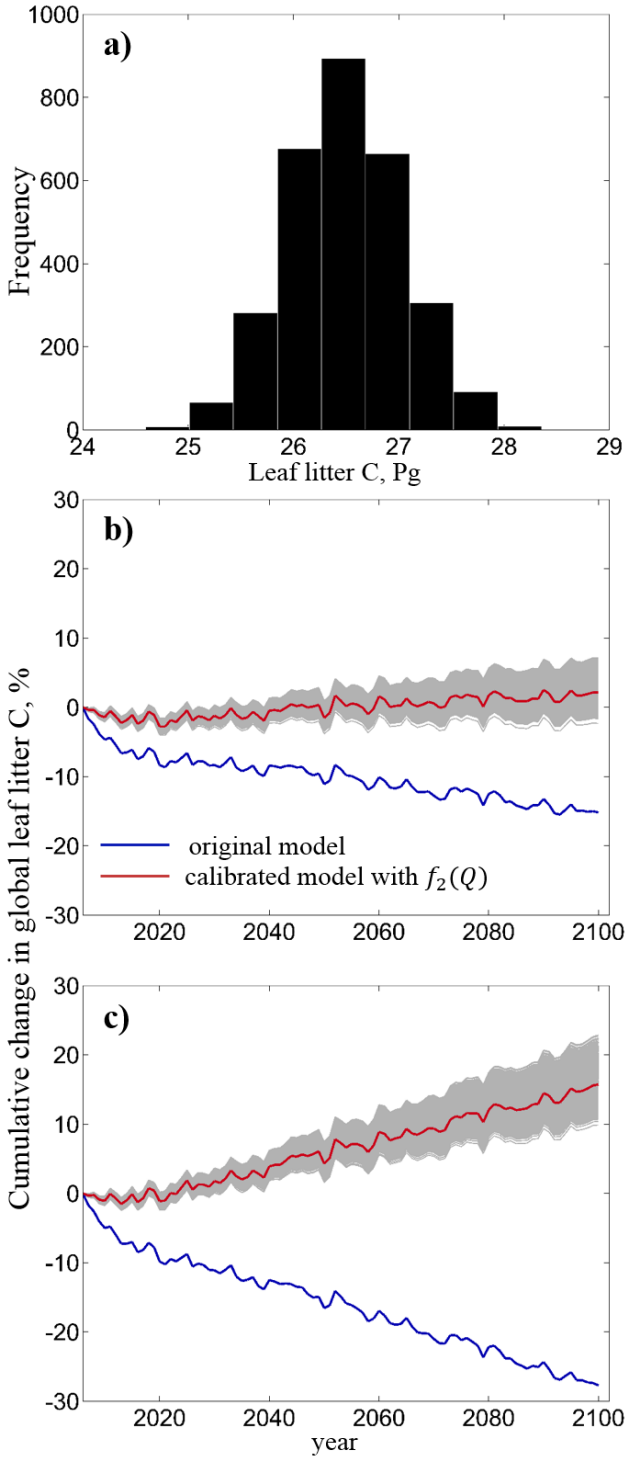


Figure 2.5 Global leaf litter pool (a), and its response to changing climate with (c) and without (b) the assumption of changing litter quality with increasing CO_2.

Changing litter quality limitation function along with model calibration resulted in the opposite direction of litter pool change in response to 95 years of increasing temperatures and CO_2 concentrations. Original model predicted a 16% decrease in leaf litter pool, whereas the calibrated model simulated a 2% increase (with the range of 2% decrease to an 8% increase) in the leaf litter pool after 95 years of climate change (Fig. 2.5b). Including the assumption about decreasing litter quality with increasing CO_2 resulted in a 28% decrease in the leaf litter pool in the original model after 95 years of climate change, and a 16% increase (with the range of 10% to an 23% increase) in the leaf litter pool in the calibrated model (Fig. 2.5c). The counter-intuitive response of the leaf litter pool to decreasing litter quality in the original model was due to a larger rate of change in the numerator of the equation 2.5 than in its denominator. Overall, unlike the original model, the calibrated model simulated a negative feedback of leaf litter to changing climate, which was amplified by CO_2-induced decrease in litter quality.

Increasing atmospheric CO_2 concentration leads to reduction of litter quality, however its effect on litter decomposition remains controversial (Norby *et al.*, 2001). Some experiments revealed that CO_2-induced decrease in litter quality decreased litter turnover rates and caused litter mass accumulation (Cotrufo & Ineson, 1996, Cotrufo *et al.*, 1994), others showed that decrease in litter quality might increase or have no effect on litter decomposition (Coûteaux *et al.*, 1999, Finzi & Schlesinger, 2002). Additionally, there is evidence that elevated CO_2 increases fungi-to-bacteria ratio (Carney *et al.*, 2007), which may offset the negative effect of increased litter quality on its turnover rate as fungi decompose low-quality litter better than bacteria (Beare et al., 1992). The varying reports of CO_2 effect on litter decomposition along with the

evidence of significant regulation of decomposition by microbial community composition shifts that current models do not account for, indicates the need for adjusting the first-order decay models to account for microbial community dynamics.

2.4 Conclusion

We calibrated a leaf litter decomposition model against the global observed leaf litter turnover rate dataset using a Bayesian MCMC approach. Changing the assumption about litter quality limitation of litter decomposition along with parameter calibration in the first-order decay model substantially improved the fitness between observed and modeled leaf litter turnover rates. When propagated in time under RCP8.5 scenario posterior parameter uncertainties resulted in a 0.26 Pg C loss to a 1.99 Pg C gain in leaf litter pool after 95 years of climate change. Including the change in litter quality adjustment to increasing atmospheric CO_2 increased the amount of carbon that remained in the leaf litter pool to a range of 2.5 Pg C to 6.1 Pg C.

As we show in this study, data assimilation is a useful tool for improving and gaining insights into carbon cycle modeling and uncertainty analysis of the model projections. The inability of our constrained model to accurately represent surface litter pools despite plausible representation of litter input indicates the potential flaws in global leaf litter quality data used in CESM, and/or importance of microbial community composition on leaf litter decomposition. More data is needed on global leaf litter quality to develop a reliable model input dataset. Also, we need more data on climate effects on microbial community compositions around the globe, and whether the

changes in the microbial composition result in the turnover rates significantly different from those predicted by the best-fitting first-order decay model.

Chapter 3

Evaluation and improvement of a global land model against soil carbon data using a Bayesian MCMC method[1]

[1] This chapter was published in Journal of Geophysical Research – Biogeosciences doi: 10.1002/2013JG002535

Abstract:

Long-term land carbon-cycle feedback to climate change is largely determined by dynamics of soil organic carbon (SOC). However, most evaluation studies conducted so far indicate that global land models predict SOC poorly. This study was to evaluate predictions of SOC by the Community Land Model with Carnegie-Ames-Stanford Approach biogeochemistry sub-model (CLM-CASA'), investigate underlying causes of mismatches between model predictions and observations, and calibrate model parameters to improve the prediction of SOC. We compared modeled SOC to the SOC pools provided by IGBP-DIS globally gridded data product and found that CLM-CASA' on average underestimated SOC pools by 65% (r^2=0.28). We extracted the C cycle component from CLM-CASA' and applied data assimilation to it to estimate SOC residence times, C partitioning coefficients among the pools, as well as temperature sensitivity of C decomposition. The model with calibrated parameters explained 41% of the global variability in the observed SOC, which was substantial improvement from the initial 27%. The SOC and litter C feedbacks to changing climate differed between models with original and calibrated parameters: after 95 years of climate change (2006-2100) the amount of C released from soils was 48 Pg C lower in the calibrated than in the non-calibrated model, and the amount of C released from litter was 6.5 Pg C lower in the calibrated than the non-calibrated model. Thus, assimilating estimated soil carbon data into the model improved model parameterization and reduced the amount of C released under changing climate. To further reduce the uncertainty in the soil carbon prediction, we need to explore alternative model structures, improve representation of ecosystems, and develop additional global datasets for constraining model parameters.

3.1 Introduction

Soils contain a large fraction of ecosystem carbon (C) (House *et al.*, 2002), which can be affected by climate change (Kirschbaum, 1995). Accurate prediction of soil organic C (SOC) content is important, as emission of CO_2 from soils greatly depends on the amount of carbon stored in it (Luo & Zhou, 2006). Being a greenhouse gas, naturally and anthropogenically emitted CO_2 likely leads to climate warming (Houghton *et al.*, 2001), which may further stimulate net release of soil carbon, forming a positive feedback loop between carbon cycle and climate warming (Friedlingstein *et al.*, 2006, Luo, 2007, Melillo *et al.*, 2002, Oechel *et al.*, 2000, Rustad *et al.*, 2001). Analysis of the output generated by 11 earth system models (ESMs) participating in the 5th Coupled Model Intercomparison Project (CMIP5, (Taylor *et al.*, 2011)) illustrates the uncertainty in the simulated SOC, which, despite the similarities in model structures, varies 6 fold among the models with the estimates ranging from 510 to 3040 Pg C (Todd-Brown *et al.*, 2013b). Only six out of 11 model estimates were within the range of the Harmonized World Soil Database (HWSD) estimate of 1260 Pg C (with a 95% confidence interval (CI) of 890–1660 Pg C) and none of the models' correlation coefficients for the grid-based comparisons between modeled and empirical SOC estimates exceeded 0.4 (Todd-Brown *et al.*, 2013b).

A few other studies that evaluated ESM's simulation of soil C stocks and dynamics indicated great differences between simulated and observed SOC stocks. Kucharik et al. (2000) evaluated the SOC stocks modeled by the Integrated Biosphere Simulator (IBIS) to be 1408 Pg C, which was lower than the estimate generated by International Geosphere-Biosphere Programme - Data and Information System (IGBP-

DIS) (1567 Pg, (Group, 2000)). IBIS underpredicted soil C stocks by up to 8 kg C/m² in deserts, open shrublands, polar regions, and tropical forests; and overpredicted soil C stocks by up to 11 kg C/m² in tundra regions. The second version of the IBIS model did not yield a better soil C fit (Delire *et al.*, 2003). Thum et al. (2011) evaluated SOC in two global models: CBALANCE-JSBACH-ECHAM5-Model (Cba-JEM (Raddatz *et al.*, 2007, Roeckner *et al.*, 2003)) and Yasso07-JSBACH-ECHAM5-Model (Y07-JEM (Liski *et al.*, 2005, Raddatz *et al.*, 2007, Roeckner *et al.*, 2003)). The total global SOC simulated by Cba-JEM was 2853 Pg C and was higher than IGBP-DIS and HWSD estimates, whereas Y07-JEM simulated 1477 Pg C, which was lower than IGBP-DIS estimate, but within the range of the HWSD estimates (Group, 2000, Todd-Brown *et al.*, 2013b). Both models overpredicted soil C in the northern conifer forests, tropical forests, grasslands, and savannas and underpredicted soil C in taiga regions. These differences in simulated SOC pools are likely propagated in the ESMs and cause substantial uncertainties in modeled carbon-climate feedback (Friedlingstein *et al.*, 2006).

ESM structures for simulating terrestrial carbon cycle are highly similar: they partition the photosynthetically fixed C among the live and dead carbon pools, from which C escapes via autotrophic or heterotrophic respiration (Todd-Brown *et al.*, 2013b, Weng & Luo, 2011). Todd-Brown et al. (2013b) showed that the differences in simulated SOC stocks among the models are mainly caused by their parameterizations. Parameter calibration is a commonly practiced procedure in model development, and parts of global land models have been calibrated with the site level data (Kuppel *et al.*, 2012, Wang *et al.*, 2007). Calibration of land models against global datasets has not

been widely implemented largely due to computational cost and methodological difficulty. For example, because of the lengthy spin-up, calibration of ESM parameters associated with soil carbon pools is computationally costly (Rayner *et al.*, 2005). Methodological difficulty arises from the fact that soil C content is a convoluted product of net primary productivity (NPP), C partitioning coefficient from plant litter to soil C pools, and SOC residence time (Luo *et al.*, 2003, Xia *et al.*, 2013a, Zhou & Luo, 2008). Even with the fixed NPP, same SOC content can be achieved by many combinations of C partitioning coefficients from plant litter to soil pools and SOC residence times (see Figure S3.1). In a model with three soil C pools and nine other C pools (plant and litter) that contribute to the soil pools it is difficult to identify the right set of parameters to adjust so as to make the simulated SOC stocks match closely to the observed ones.

Data assimilation techniques for parameter estimation have been successfully applied in the carbon cycle research at ecosystem, regional and global scales. Wu et al. (2009) demonstrated significant improvement in modeled net ecosystem exchange (NEE) after applying conditional inversion method to the flux-based ecosystem model (FBEM) to integrate observed NEE data into FBEM at Harvard forest. At the same site, Keenan et al. (2012) assimilated 15 carbon pool and flux datasets into Forest Biomass, Assimilation, Allocation and Respiration (FoBAAR) model to constrain 42 model parameters using adaptive multiple constraints Markov Chain Monte Carlo optimization, and was able to significantly reduce the forecast uncertainty of the C pool dynamics. Xu et al. (2006) used Bayesian Markov Chain Monte Carlo technique to assimilate six carbon datasets from Duke Forest into the Terrestrial ECOsystem

(TECO) model to constrain C residence times and partitioning coefficients, producing good prediction of soil respiration, foliage biomass, and woody biomass.

On the regional scale, Zhou and Luo (2008) successfully constrained ecosystem residence times against regional data of biomass, net primary production (NPP), litter and soil C across the continental USA by using genetic algorithm to find the best parameters. Zhou et al. (2009) and Ise and Moorcroft (2006) presented global constrained SOC temperature sensitivities resulting from assimilating global soil C data (Group, 2000) into C cycle models. Smith et al. (2013) developed a fully data-informed C cycle model and generated C cycle projections with the parameter and structural uncertainties propagated in time.

Despite the obvious advantages to data assimilation for calibrating model parameters, not much research has been carried out with global land models due to their complexity. Meanwhile, the modeling community still focuses on including more processes into ESMs to represent the C cycle as realistically as possible (Koven *et al.*, 2009, Koven *et al.*, 2011, Lawrence *et al.*, 2011a, Luo *et al.*, 2009, Oleson *et al.*, 2008). It is equally important to develop the capacity to calibrate models to improve their predictive ability. In this study we take advantage of Bayesian MCMC to (1) calibrate the C cycle component of CLM-CASA'; (2) identify the regions with the highest uncertainty in SOC; (3) gain insight about the model structure through maximum likelihood parameter estimates and their correlations; (3) generate data-constrained estimates of C residence time; and (4) generate uncertainties of SOC change under climate change.

3.2 Methods

3.2.1 CLM-CASA' and its C-only version for data assimilation

In this study we calibrated Community Land Model coupled with Carnegie-Ames-Stanford Approach biogeochemistry sub-model (CLM-CASA') (Oleson *et al.*, 2004, Oleson *et al.*, 2008, Parton *et al.*, 1993). CLM-CASA' consists of biogeophysics and biogeochemistry sub-models. Biogeophysics sub-model simulates solar and longwave radiation dynamics in vegetation canopy and soil; momentum and turbulent fluxes from canopy and soil; heat transfer in soil and snow layers; hydrology of canopy, soil, and snow; stomatal physiology, and photosynthesis (more details can be found in (Oleson *et al.*, 2004, Oleson *et al.*, 2008)). Biogeochemistry sub-model simulates carbon transfer among various plant, litter, and soil pools (Fig. 3.1) (Parton *et al.*, 1993). Change in carbon content in each pool is determined by a balance between influx into and efflux from the pool. Carbon influx into the ecosystem is defined by NPP, which is partitioned among three live biomass pools. Carbon efflux is heterotrophic respiration (autotrophic respiration in CLM-CASA' is assumed to be 50% of gross primary productivity, GPP) as determined by decomposition rate of organic C in each pool. Heterotrophic respiration is influenced by environmental conditions (specifically, temperature and soil moisture) as well as soil texture, tissue lignin, and tissue available nitrogen content.

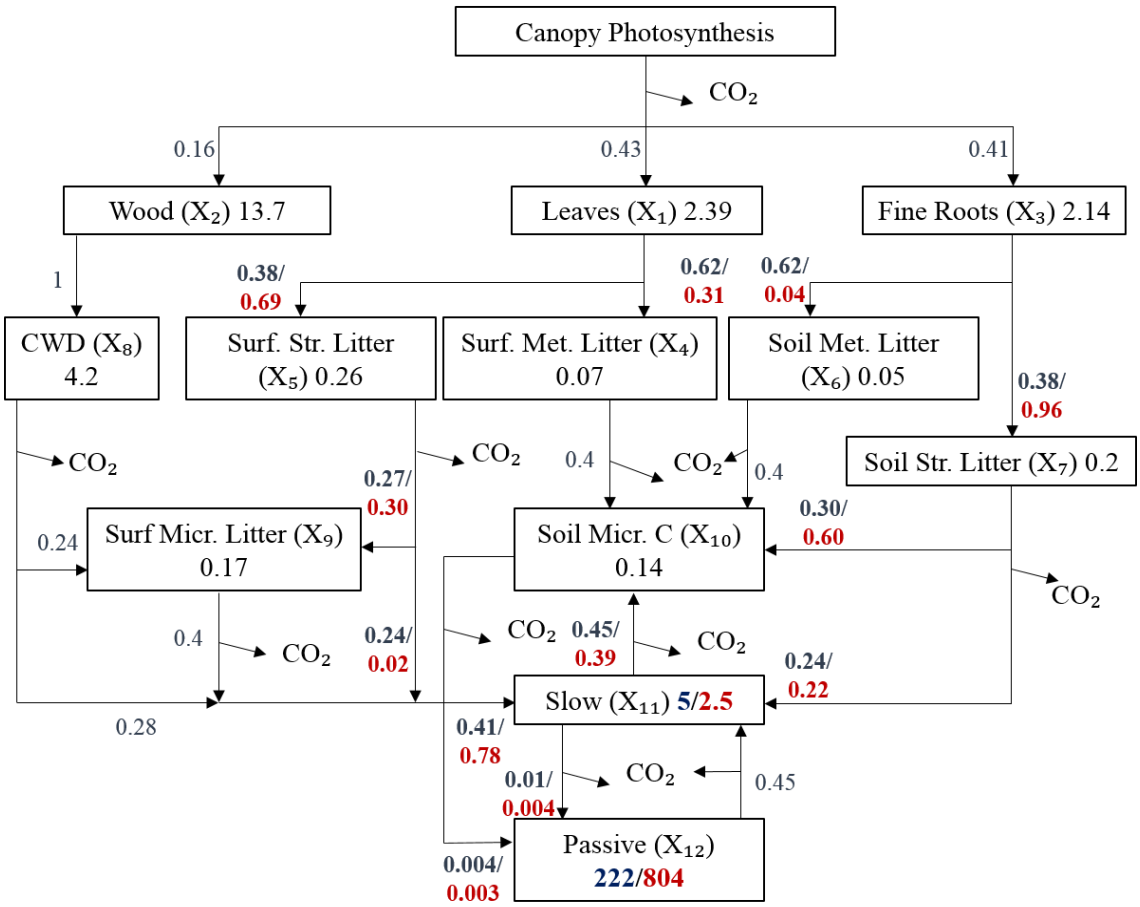


Figure 3.1 CLM-CASA' model structure and global parameter averages. Carbon enters the system through photosynthesis and is partitioned among three live pools. From the live pools carbon is transferred to the five litter pools, and from the litter pools it is transferred to the three soil pools. Values in the boxes are pool residence times; values outside the boxes are partitioning coefficients (values in blue are initial model values, values in red are results of parameter optimization). After data assimilation soil passive C residence time increased by 36 years, whereas slow C residence time decreased by 2.1 years; compared to initial values, more C was transferred to soil C pools from soil litter and less C is transferred to soil C from surface litter.

Based on the theoretical analysis, carbon cycle components of any land models can be represented by a matrix equation (i.e., a set of carbon input-output equations) (Luo & Weng, 2011, Luo *et al.*, 2003, Xia *et al.*, 2013a) as:

$$\frac{d\boldsymbol{X}(t)}{dt} = \mathbf{A}\boldsymbol{\xi}(t)\mathbf{C}\boldsymbol{X}(t) + \boldsymbol{B}U(t) \quad (3.1)$$

where $\frac{d\boldsymbol{X}(t)}{dt}$ is the change in C pools at each time step; $\boldsymbol{\xi}(t)$ is the diagonal matrix of environmental scalars, which represents the influence of climate on mortality or decomposition rate of organic C in each pool in a gridcell (the elements of $\boldsymbol{\xi}(t)$ for all non-live pools are the same) ; $\mathbf{A}$ is a matrix of partitioning coefficients among C pools, which is influenced by soil texture, lignin, and tissue lignin to nitrogen ratio; $\mathbf{C}$ is the diagonal matrix of baseline C exit rates at a given temperature (25°C in CLM-CASA'); $\boldsymbol{X}(t)$ is a vector of C pools; $\boldsymbol{B}$ is the vector of partitioning coefficients of NPP to the three live biomass pools; and $U(t)$ is NPP.

Based on equation 3.1, we extracted the carbon cycle component of CLM-CASA' into a set of equations that described carbon transfer among pools, solved the steady state solution for the set of equations, and encoded them into MATLAB to perform data assimilation. Steady state solution for each gridcell was calculated as proposed by Xia et al. (Xia *et al.*, 2012):

$$X_{ss} = -(\mathbf{A}\bar{\boldsymbol{\xi}}\mathbf{C})^{-1}\bar{\boldsymbol{B}}\bar{U} \quad (3.2)$$

where $\bar{\boldsymbol{\xi}}$, $\bar{\boldsymbol{B}}$, and $\bar{U}$ were long-term averages of the environmental scalars, C partitioning among the three live pools, and NPP correspondingly. Before proceeding to data assimilation we verified the MATLAB version of CLM-CASA' by comparing its simulated soil C content with that of the original model. As illustrated in Figure S3.2, steady-state soil C simulated by the original CLM-CASA' model matched closely to

that produced by the MATLAB version despite that the latter used the semi-analytic spin-up approach (Xia *et al.*, 2012). Extracting the carbon component from the original model for data assimilation saved computational time and made it easier to fit our data-assimilation workflow.

3.2.2 Data assimilation for parameter estimation

To calibrate parameters associated with soil C we used Bayesian probabilistic inversion, which states that posterior probability density functions of parameters can be obtained from prior knowledge about the parameters and the error between model and observations. According to Mosegaard and Sambridge (2002) Bayesian inversion can be summarized by the following equation:

$$p(\mathrm{c}|\mathrm{Z}) = v_c \times \mathrm{p}(\mathrm{Z}|\mathrm{c}) \times \mathrm{p}(\mathrm{c}) \tag{3.3}$$

where p(c|Z) is posterior probability density function of model parameters c; p(Z|c) is a likelihood function of parameters c; p(c) is prior probability density function of parameters c; v_c is a normalization constant. We assumed that the prediction errors were normally distributed and uncorrelated, hence, the likelihood function, p(Z|c), was calculated as:

$$p(Z|c) = v_L \times \exp\left\{-\sum_{i=1}^{k} \frac{(Z_i - X_i)^2}{2\sigma_i^2}\right\} \tag{3.4}$$

where Z_i is soil C reported by IGBP-DIS at *i*th gridcell, X_i is soil C simulated by CLM-CASA' at a corresponding gridcell; ${\sigma^2}_i$ is the variance of a measurement at a gridcell; k is the total number of gridcells; and v_L is a constant. The issue of uncertainty in the global gridded data products is much understudied, and IGBP-DIS database, like many other global gridded data products, does not include the uncertainty estimates. In light

of the absent uncertainties we employed a concept used in Harmon and Challenor (1997), and assumed a standard deviation of 30% of a reported value at each grid cell, which we then used to calculate the variance.

To generate the posterior distributions we first specified the priors of the parameters to be uniformly distributed over the intervals specified in Table 3.1. We put constraints on parameters based on the literature, educated guess, and hypothesis testing. For instance, temperature sensitivity was varied between 1, assuming that respiration was insensitive to temperature, and 3 (average temperature sensitivity calculated from the global soil respiration database (Bond-Lamberty & Thomson, 2012a)). The minima for maximum turnover rates in slow and passive pools were within the reported range (Parton *et al.,* 2010, Parton *et al.,* 1993), however, because the maximum turnover rates for slow SOC pool in CLM-CASA' was the highest reported, we made upper limits in this study 2.5 times the original value to test whether the maximum turnover rate could be higher on the global level than those reported to date. The upper limit for the passive SOC turnover rate was set at 0.005 years^{-1}, which was higher than the rate used in (Parton *et al.,* 1993) by 0.0005 years^{-1}. Literature search yielded no information about the range for the other 17 parameters from Table 3.1, therefore, we assigned the limits to the parameters so that they would preserve the original relationship with the dependent variables (e.g. soluble fraction, structural lignin, clay, and sand), but adjust the degree of influence of the dependent variables on a C partitioning coefficient. As an additional constraint C partitioning coefficients had to be a value between 0 and 1, and sum of partitioning coefficients for C efflux from the same pool could not be larger than 1.

Table 3.1 CLM-CASA' parameter characteristics. MLE is maximum likelihood estimate, and G-R is the result of Gelman-Rubin chain convergence diagnostics.

Parameter description	Symbol	Prior			MLE $\times 10^{-3}$	G-R	Lower 95% bounds $\times 10^{-3}$	Upper 95% bounds $\times 10^{-3}$
		minimum $\times 10^{-3}$	maximum $\times 10^{-3}$	Value $\times 10^{-3}$				
Exit rate from slow pool, $years^{-1}$	c(11,11)	100	500	200	400	1.00	287	488
Exit rate from passive pool, $years^{-1}$	c(12,12)	1	5	4.5	1	1.00	1.04	2
Temperature sensitivity of C decomposition	Q_{10}	1000	3000	2000	1837	1.00	1751	1989
Labile C fraction effect on C partitioning from leaves to surface metabolic litter	w_1	0	1000.1	1000	503	1.00	120	920
Labile C fraction effect on C partitioning from roots to soil metabolic litter	w_2	0	1000.1	200	75	1.00	13	366
Partitioning from surface structural to surface microbial pool if no lignin in surface structural litter	l_1	350	2300	400	461	1.00	369	811
Lignin effect on partitioning from surface structural litter to surface microbial litter	l_2	0	700	400	462	1.00	166	674
Lignin effect on partitioning from surface structural litter to soil slow pool	l_3	0	800	700	59	1.00	11	302
Partitioning from soil structural to soil microbial pool if no lignin in soil structural litter	l_4	400	700	450	648	1.00	573	695
Lignin effect on partitioning from soil structural litter to soil microbial pool	l_5	0	900	450	135	1.00	23	557
Lignin effect on partitioning from soil structural litter to soil slow pool	l_6	0	900	700	654	1.00	334	875
C partitioning from soil microbial pool to slow pool if no sand or clay	t_1	19	750	169	655	1.00	527	741
Clay effect on C partitioning from soil microbial pool	t_2	0	30	5.44	25	1.00	17	30
Sand effect on C partitioning from soil microbial to slow pool	t_3	0	1000	678	343	1.00	57	844
Combined effect of sand and clay on C partitioning from soil microbial pool	t_4	0	100	22	67	1.00	34	96
C partitioning from soil microbial to passive pool if no sand or clay	t_5	0.1	0.8	0.51	0.39	1.00	0.14	0.73
Sand effect on C partitioning from soil microbial to passive pool	t_6	0	30	2.04	9	1.00	3.26	21
Clay effect on C partitioning from slow pool to passive pool	t_7	0	700	4.05	22	1.00	11	52
C partitioning from slow to passive pool if no clay	t_8	0	20	14	1	1.01	0.11	3
C partitioning from slow to soil microbial pool if no clay	t_9	300	700	449	384	1.00	314	599

Once we specified parameter ranges, we used adaptive Metropolis (AM) algorithm (Haario *et al.*, 2001), a Markov Chain Monte Carlo method, to sample from the posterior parameter distribution. To generate a parameter chain we ran AM algorithm in two steps: proposing step and a moving step. In the proposing step a new parameter set c^{new} was generated from a previously accepted parameter set c^{k-1} through a proposal distribution $(c^{new}|c^{k-1})$. In the moving step a probability of acceptance $P(c^{k-1}|c^{new})$ was calculated as in (Marshall *et al.*, 2004):

$$P(c^{k-1}\big|c^{new}) = min\left\{1, \frac{p(Z|c^{new})p(c^{new})}{p(Z|c^{k-1})p(c^{k-1})}\right\} \tag{3.5}$$

The value of $P(c^{k-1}|c^{new})$ was then compared with a random number *U* from 0 to 1. Parameter set c^{new} was accepted if $P(c^{k-1}|c^{new}) \geq U$, otherwise c^k was set to c^{k-1}.

The AM algorithm requires an initial parameter covariance matrix, which we generated from a test run of 50,000 simulations using a uniform proposal distribution following the example in Xu et al. (2006):

$$c^{new} = c^{k-1} + r \times \frac{c^{max} - c^{min}}{D} \tag{3.6}$$

where c^{max} and c^{min} are upper and lower parameter limits *r* is a random number between -0.5 to 0.5, and D=5. Out of 50,000 simulation the test run accepted about 2,500 updated samples. We constructed a covariance matrix C_0 on the basis of the test run and modified the proposal step to be

$$c^{new} = N(c^{k-1}, C_k) \tag{3.7}$$

$$C_k = \begin{cases} C_0 & k \leq k_0 \\ s_d Cov(c_0, \dots, c_{k-1}) & k > k_0 \end{cases} \tag{3.8}$$

where k_0 = 2000; $s_d = 2.38/\sqrt{20}$ (Gelman *et al.*, 1996).

We made five parallel runs starting at dispersed initial points in the parameter space, each run contained 1,000,000 simulations. During each simulation we solved equation 3.2 with newly proposed parameters and proceeded to the moving step. We discarded the first half of the simulations (as burn-in phase) and tested the second half for convergence to stationary distributions with Gelman-Rubin diagnostics (Gelman & Rubin, 1992).

3.2.3 Parameters to be estimated

Soil C content is determined by soil C influx (proportional to NPP) and its residence time, or the inverse of the turnover rate (Luo *et al.*, 2003). The point-by-point comparisons of the NPP produced by CLM-CASA' and the observed NPP from Ecosystem Model Data Intercomparison Initiative (EMDI) dataset (Olson *et al.*, 2001) showed good agreement between modeled and observed NPP (Randerson *et al.*, 2009b). Due to this good agreement we used modeled NPP to drive the CASA' biogeochemistry sub-model, and focused on improving the SOC through calibrating the parameters associated with turnover rates of carbon pools and C partitioning coefficients among pools as influenced by climate, tissue lignin, available nitrogen, and soil texture. Specifically, we calibrated the baseline C exit rates from slow and passive C pools (these two parameters strongly regulated SOC pool size due to their long residence times), temperature sensitivity of heterotrophic respiration (Q_{10}), and the parameters in (i.e., elements of) matrix **A**, which we described in the following equations:

$$a(4,1) = w_1 * lf \quad (3.9)$$

$$a(5,1) = 1 - a(6,3) \quad (3.10)$$

$$a(6,3) = w_2 * lf \quad (3.11)$$

$$a(7,3) = 1 - a(6,3) \quad (3.12)$$

where $a(4,1)$ is the partitioning coefficient from leaves to surface metabolic litter; $a(5,1)$ is the partitioning coefficient from leaves to surface structural litter; $a(6,3)$ is the partitioning coefficient from roots to soil metabolic litter; $a(7,3)$ is the partitioning coefficient from roots to soil structural litter; lf is the labile C fraction in leaves and roots at a given grid cell; and w_1 and w_2 are empirical coefficients.

$$a(9,5) = l_1 - l_2 * sl \quad (3.13)$$

$$a(11,5) = l_3 * sl \quad (3.14)$$

$$a(10,7) = l_4 - l_5 * sl \quad (3.15)$$

$$a(11,7) = l_6 * sl \quad (3.16)$$

where $a(9,5)$ is C partitioning coefficient from surface structural litter to surface microbial pool; $a(11,5)$ is C partitioning coefficient from surface structural litter to soil slow pool; $a(10,7)$ is C partitioning coefficient from soil structural litter to soil microbial pool; and $a(11,7)$ is C partitioning coefficient from soil structural litter to soil slow pool. sl is structural lignin content in the structural litter pools in a given grid cell expressed as a fraction of one, and l_1 through l_6 are empirical parameters.

$$a(11,10) = t_1 - t_2 * clay + t_3 * sand - t_4 * clay * sand \quad (3.17)$$

$$a(12,10) = t_5 + t_2 * clay + t_6 * sand - t_4 * clay * sand \quad (3.18)$$

$$a(12,11) = t_8 + t_7 * clay \quad (3.19)$$

$$a(10,11) = t_9 - t_7 * clay \quad (3.20)$$

where $a(11,10)$ is C partitioning coefficient from soil microbial to soil slow pool; $a(12,10)$ is C partitioning coefficient from soil microbial to soil passive pool; $a(12,11)$ is C partitioning coefficient from soil slow to soil passive pool; $a(10,11)$ is C

partitioning coefficient from soil slow to soil microbial pool; $clay$ and $sand$ are the soil clay and sand fractions in a given gridcell; and t_1 through t_9 are empirical parameters. Those parameters together with their prior values are listed in Table 3.1.

3.2.4 Database

IGBP-DIS database provides a high-resolution (5x5 arc minutes) global map of soil C for the top 1 m of soil. The map is the result of linking the pedon records from the extensive Global Pedon Database (Batjes, 1995), which contains soil texture classes, terrain slopes, and 106 soil units, and FAO/UNESCO Digital Soil Map of the world (containing the abovementioned data in addition to pH, organic C and nitrogen, bulk density, cation exchange capacity, etc. (FAO, 1995)) by statistical bootstrapping. The database features data on soil bulk density, field capacity, thermal capacity, soil carbon, and nitrogen density. Many studies used the soil carbon data from this dataset to produce new datasets (House *et al.*, 2002), as an assessment of terrestrial C uptake (Freibauer *et al.*, 2004), to evaluate models (Delire *et al.*, 2003, Kucharik *et al.*, 2000), and to improve models (Ise & Moorcroft, 2006, Smith *et al.*, 2013, Zhou *et al.*, 2009). Prior to using the data set in the data assimilation routine we randomly separated all the grid cells into halves similar to Smith et al. (2013). One half of grid cells was used for model calibration, and the second group – for validation to avoid overfitting. All further discussion about the fit of modeled data to observed data will refer to the validation subsample of the data.

In this study, for the purpose of estimation of the parameters related to carbon processes, soil C in the IGBP-DIS database was assumed to be in a steady state. This assumption could not be verified at present because of the lack of time series data over the globe. However, Zhou and Luo [2008] researched parameter uncertainty resulting from the steady state assumption by calibrating parameters two times: with equilibrium SOC values, and with SOC reduced 40% from the equilibrium values. Their finding was that the parameters obtained with the equilibrium assumption were within the margin of error of the parameters obtained after 40% reduction in soil C (except for the C partitioning coefficients from the soil pools), therefore, even with equilibrium assumption we were likely to constrain the parameters at their true values. Moreover, 40% from the steady-state soil C content represents extreme cases as most of the disturbances do not cause such big changes in soil C. The steady-state assumption made our MATLAB model version particularly effective for data assimilation as it used averaged values of environmental responses (scalars) to repeated environmental forcing (Xia *et al.*, 2012).

The litter lignin content was in CLM-CASA' by default: the plant-functional-type-level estimates of lignin were applied to MODIS-derived distribution of plant functional types used in CLM (Lawrence & Chase, 2007, Oleson *et al.*, 2007). Labile C pool fraction in roots was estimated by the CLM-CASA' model from the lignin to nitrogen ratio, and the latter, similarly to litter lignin content, was by default assigned to each plant functional type in CLM-CASA'. The maps of soil sand and clay content were originally developed by the International Geosphere-Biosphere Programme (Group, 2000) and were provided as a part of CLM-CASA' package.

3.2.5 Forward analysis of carbon dynamics with original and optimized parameters

After parameter calibration we ran forward simulations with the original and calibrated parameters, and examined the change in ecosystems' responses to increasing temperature and atmospheric CO_2 concentrations. We used the Community Earth System Model (CESM) output for the Representative Concentration Pathway 8.5 (RCP8.5) experiment (specifically, temperature and live pool sizes) to drive the litter and soil C pools in CLM-CASA' with original and calibrated parameters. The CESM model output was provided as a part of Coupled Model Intercomparison Project Phase 5 (CMIP5), and could be accessed at http://pcmdi9.llnl.gov. Over 95 years CESM simulated a 3.5 K increase in mean global temperature and atmospheric CO_2 increase to 1150 ppm by the year 2100 (Keppel-Aleks *et al.*, 2013). We ran the CASA' model forward in time starting from the year 2006 to 2100, using the average live pools, their residence times, and temperatures for the years 2006-2010 to generate initial litter and soil C pools using the semi-analytical spin-up (Xia *et al.*, 2012).

3.3 Results

3.3.1 Evaluation and improvements of modeled SOC

We estimated SOC pool sizes at steady states (equation 3.2), which we then compared to observed SOC pools provided by IGBP-DIS global gridded product (Fig. 3.2b, (Group, 2000)). CLM-CASA' explained 27% of spatial variability in the observed data (Fig. 3.3a and b). The model on average underestimated soil C pools by 8.8 kg/m² with the largest deviations in the northern regions, where soil C was underestimated by

more than 30 kg/m² (Fig. 3.3a). The root-mean-square error (RMSE) of CLM-CASA' was 11.34 kg/m².

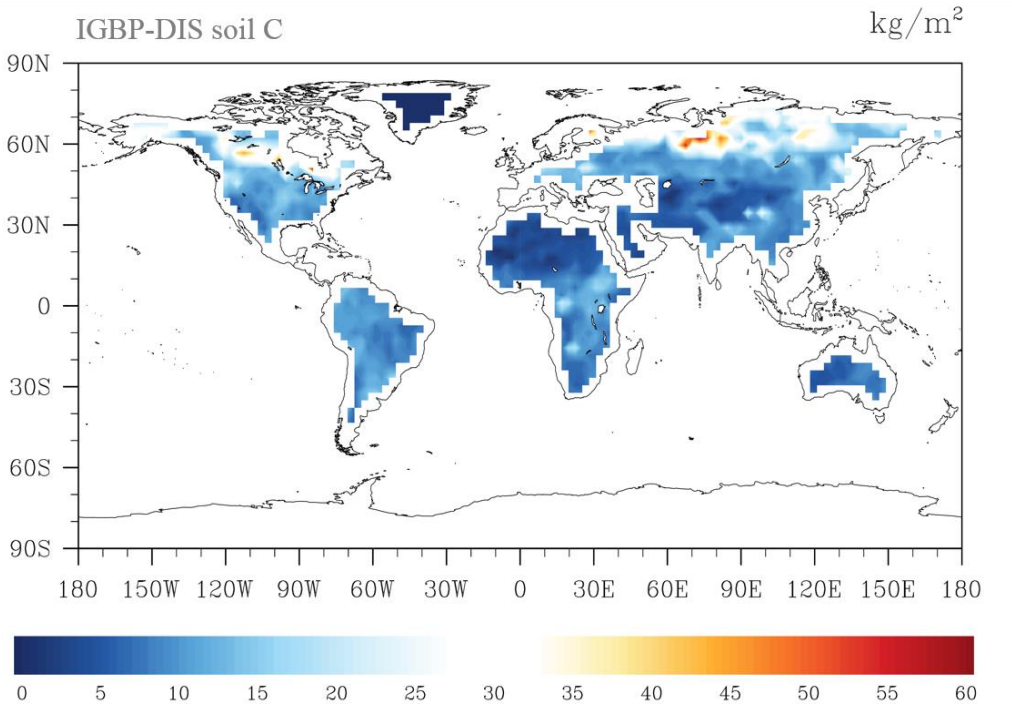


Figure 3.2 IGBP-DIS soil carbon distribution. Soil carbon varies from 0 kg/m² in deserts to 60 kg/m² in the northern regions.

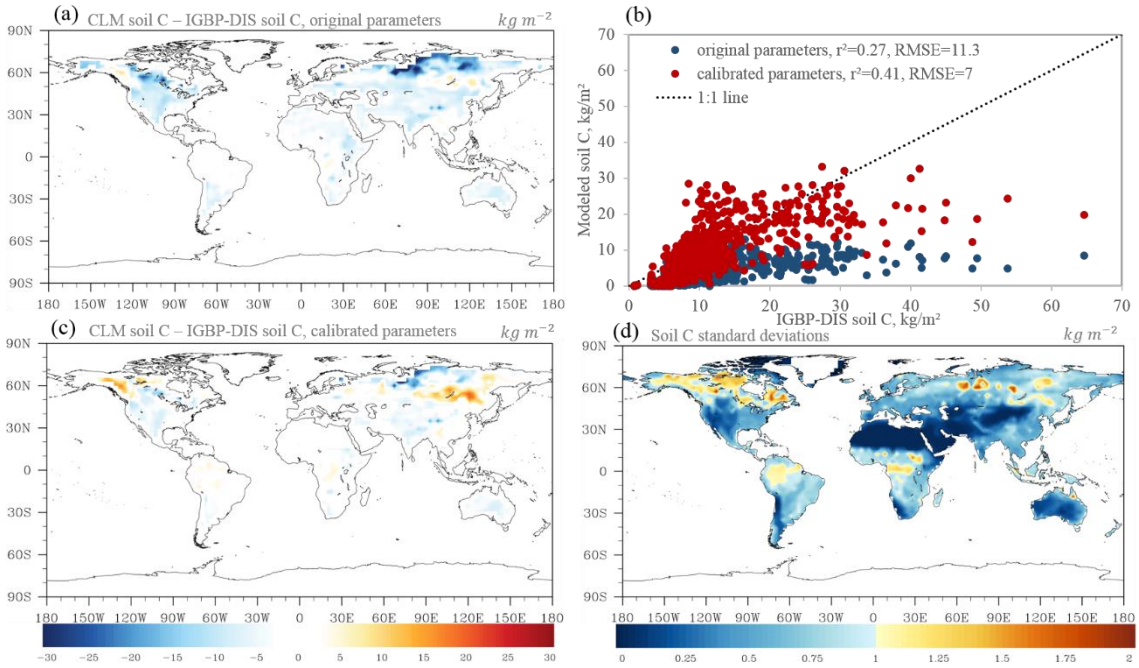


Figure 3.3 Spatial correspondence of CLM-CASA' produced SOC to the IGBP-DIS reported SOC before (a,b) and after (b,c) data assimilation; and standard deviations in the modeled soil C after data assimilation (d). The points in panel (b) represent the grid cell values. Model with default parameters explained 27% of variation in the observed soil C, whereas model with calibrated parameters explained 41% of variability in the observed soil C. The regions with highest uncertainty were located in the northern latitudes and in the tropics.

3.3.2 Constrained parameters related to soil carbon dynamics

All estimated parameters converged to stationary distributions as indicated by the Gelman-Rubin diagnostics (Gelman & Rubin, 1992) (Table 3.1). Assimilation of SOC dataset into CLM-CASA' model resulted in tight constraints on temperature sensitivity of heterotrophic respiration (Q_{10}) and clay effect on C partitioning from slow to passive pool (t_7) (Fig. 3.4). Sharp posterior distributions indicated that model

predictions were highly sensitive to changes in those parameters, which agreed with findings of Post et al. (2008). Sand effect on C partitioning from soil microbial to passive pool (t_6) was also constrained within the prior values, however, its distribution was not as sharp as distributions of Q_{10} and t_7. Many posterior parameters were skewed against their assigned maximum or minimum values. This was especially the case for baseline passive C exit rate, which was skewed against its minimum value, suggesting that residence time of passive SOC under optimum environmental conditions was larger than 1000 years. Parton et al. (2010) assigned passive pool residence times up to 5000 years in DayCent model (a model with similar carbon sub-model to CLM-CASA'). However, in the same study they presented the improved model version, ForCent, which had the highest value of 1000 years, hence we did not increase the maximum boundary for baseline passive pool's residence time above 1000 years.

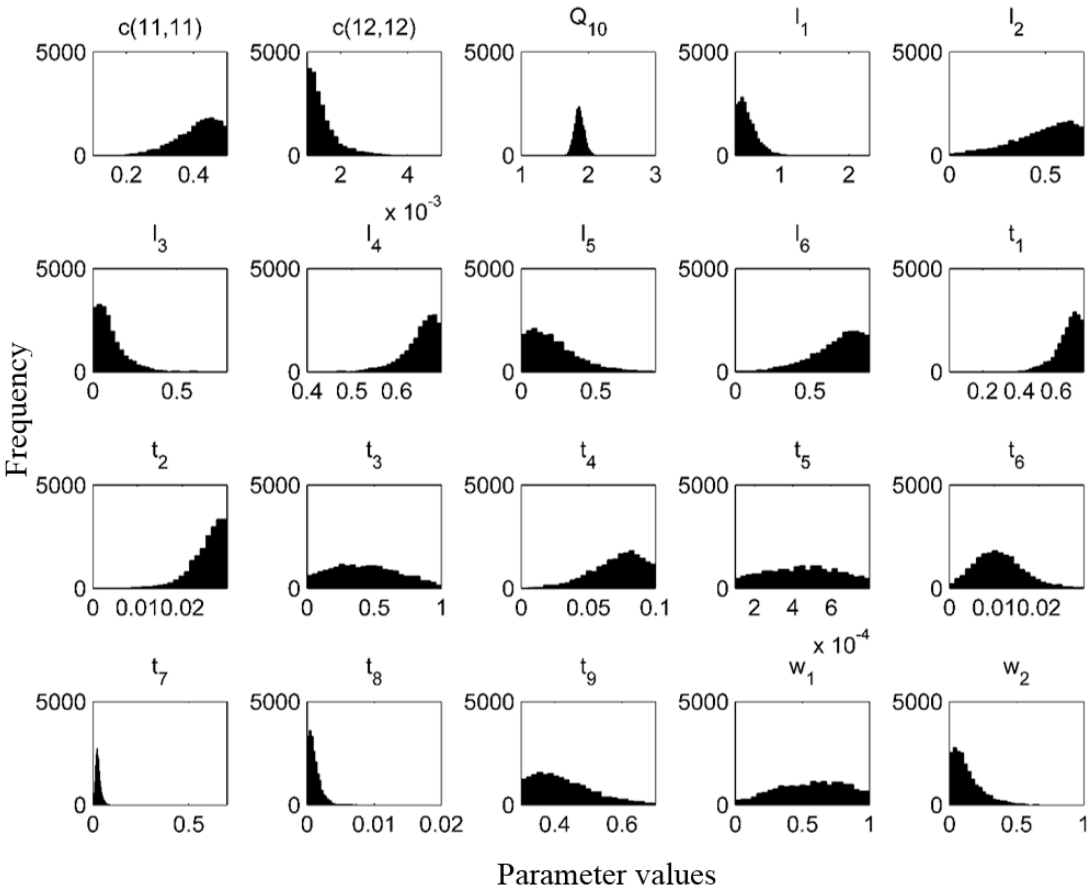


Figure 3.4 Frequency distributions of 20 calibrated parameters. The most constrained parameters were temperature sensitivity of heterotrophic respiration ($\mathbf{Q_{10}}$), clay effect on C partitioning from slow to passive pools ($\boldsymbol{t_7}$), and sand effect on C partitioning from soil microbial to passive pool.

Many parameters associated with soluble fraction, structural lignin, and soil texture effects on C partitioning coefficients were skewed against their boundaries. For instance, posterior distribution of t_8, was skewed against its minimum (Fig. 3.4), indicating increased effect of clay on C partitioning from slow to passive pool as lower fraction would be transferred to passive pool in soils with low clay content (eq. 3.20). Skews of $t_1$and t_2 against their maxima indicated that more C was transferred from fast

to slow pool under low clay and sand content, and clay effect on C transfer from fast to slow and passive pools was larger than modeled originally (eq. 3.17,3.18). Skew of l_3 against its minimum indicated either no relationship between structural lignin and C partitioning between surface structural and slow pools, or very low C transfer between these pools (eq. 3.14). Similarly, calibration revealed absence of relationship between labile C fraction and C transfer from roots to soil litter pools as indicated by the skew in w_2, however it could also mean that under the given model formulation all C from roots had to go to soil structural litter to reproduce spatial patterns of SOC. Some of the parameter skews pointed at limitations of the model formulation, for instance, skews of l_1 and l_2 were pushing a(9,5) to negative values, and skew in l_6 pushed a(11,7) to be larger than one, which was unrealistic. Lastly, observed SOC contributed little information for C partitioning from leaves to surface metabolic litter (w_1) and C partitioning from soil microbial to passive pool if no sand or clay content (t_5).

After parameter calibration temperature sensitivity, Q_{10}, decreased from 2 to 1.84 (with 95% CI of 1.75-1.99); clay effects on partitioning coefficients were increased, whereas the effects of lignin and labile C fractions were decreased (Table 3.1). Less carbon reached slow pool from surface litter and soil litter (Fig.3.1), which, in combination with the decreased optimum slow pool's residence time, decreased the size of the slow pool from 361 Pg to 275 Pg (95% CI 205-459 Pg). Soil microbial and passive pools, on the contrary, increased. Soil microbial pool increased from 10.6 Pg to 11.8 Pg (95% CI 9.2-18.1 Pg) due to higher partitioning coefficient from soil litter and lower Q_{10}. Passive pool increased from 346 Pg to 865 Pg (95% CI 735-1047 Pg), and the change was due to increase in its residence time (Fig.3.1) and decrease in Q_{10}.

Calibrated parameters also showed that, compared to original parameters, a higher fraction of the fine root carbon was allocated to the structural litter, and a lower fraction was allocated to the metabolic litter (Fig.3.1). Increase in the abovementioned C allocation coefficient to soil structural litter together with lower Q_{10} led to an overall increase in soil litter pool due to longer C residence time in the soil structural than soil metabolic litter pool. In accordance with C cycle theory (Luo & Zhou, 2006) larger soil litter pool contributed more C to the soil pool, which in combination with lower C partitioning from surface litter to soil led to a shift in importance of C inputs from surface and belowground litter for SOC pool formation. For instance, the initial ratio of soil structural litter to surface structural litter inputs into soil slow pool was 0.66, and after parameter optimization the ratio became 6.69. Thus, soil litter contribution to slow soil C pool formation was more important than modeled originally. The implication of more root inputs rather than shoot inputs into soil C pool is supported by long term experimental (Kätterer *et al.*, 2011) and meta-analysis (Rasse *et al.*, 2005) studies.

3.3.3 Parameter correlations

We used the posterior distributions to calculate correlations among the model parameters (Table 3.2). Out of 190 parameter pairs, 3 pairs had strong correlations. The strongest positive correlations were between passive C pool turnover, c(12,12), and clay effect on partitioning from slow to passive pool, t_7; and between sand effect on C partitioning from soil microbial to passive pool, t_6, and t_7. The strongest negative correlation was between t_6 and sand effect on C partitioning from soil microbial C pool to slow pool, t_3.

Table 3.2 Correlations between the calibrated parameters. The darker the shade of the cell the stronger the correlation .The strongest positive correlations were between passive C pool turnover and clay effect on partitioning from slow pool and passive pool; and between sand effect on C partitioning from soil microbial C pool to passive pool and clay effect on partitioning from slow pool to passive pool. The strongest negative correlations were between sand effect on C partitioning from soil microbial C pool to passive pool and sand effect on C partitioning from soil microbial C pool to slow pool.

Parameter correlations

	c(11,11)	c(12,12)	Q_{10}	l_1	l_2	l_3	l_4	l_5	l_6	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	w_1	w_2
c(11,11)	1.00																			
c(12,12)	-0.04	1.00																		
Q_{10}	0.10	0.21	1.00																	
l_1	0.05	-0.01	0.19	1.00																
l_2	-0.03	-0.01	-0.09	0.07	1.00															
l_3	0.07	0.14	0.25	-0.09	0.04	1.00														
l_4	0.02	0.03	-0.20	0.04	-0.05	0.03	1.00													
l_5	-0.07	-0.04	0.30	-0.01	0.07	-0.04	0.02	1.00												
l_6	-0.02	0.08	-0.27	-0.06	-0.01	-0.01	-0.08	0.14	1.00											
t_1	0.00	0.07	-0.15	-0.09	-0.02	-0.05	0.01	0.03	-0.10	1.00										
t_2	-0.03	-0.23	-0.07	0.04	-0.01	0.02	-0.02	-0.03	-0.05	-0.12	1.00									
t_3	0.01	-0.07	0.07	0.10	-0.05	0.01	0.02	-0.02	-0.06	-0.02	0.06	1.00								
t_4	0.06	0.10	-0.22	-0.06	0.08	-0.14	0.02	-0.02	0.11	0.13	-0.01	0.13	1.00							
t_5	0.00	-0.03	-0.06	0.00	0.03	-0.03	0.08	-0.03	-0.07	0.04	0.00	0.00	0.01	1.00						
t_6	0.13	0.30	-0.05	-0.08	0.05	-0.04	-0.01	0.11	0.07	-0.08	-0.15	-0.56	0.30	-0.04	1.00					
t_7	0.01	0.80	-0.01	-0.17	0.11	-0.01	0.02	0.01	0.09	0.01	-0.38	-0.14	0.37	-0.02	0.55	1.00				
t_8	0.00	0.32	-0.07	-0.03	0.03	0.03	0.01	-0.05	0.03	0.00	-0.05	-0.05	-0.05	0.00	-0.05	0.17	1.00			
t_9	0.07	-0.08	-0.26	-0.06	0.01	-0.15	-0.01	-0.05	0.06	0.00	0.05	-0.12	0.07	-0.04	-0.30	-0.26	-0.11	1.00		
w_1	0.08	0.05	0.29	-0.19	0.02	0.00	-0.09	0.01	-0.03	-0.07	-0.05	-0.08	0.05	-0.02	0.00	0.00	-0.03	-0.03	1.00	
w_2	-0.01	-0.11	0.15	-0.08	-0.07	0.02	-0.10	0.04	-0.06	0.02	0.03	0.01	-0.13	0.01	-0.04	-0.13	-0.07	0.00	0.09	1.00

3.3.4 Improvement of soil carbon estimation with data assimilation

After parameter calibration, the fitness between observed and modeled SOC improved substantially (Fig. 3.3b and c): calibrated model explained 41% of the variability in the observed SOC whereas uncalibrated model explained 27%. The model with calibrated parameters also reduced the magnitude of the average underprediction to 2.1 kg/m² and decreased RMSE by 38%. Using posterior parameter distributions we generated the standard deviations of SOC for each grid cell to illustrate the uncertainty in the estimated SOC. The regions with the highest uncertainties were located in the northern latitudes and tropics (Fig. 3.3d). Global SOC content in CLM-CASA' increased from 762 Pg to 1205 Pg (with 95% CI of 1150-1293 Pg) after data assimilation. The global SOC range was within the 95% CI of the observed HWSD SOC estimates presented by Todd-Brown et al. (2013b), however, it was still lower than global observed soil C content in IGBP-DIS.

3.2.5 Carbon pool responses to environmental change

To illustrate the impact of parameter calibration we performed the forward runs using a climate change scenario (RCP8.5). Over 95 years soils in the calibrated model released 48 Pg C less than soils in the original model (Fig. 3.5a). Similarly, carbon loss from litter decreased by 6.5 Pg in the calibrated model, compared to the original model (Fig. 3.5b). The decreases in the C loss rates in litter and soil C pools were caused by reduction in temperature sensitivity. The decreased soil C loss was also caused by the decrease in the calibrated passive pool's turnover rate. We also generated the uncertainties of changes in C pools from the posterior parameter distributions to gain

perspective about the magnitude of the range for the cumulative C change after 95 years (gray lines in Fig. 3.5). The uncertainty ranges were quite large: from 15 to 100 Pg for cumulative soil C loss, and from -30 to +5 Pg for cumulative litter C change. The magnitude of uncertainties was caused mostly by the uncertainties in maximum turnover rates and Q_{10}, as soil C dynamics was most sensitive to those parameters (Post *et al.*, 2008).

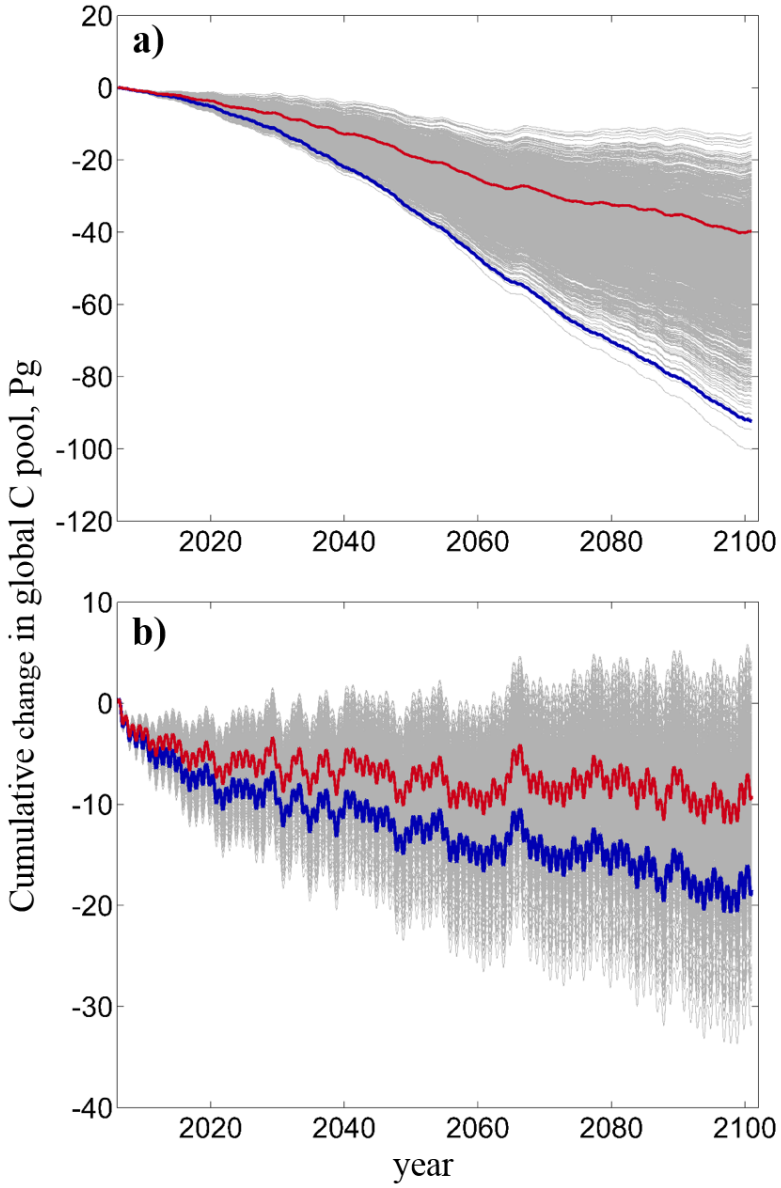


Figure 3.5 Change in CLM-CASA's soil (a) and litter (b) carbon under RCP 8.5 climate change scenario. Blue lines are projections of the model with original parameters, red lines – model with maximum likelihood parameters, and gray lines are projections of the models with parameter samples from the posterior distributions (sample size = 2000), representing the uncertainties of projections. Soils released 45 Pg C less in the model with calibrated parameters than the model with original parameters.

Similarly, litter released 6.5 Pg C less in the calibrated model's projections than in the original model's projection. Despite the differences in models' projections, the original model's projections were within the uncertainty range.

3.4 Discussion

3.4.1 Current status of soil carbon modeling

Currently most carbon cycle models are similarly formulated: they partition the photosynthetically fixed C among the live and dead carbon pools, from which C is released via autotrophic or heterotrophic respiration (Todd-Brown *et al.*, 2012, Weng & Luo, 2011). Despite the conceptual similarity of the models their estimates of global soil C range from 514 to 3046 Pg with poor spatial correlation with the empirical estimates (up to r^2=0.16 (Todd-Brown *et al.*, 2012)). While the C influx is important determinant of the C pool size, the capacity and sustainability of a C sink (e.g. soil C sink) under changing environmental conditions are strongly regulated by the C sink's residence time (Luo *et al.*, 2003, Luo *et al.*, 2001b), hence, it is important to calibrate the parameters regulating it.

Since the start of global soil C map production in 1982 (Post *et al.*, 1982), to our knowledge, only very few studies have made an attempt to use global soil C distributions to constrain global soil C residence times. For example, Zhou et al. (2009) and Ise and Moorcroft (2006) used the IGBP-DIS dataset (Group, 2000) to constrain global soil C temperature sensitivities, and Smith et al. (2013) developed the C cycle model constrained with multiple data streams. This study showed that the poor model performance in simulating soil C could be substantially improved via parameter

calibration against the observed soil C data. To model the C cycle as realistically as possible, we not only need to incorporate more processes into ESMs as the modeling community currently focuses on (Koven *et al.*, 2009, Koven *et al.*, 2011, Lawrence *et al.*, 2011a, Luo *et al.*, 2009) but also need to develop our capacity to calibrate parameters effectively against the observed data.

3.4.2 Improvement of soil carbon modeling

Our results showed that applying data assimilation could greatly reduce the mismatches between modeled and observed soil C. This error reduction was achieved largely due to increase in the soil C residence times. The implication that residence time in the original model was too low was in agreement with literature. For instance, Rumpel and Kögel-Knabner (2011) reported isotope-derived passive C pools' residence times for the top meter of different soil types ranging from 1026 years to 3030 years with samples from South America, Australia, Eurasia, and Africa. In comparison, the original model underestimated passive pool C residence time by 42-2046 years, whereas the calibrated model produced an estimate closer to the reported values (global average C residence time of passive pool was 2627 years with 95% CI 1583-3581 years).

To generate the estimates comparable to the ones in Todd-Brown et al. (2013b) we calculated global weighted averages of soil residence time as the ratio of soil C pool to NPP. After parameter calibration soil C residence time increased from 14 to 26 years (with 95% CI 25-28 years), which was within the range of observed soil C residence times reported in Todd-Brown et al. (2013b). We then calculated soil C residence time

as the ratio of soil C pool to soil heterotrophic respiration, and observed that global mean residence time increased from 32.3 years to 51 years (95% CI 45-61 years). The posterior estimates were too high compared to a global soil residence time of 32 years reported in (Raich & Schlesinger, 1992). However, in the above-mentioned study soil residence time was obtained with the assumption that 30% of soil respiration originated from roots, which was lower than the average global estimate derived from Bond-Lamberty and Thomson (2010) (54%). We corrected the assumption in Raich and Schlesinger (1992) and SOC residence time became 48.6 years, which was within the 95% CI for our calibrated estimate. Combining more up-to-date range of estimates for SOC (Todd-Brown *et al.*, 2013b) and global soil respiration (Bond-Lamberty & Thomson, 2010) we obtained a range for global soil residence times of 16.5 to 41 years, which was outside of our posterior 95% CI. Such spread in the observed estimates highlights the need for datasets with better confidence.

We tested how new residence times affected C flux predictions by comparing the modeled soil heterotrophic respiration to the observed values from Bond-Lamberty and Thomson (2010, 2012a). The global observed range for soil heterotrophic respiration (calculated from soil respiration) was 41-54 Pg/year. Both model predictions were within the observed estimates: original model simulated soil heterotrophic respiration at 46 Pg/year and calibrated model predicted 40 Pg/year with 95% CI 36-52 Pg/year. The spatial predictions of the calibrated model were significantly different from the original model ($P<0.05$), and although there was no improvement in the explained variability of the observed data after calibration, RMSE reduced by 10% (Fig. 3.6). The decrease of soil respiration was a logical outcome of the decrease in Q_{10} and

maximum turnover rate of soil passive pool although the latter is somewhat offset by increase of maximum turnover rate in slow pool. The lack of large improvement in prediction of the observed soil heterotrophic respiration may be due to interannual variability of the observations, which was not accounted for as we assumed equilibrium conditions for the purpose of parameter calibration; differences in scale (model grid was around 2.8 by 2.8º, whereas the observation was a point on a grid); and limitations in model structure, which was also indicated by parameters approaching unrealistic values by skewing against their maxima or minima.

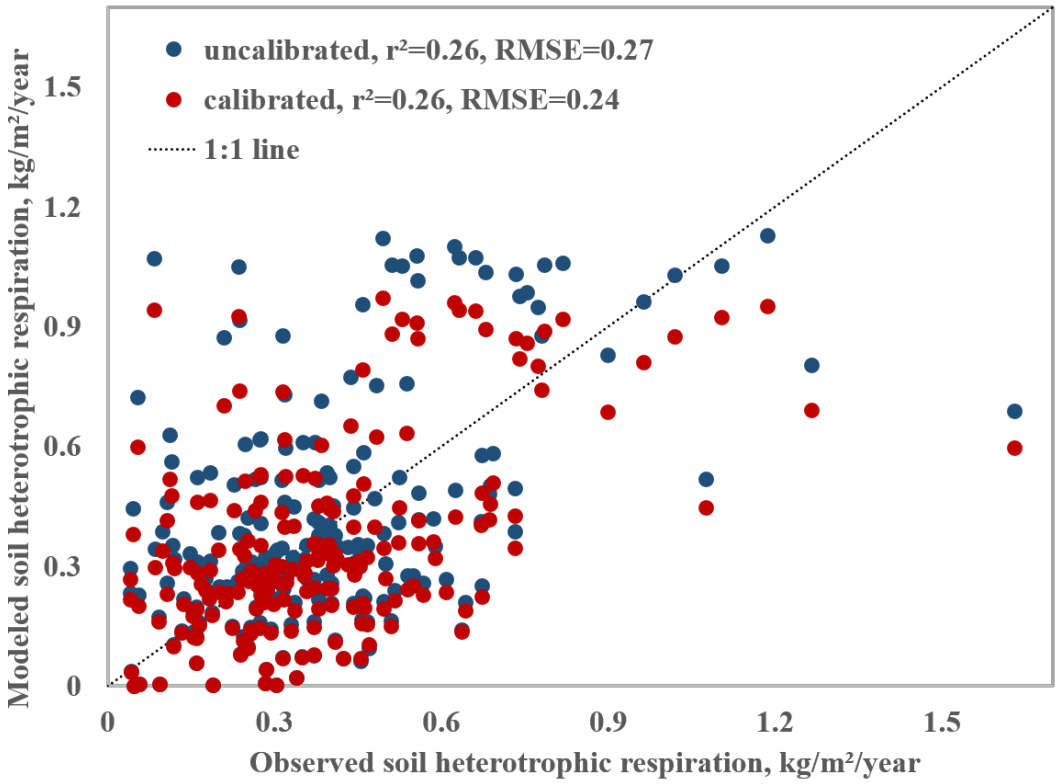


Figure 3.6 Point-by-point comparison of the modeled soil heterotrophic respiration to the observed data derived from Bond-Lamberty and Thomson (2012a). Parameter

calibration did not change the explained variability in the observations, it reduced the RMSE by 10%.

The large role of chemical stabilization in soil C pool formation illustrated by increase in clay effect on C partitioning to more recalcitrant pools has been supported by the observations and is described in the literature (Krull *et al.*, 2003, Paul *et al.*, 2008, Torn *et al.*, 1997). Parameter calibration also revealed a larger regulation of soil C by soil litter inputs than surface litter inputs, which was also supported by experimental results (Kätterer *et al.*, 2011, Rasse *et al.*, 2005). Hence, through model calibration we were able to not only improve the fitness between modeled and observed soil C, but also gain insights into the C cycle processes.

3.4.3 Uncertainty in future C projections

With the prospect of global warming it is important to constrain the magnitude of SOC feedback (which depends on SOC residence time) as fast CO_2 release from soils may accelerate the increases in temperature. We discussed earlier that parameter calibration decreased both the maximum turnover rate and temperature sensitivity of SOC. Our maximum likelihood temperature sensitivity was within the range of the estimates constrained against the atmospheric CO_2 concentrations by Jones and Cox (2001) (2.1 ±0.7 and 0.91±0.4); higher than the average global Q_{10} reported in Zhou et al. (2009) (1.72); and higher than the estimate in Ise and Moorcroft [2006] (1.37). The Q_{10} from Ise in Moorcroft [2006], however, was likely underestimated as authors used NPP rather than C influx from litter for SOC input. According to Appendix Fig. 1, a higher input would decrease residence time for the same SOC under equilibrium

conditions. The range in the posterior SOC residence times caused large uncertainties in the cumulative soil C change (a 15-100 Pg loss), which was within the range reported in Smith et al. (2013) (from a 138 Pg C loss to a 82 Pg C gain). The smaller uncertainty range in this study did not indicate better constraint, it was most likely due to the lower number of parameters: the uncertainties in Smith et al. (2013) resulted from parameters regulating soil C dynamics as well as parameters regulating plants' carbon fixation. The large uncertainty ranges clearly indicated the need for additional data sets to further constrain model parameters.

3.5 Conclusion

We calibrated a global land model against the global observed soil C dataset using Bayesian MCMC approach. By adjusting parameters that influenced residence times of C pools we were able to substantially improve the fitness between observed and modeled soil C. After parameter calibration we discovered that the regions with the highest uncertainty in soil C were located in the northern latitudes and tropics, indicating that we need to put more efforts in researching soil C dynamics in those regions. When propagated over time under RCP8.5 scenario posterior parameter uncertainties resulted in 15-100 Pg C loss from soils over 95 years, however the maximum likelihood C loss (40 Pg) was 53% lower than SOC loss predicted with original parameters. Similarly, maximum likelihood litter loss decreased 53% from the original predictions, but the range of change was from a 30 Pg C loss to a 5 Pg C gain.

As we show in this study, data assimilation is a useful tool for gaining insights into carbon cycle modeling, identifying regions that need additional data collection and

more research to improve representation of ecosystem carbon processes, as well as for uncertainty analysis of the model projections. More global level C cycle data (e.g. litter pools, litter residence times, soil C residence times, live C pools, and their residence times) along with their uncertainties are needed to constrain the processes in carbon cycle. Assimilating the data into the model at global level will constrain C cycle feedbacks and provide more confidence in the future carbon budget, which will facilitate more effective management of the natural resources.

Supplemental Materials

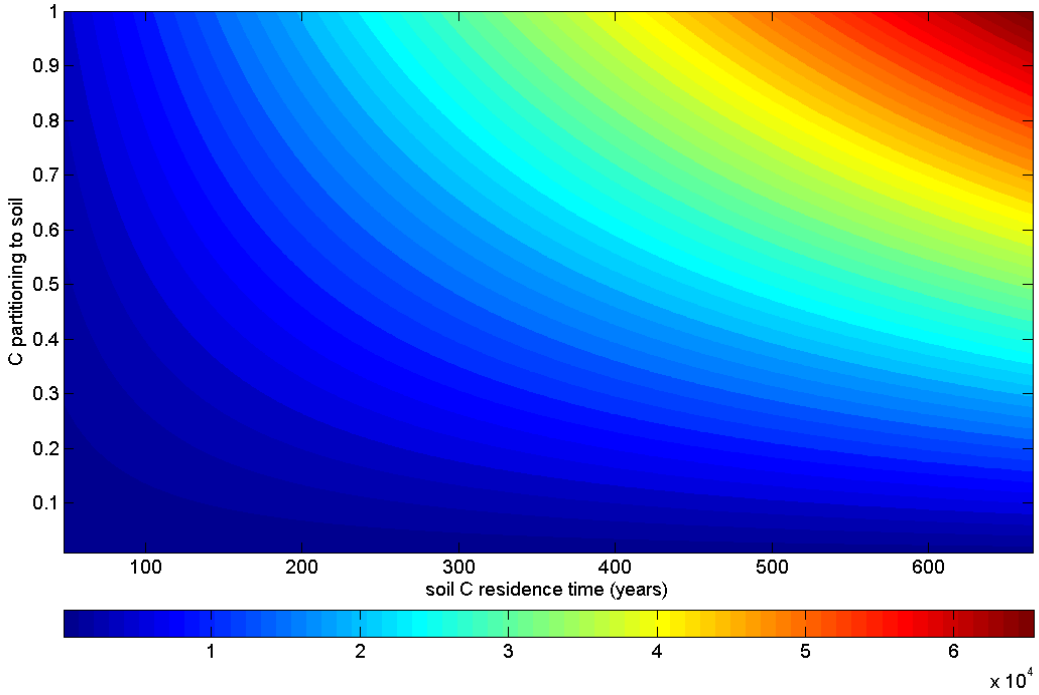


Figure S3.1 SOC stocks under different combinations of C transfer coefficients from plant to soil pools (or C partitioning to soil) and soil C residence times. For the purpose of illustration, NPP in this figure is fixed at 0.1 $\mathbf{kg\ m^{-2}\ year^{-1}}$ and Soil C equals the product of NPP, C transfer coefficient and residence time. The color bar represents soil C pools in $\mathbf{g\ m^{-2}}$. In this figure we illustrate the equifinality of SOC content in a gridcell: same SOC value can be obtained with different combinations of C transfer coefficients from plant to soil and SOC residence times under fixed (observed) NPP.

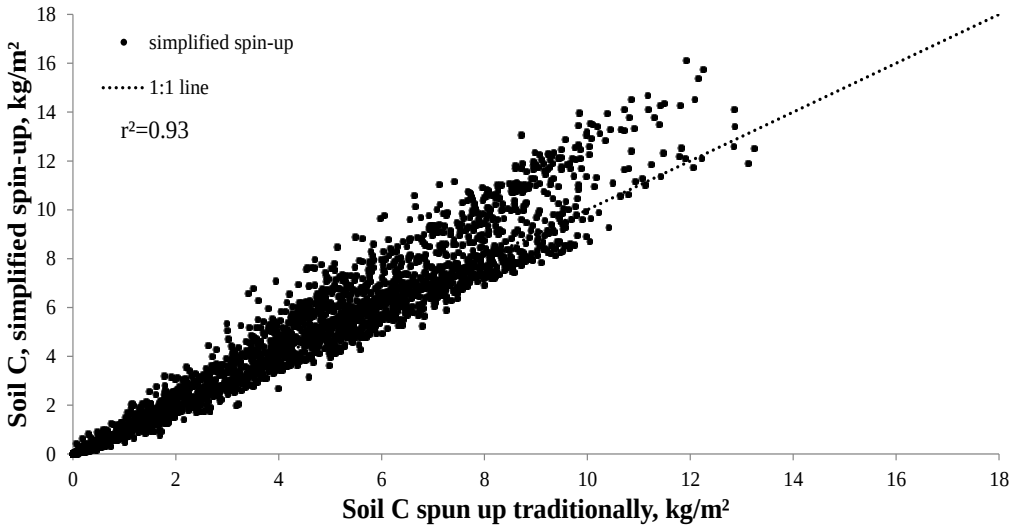


Figure S3.2 Comparison of traditional and speeded up spin up of soil C.

Chapter 4

Modelling microbial processes in soil improves global carbon stocks predictions

Abstract

Long-term carbon (C) cycle feedback to changing climate depends on future soil C dynamics. Current models do not represent soil C dynamics well, and although assimilating observations improves model performance, there is still substantial uncertainty in the future soil C feedbacks to climate change. Soil C feedbacks depend on microbial activity responses to changing climate, which in most ("conventional") models is simulated as a constant modified by environmental functions. Including microbial biomass dynamics into C cycle model formulation has shown potential to improve the representation of the global soil C dynamics. This study was to calibrate the parameters in two soil microbial models; evaluate the calibrated microbial models' performance; and compare the soil C climate change feedbacks and their uncertainties between microbial and conventional models. Microbial models with calibrated parameters explained 51% of variability in the observed total soil organic C, whereas conventional model explained 41%. The 2-pool microbial model exhibited unrealistic oscillations in soil C dynamics, however the oscillations were less prominent or absent in the 4-pool microbial model. Microbial models produced stronger soil C feedbacks to 95 years of climate change than any of the 11 conventional models used in CMIP5, simulating an 8% (2-pool model) and 11% (4-pool model) losses whereas CMIP5 models projected from a 7% loss to 22.6% gain in soil C. To further improve model performance and reduce uncertainty in future feedbacks, we need more data on microbial dynamics, such as controls over microbial carbon use efficiency and substrate quality limitation in various ecosystems.

4.1 Introduction

Soils contain the largest fraction of global terrestrial carbon (C), storing more C than vegetation and atmosphere combined (Falkowski *et al.*, 2000, House *et al.*, 2002). Changing climate may accelerate soil organic carbon (SOC) decomposition (Fang *et al.*, 2005), increasing atmospheric CO_2 concentrations, which may cause further climate warming (Falkowski *et al.*, 2000). Such potential interdependencies between SOC and climate highlight the importance of accurate predictions of global SOC distributions and their feedbacks to increasing temperatures and atmospheric CO_2 concentrations.

Predictions of the global SOC pools from 11 earth system models (ESMs) participating in the 5th Coupled Model Intercomparison Project [CMIP5, (Taylor *et al.*, 2011)] vary six-fold, ranging from 510 to 3040 Pg C (Todd-Brown *et al.*, 2013b). Out of 11 models only six produced SOC pools that were within the range of the Harmonized World Soil Database estimates (HWSD, 890–1660 Pg C) and none of the models explained more than 16% of spatial variability in the HWSD soil C (Todd-Brown *et al.*, 2013b). Current ESMs' performance calls for improving the models used for soil carbon cycle simulation.

All CMIP5 models simulate soil carbon decomposition as first-order decay process (Todd-Brown *et al.*, 2013b). Such model formulation (we will call it "conventional") represents the decomposing activity of microbes as decay constants, modified by environmental functions, and assumes that amount of the decomposed SOC is linearly dependent on the SOC stocks. Conventional models do not account for microbial processes observed in experimental studies, such as priming effect (Fontaine *et al.*, 2004, Fontaine *et al.*, 2007, Kuzyakov *et al.*, 2000), microbial acclimation to

increasing temperatures (Chen & Tian, 2005, Luo *et al.*, 2001a, Peng *et al.*, 2009), and CO_2-induced change in the microbial community composition (Carney *et al.*, 2007). Schimel and Weintraub (2003) argue that SOC decomposition shouldn't be represented by decay constants as it is catalyzed by extracellular enzyme concentration, and propose to model SOC dynamics by Michaelis-Menten equation, modified to include the concentration of the catalyst (enzyme).

In the recent years several enzyme-driven decomposition models (we will call them "microbial models") were developed (Allison *et al.*, 2010, German *et al.*, 2012, Schimel & Weintraub, 2003, Wang *et al.*, 2012b). The emerged microbial models simulate the acclimation of soil respiration to elevated temperatures (Allison *et al.*, 2010) as well as priming effect (Schimel & Weintraub, 2003). Once applied to the Community Land Model the new additions improved the prediction of global SOC distribution (Wieder *et al.*, 2013). However, microbial models may also produce responses not observed in nature: a recent stability analysis illustrates that microbial models produce unrealistic oscillatory responses to small environmental perturbations (Wang *et al.*, 2013). Such unrealistic model properties emphasize the importance of the thorough analysis of the improvements made to the carbon cycle models.

Apart from changing model assumptions about ecosystem processes, model improvement can be achieved by calibration of the model parameters. With the increase in the globally observed data, more studies have focused on assimilating global datasets into carbon cycle models. For instance, Zhou et al. (2009) and Ise and Moorcroft (2006) assimilated global SOC (Group, 2000) data into a C cycle model to constrain SOC temperature sensitivities; Hararuk et al. (2014) researched how assimilating global SOC

data changed model parameters and SOC feedbacks to changing climate; and Smith et al. (2013) assimilated multiple data into a carbon cycle model fully constraining it. All of these studies parameterized conventional carbon cycle models, and to date, no research had been done on calibrating and researching the properties of data-constrained global microbial decomposition models.

This study was to (1) calibrate two microbial models to the global estimated distributions of total soil organic carbon and microbial biomass carbon; (2) compare the performance of the calibrated microbial and conventional models; (3) test for presence of unrealistic oscillations in the microbial model dynamics; and (4) compare the SOC climate change feedbacks and their uncertainties between microbial and conventional models.

4.2 Methods

4.2.1 Models

We performed Bayesian parameter estimation on two microbial models and compared the parameter estimates and model predictions to those of a calibrated conventional model (Hararuk *et al.*, 2014). For the conventional model we used Community Land Model coupled with Carnegie-Ames-Stanford Approach biogeochemistry sub-model (CLM-CASA') (Oleson *et al.*, 2004, Oleson *et al.*, 2008, Parton *et al.*, 1993) with the soil carbon cycle compartment modeled as a 3-pool system and C transfers among the pools regulated by temperature, soil clay content, and pool sizes. For the 2-pool microbial model formulation (Fig. 4.1a) we used the model described in German et al. (2012), with altered calculations of half saturation constants

and temperature limitation of C uptake so as to make them comparable to the ones in the 4-pool microbial model (Allison *et al.*, 2010). For the 4-pool microbial model we used the model introduced by Allison et al. (2010) (Fig. 4.1b).

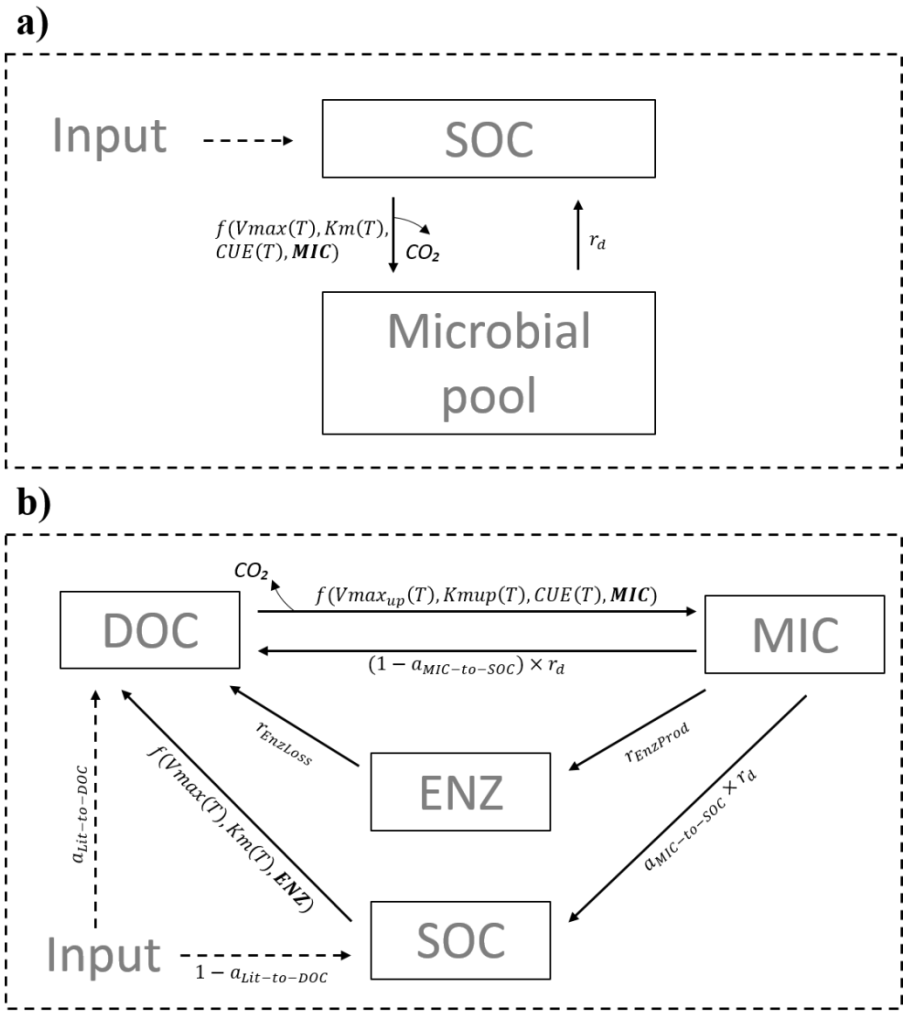


Figure 4.1 Schematic representation of the 2-pool (a) and 4-pool (b) microbial models

The two-pool microbial model by German et al. (2012) is

$$\frac{dMIC}{dt} = CUE \times V_{max} \times MIC \frac{SOC}{Km+SOC} - r_d \times MIC \quad (4.1)$$

$$\frac{dSOC}{dt} = Input_{soil} + r_d \times MIC - V_{max} \times MIC \frac{SOC}{K_m+SOC} \quad (4.2)$$

with

$$CUE = CUE_{slope} \times T_s - CUE_0 \quad (4.3)$$

$$V_{max} = V_{max_0} \times exp(-\frac{E_a}{R \times (T_s + 273)}) \times \exp(-par_{clay} \times clay) \quad (4.4)$$

$$\text{Km} = Km_{\text{slope}} \times T_s + Km_0 \times \exp(par_{lig} \times lignin) \quad (4.5)$$

where *MIC*, and *SOC* are microbial biomass and soil organic carbon pools (Fig. 4.1a); $Input_{soil}$ is carbon transferred to the soil from the litter pool; V_{max} is the temperature adjusted rate of SOC decomposition; K_m is the half-saturation constant for substrate limited soil organic carbon decomposition rate; r_d is the microbial death rate, and CUE is the microbial carbon use efficiency; T_s is soil temperature, *R* is gas constant (8.31 J/K/mol); CUE_0 and CUE_{slope} are baseline microbial carbon use efficiency and its dependency on temperature; V_{max_0} is maximum rate of microbial carbon uptake; E_a is activation energy of SOC decomposition; Km_0 and Km_{slope} are baseline half saturation constant and its dependency on temperature.

We complemented the model with exponential function of clay limitation of decomposition and estimated a parameter par_{clay} to test whether it were different from 0 in the microbial model formulation. Microbial respiration, when normalized by microbial biomass ($(1 - CUE) \times V_{max} \times \frac{SOC}{Km + SOC}$), has been reported to be non-linearly dependent on soil clay content (Müller & Höper, 2004). We also mimic the substrate quality limitation by adjusting baseline half-saturation constant by a lignin-dependent correction factor with its magnitude regulated by par_{lig} as there have been reports about substrate quality limitation of decomposition (Cusack *et al.*, 2009, Taylor *et al.*, 1989, Vance & Chapin Iii, 2001).

The 4-pool microbial model from Allison et al. (2010) is

$$\frac{dMIC}{dt} = V_{maxup} \times MIC \frac{DOC}{Kmup+DOC} \times CUE - r_d \times MIC - r_{EnzProd} \times MIC \quad (4.6)$$

$$\frac{dDOC}{dt} =$$

$$a_{lit\text{-}to\text{-}DOC} \times Input_{soil} + r_d \times MIC \times (1 - a_{MIC\text{-}to\text{-}SOC}) + V_{max} \times ENZ \frac{SOC}{Km+SOC} +$$

$$r_{EnzLoss} \times ENZ - V_{maxup} \times MIC \frac{DOC}{Kmup+DOC} \quad (4.7)$$

$$\frac{dSOC}{dt} = a_{lit\text{-}to\text{-}DOC} \times Input_{soil} + r_d \times MIC \times a_{MIC\text{-}to\text{-}SOC} - V_{max} \times ENZ \frac{SOC}{Km+SOC} \quad (4.8)$$

$$\frac{dENZ}{dt} = r_{EnzProd} \times MIC - r_{EnzLoss} \times ENZ \quad (4.9)$$

where *ENZ* and *DOC* are enzyme and dissolved organic carbon pools; V_{maxup} is temperature adjusted rate of *DOC* uptake by microbes; $Kmup$ is a half-saturation constant limiting microbial uptake of *DOC*; $r_{EnzProd}$ is a rate of enzyme production; $Input_{soil}$ is C transferred from litter to soil, and $a_{lit\text{-}to\text{-}DOC}$ is fraction of that input transferred to *DOC*; $a_{MIC\text{-}to\text{-}SOC}$ is fraction of dead microbes transferred to soil; and $r_{EnzLoss}$ is the rate of enzyme loss. Functions were dependent on temperature as follows:

$$CUE = CUE_{slope} \times T_s - CUE_0 \quad (4.10)$$

$$V_{maxup} = V_{maxup_0} \times exp(-\frac{E_{aup}}{R \times (T_s+273)}) \quad (4.11)$$

$$\text{Kmup} = Kmup_{\text{slope}} \times T_s + Kmup_0 \quad (4.12)$$

$$V_{max} = V_{max_0} \times exp(-\frac{E_a}{R \times (T_s+273)}) \times \exp(-par_{clay} \times clay) \quad (4.13)$$

$$\text{Km} = Km_{\text{slope}} \times T_s + Km_0 \times \exp(par_{lig} \times lignin) \quad (4.14)$$

where V_{maxup_0} was maximum rate of microbial *DOC* uptake; E_{aup} – activation energy of *DOC* uptake; $Kmup_0$ and $Kmup_{slope}$ were baseline half saturation constant for substrate limitation of *DOC* uptake and its dependency on temperature. As for the 2-pool model, we adjusted the rate of *SOC* degradation and the baseline half saturation constant of substrate limitation for *SOC* degradation by soil clay content and lignin content.

4.2.2 Data

We used two soil carbon datasets for Bayesian parameter estimation: a global map of total soil carbon content for the top 1 m of soil generated by International Geosphere-Biosphere Programme - Data and Information System [IGBP-DIS (Group, 2000)]; and a global map of soil microbial biomass distribution for the top 1 m (Xu *et al.*, 2013). The IGBP-DIS dataset has been widely used for production of new datasets (House *et al.*, 2002), the assessment of terrestrial C uptake (Freibauer *et al.*, 2004), for model evaluation (Delire *et al.*, 2003, Kucharik *et al.*, 2000), and for model improvement (Ise & Moorcroft, 2006, Smith *et al.*, 2013, Zhou *et al.*, 2009). Global microbial dataset has been used for parameterization of a biogeochemical model (Waring *et al.*, 2014). Prior to using the datasets in the data assimilation routine we randomly separated all the grid cells into two groups as in Smith et al. (2013) and Hararuk et al. (2014), and used one group for model parameter estimation, and the other – for evaluation to avoid overfitting.

The global litter lignin content was provided as part of the CLM-CASA' package: the plant-functional-type-level estimates of lignin were applied to MODIS-

derived distribution of plant functional types used in CLM (Lawrence & Chase, 2007, Oleson *et al.*, 2007). The map of soil clay content was originally developed by the International Geosphere-Biosphere Programme (Group, 2000) and was also provided as a part of CLM-CASA' package. To drive the models we used the 30-year averages of soil temperatures and soil C input produced by CLM-CASA'. Soil temperatures were calculated using air temperatures from global reanalysis data (Qian *et al.*, 2006), and soil C input was strongly correlated to MODIS NPP data (Fig. S4.1).

4.2.3 Parameter estimation

We calibrated parameters in the two microbial models using Bayesian probabilistic inversion. According to Mosegaard and Sambridge (2002) Bayesian inversion can be summarized as

$$p(\mathrm{c}|\mathrm{Z}) = v_c \times \mathrm{p}(\mathrm{Z}|\mathrm{c}) \times \mathrm{p}(\mathrm{c}) \quad (4.15)$$

where p(c|Z) is posterior probability density function of model parameters c; p(Z|c) is a likelihood function of parameters c; p(c) is prior probability density function of parameters c; and v_c is a normalization constant. We assumed that the prediction errors were normally distributed and uncorrelated, and calculated the likelihood function, p(Z|c), as

$$p(Z|c) = v_L \times \exp\left\{-\sum_{j=1}^{2}\sum_{i=1}^{k}\frac{\left(Z_{i,j}-X_{i,j}\right)^2}{2\sigma_{i,j}^2}\right\} \quad (4.16)$$

where $Z_{i,j}$ is total soil C reported by IGBP-DIS (*j*=1) or soil microbial biomass reported in Xu et al. (2013) (j=2) at *i*th gridcell, $X_{i,j}$ is total soil C or microbial biomass C simulated by the models at a corresponding gridcell; $\sigma^2{}_{i,j}$ is the variance of a *j*th measurement at *i*th gridcell; *k* is the total number of gridcells; and v_L is a constant.

We assigned minimum and maximum values to the parameters and used adaptive Metropolis (AM) algorithm (Haario *et al.*, 2001), a Markov Chain Monte Carlo method, to sample from the posterior parameter distributions. We generated a parameter chain by running AM algorithm in two steps: a proposing step and a moving step. In the proposing step a new parameter set c^{new} was generated from a previously accepted parameter set c^{k-1} through a proposal distribution $(c^{new}|c^{k-1})$. In the moving step a probability of acceptance $P(c^{k-1}|c^{new})$ was calculated as in (Marshall *et al.*, 2004):

$$P(c^{k-1}\big|c^{new}) = min\left\{1, \frac{p(Z|c^{new})p(c^{new})}{p(Z|c^{k-1})p(c^{k-1})}\right\} \tag{4.17}$$

The value of $P(c^{k-1}|c^{new})$ was then compared with a random number *U* from 0 to 1. Parameter set c^{new} was accepted if $P(c^{k-1}|c^{new}) \geq U$, otherwise c^k was set to c^{k-1}.

The AM algorithm required an initial parameter covariance matrix, which we generated from a test run of 50,000 simulations with uniform proposal distribution as in Xu et al. (2006):

$$c^{new} = c^{k-1} + r \times \frac{c^{max} - c^{min}}{D} \tag{4.18}$$

where c^{max} and c^{min} are upper and lower parameter limits *r* is a random number between -0.5 to 0.5, and D=5. We constructed a covariance matrix C_0 on the basis of the test run and modified the proposal step to be

$$c^{new} = N(c^{k-1}, C_k) \tag{4.19}$$

$$C_k = \begin{cases} C_0 & k \leq k_0 \\ s_d Cov(c_0, \dots, c_{k-1}) & k > k_0 \end{cases} \tag{4.20}$$

where k_0 = 2000; $s_d = 2.38/\sqrt{8}$ for the 2-pool model and $s_d = 2.38/\sqrt{15}$ for the 4-pool model (Gelman *et al.*, 1996).

We made five parallel runs (each run containing 500,000 simulations) starting at dispersed initial points in the parameter space. During a simulation we equated eqs. 4.1-4.2 or eqs. 4.6-4.9 (depending on the model) to zero and solved the model for the two (or four) carbon pool sizes at each gridcell [semi-analytical spin-up approach (Xia *et al.*, 2012)] using 30-year averages of soil temperatures simulated by CLM-CASA' and calibrated soil C input from Hararuk et al. (2014). Because we did not have global time-variant data of soil carbon and microbial biomass pools, we assumed that spatial relationships of total soil C and microbial biomass C with environmental factors would represent the temporal relationships, and that year-to-year changes in soil C pools were close to zero. Such approach has been used before by Ise and Moorcroft (2006) and Smith et al. (2013). We discarded the first half of the simulations (as burn-in phase) and tested the second half for convergence to stationary distributions with Gelman-Rubin diagnostics (Gelman & Rubin, 1992).

4.2.4 Forward model runs and stability analysis

To evaluate the uncertainties in soil C feedbacks to climate change we ran the calibrated microbial models forward, driving them with a climate change scenario (increasing CO_2 and temperatures) and sampling from the posterior parameter distributions. We used the Community Earth System Model (CESM) output for the Representative Concentration Pathway 8.5 (RCP8.5) experiment (specifically, the simulated temperature and soil C influx) to drive soil C pools. The CESM model output

was provided as a part of Coupled Model Intercomparison Project Phase 5 (CMIP5), and could be accessed at http://pcmdi9.llnl.gov. Over 95 years CESM simulated a 3.5 K increase in mean global temperature and atmospheric CO_2 increase to 1150 ppm by the year 2100 (Keppel-Aleks *et al.*, 2013). We first used the 2006-2010 data to generate initial pools using the semi-analytical model spin-up approach (Xia *et al.*, 2012), then ran the microbial models forward in time to the year 2100, generating soil C feedbacks to the changing climate, and compared them to the feedbacks produced by a calibrated conventional model (Hararuk *et al.*, 2014).

Once we ran the forward simulations, it became evident that we needed to test the microbial models for oscillations, as the uncertainties in feedbacks of the two microbial models differed drastically. We calculated the periods of oscillations along with the time required to damp the oscillations following the technique described in Wang et al. (2013) and Svirezhev (2002): (1) we calculated the matrix of sensitivities (Jacobians) of pool changes to the pools sizes for the microbial models; (2) calculated eigenvalues of the Jacobians; (3) calculated period of oscillations from the imaginary component of an eigenvalue as $p = \frac{2\pi}{i}$, where p was the oscillation period, and *i* was the imaginary component of an eigenvalue; (4) calculated the approximate times required to damp the oscillations as $t \approx (-r)^{-1}$, where *r* was the real part of an eigenvalue (negative if the models were convergent).

4.3 Results

4.3.1 Performance of the microbial model formulations after calibration

We calibrated two microbial models using global gridded observed total SOC data product (IGBP-DIS), and global gridded microbial biomass C (Xu *et al.*, 2013). As indicated in the methods section, we used soil C input from the calibration study in Hararuk et al (2014) (Figure 4.2a) to drive the models. Soil C input was closely associated with a 7-year average of MODIS NPP (r^2=0.75, Fig. S4.1), and was 45% of NPP values, which, given the equilibrium assumption, implied that 55% of the incoming C was returned back to the atmosphere via respiration from the litter pools.

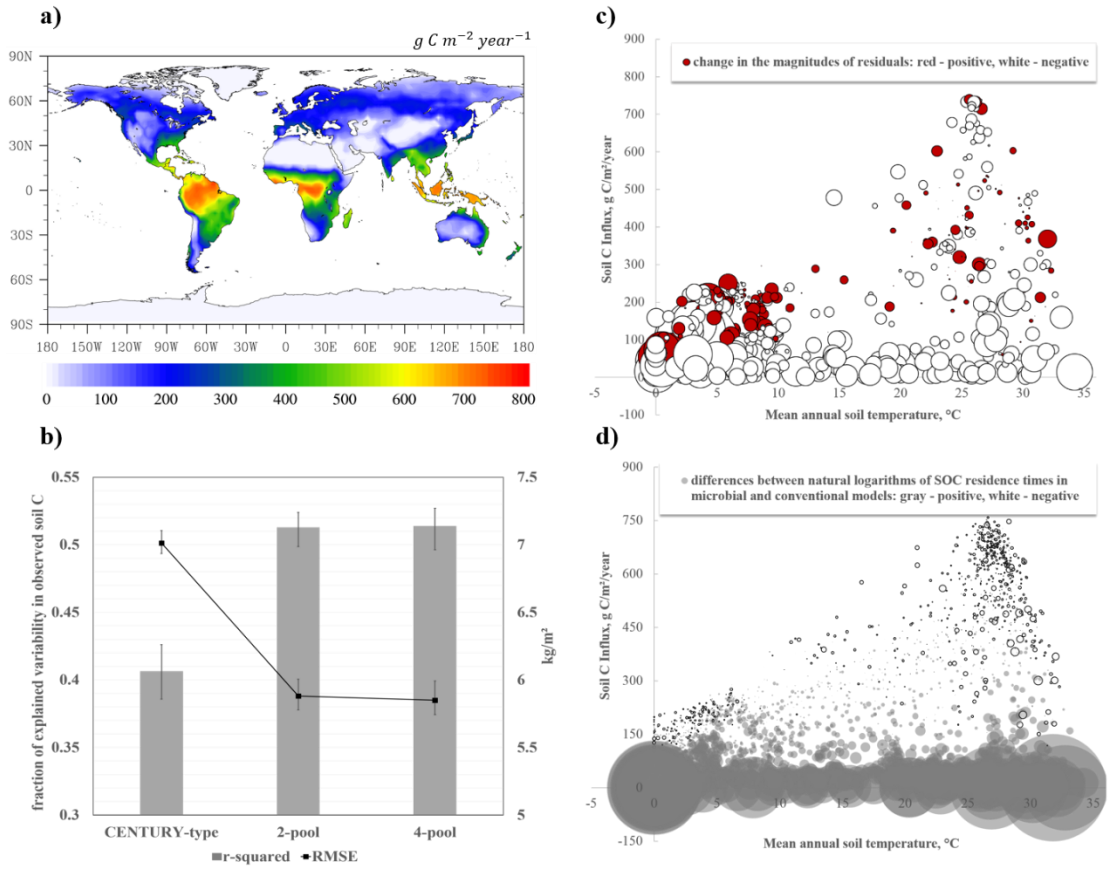


Figure 4.2 Annual soil C influx used to drive the soil sub-models (a); performance of calibrated microbial and conventional models (b); distribution of changes in the

residuals' magnitudes after switching from conventional to microbial model formulation in the environmental space (c): circle diameters represent the relative magnitudes of change, and colors indicate the direction of change; differences in the natural logarithms of calibrated soil residence times between microbial and conventional models (d): circle diameters represent the relative magnitudes of change, and colors indicate the direction of change

After calibration of the microbial models with observed total SOC and microbial biomass C we saw a higher fraction of explained variability in the SOC data than a fraction explained by the calibrated conventional model (Figure 4.2b). Additionally, microbial models had lower spatial RMSE than the calibrated conventional model. As indicated by RMSE's and r-squares, two microbial models produced similar SOC distributions, therefore, we make further total SOC fit comparisons using a conventional model and one of the microbial models giving the full illustrations in the Supplementary Information.

Microbial models performed better than the calibrated conventional model in terms of soil C prediction in the low-temperature regions and in the regions with small soil C inputs (Figures 4.2c, S4.2). Because soil C pools were determined by the soil C inputs and residence times (Luo *et al.*, 2003) and the former were identical for conventional and microbial model formulations the differences in SOC predictions were caused by the differences in the SOC residence times. In the conventional models the spatial patterns of SOC residence times are determined mainly by temperature (Todd-Brown *et al.*, 2013b) whereas in microbial models residence times are controlled by

both temperature and SOC input (mediated by microbial biomass change, Fig. 4.1). Fresh carbon input stimulates microbial biomass growth, which increases the rate of old SOC decomposition [priming effect, (Fontaine *et al.*, 2004, Fontaine *et al.*, 2007, Kuzyakov *et al.*, 2000)], therefore in the areas with high SOC input microbial models simulated lower SOC residence times than the conventional model (Fig. 4.2d, S4.3). In the regions with low SOC input, SOC residence times in the microbial models were higher than predicted by the conventional model. This was due to nonlinearity of substrate limitation in the microbial models (eqs. 4.1-4.2, and 4.6-4.8) [conventional model assumed linear effect of substrate limitation on the microbial activity (Parton *et al.*, 1993)], as well as the dependency of residence times on microbial biomass. As illustrated in Figure 4.2c and d, change in assumptions about SOC residence times most often led to the decrease in the magnitudes of models' residuals, and therefore better model performance.

Both 2-pool and 4-pool microbial models explained 30% of spatial variability of the observed microbial biomass C after calibration (Fig. 4.3). In most gridcells the values were underpredicted with the largest underpredictions located in the low-temperature regions with low annual soil C input. Interestingly, the largest overpredictions were located in the similar climatic zones. Such pattern suggests the presence of controls not included in the model, such as the distribution of plant functional type other than assumed in the model, water table controls (the model does not include peatlands), or effects of various disturbances on microbial biomass.

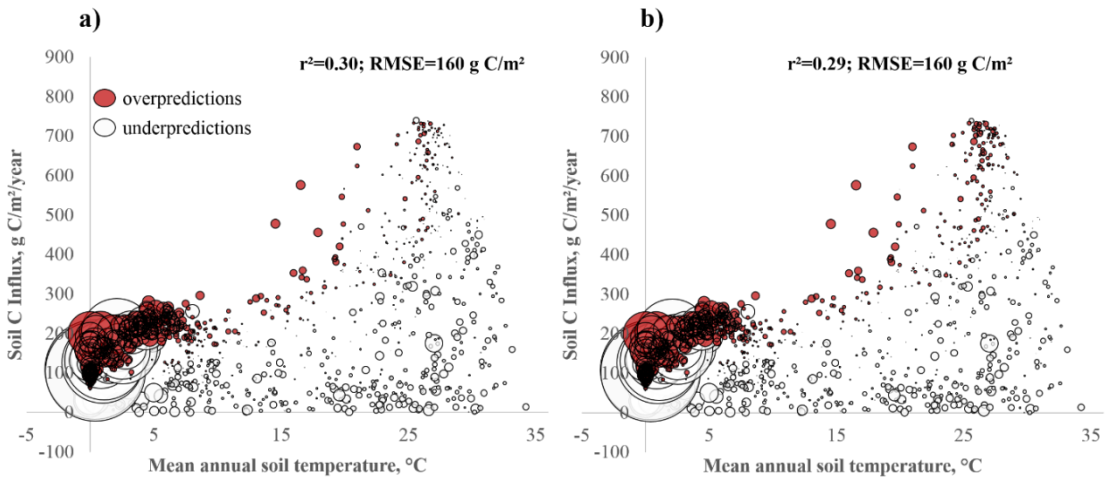
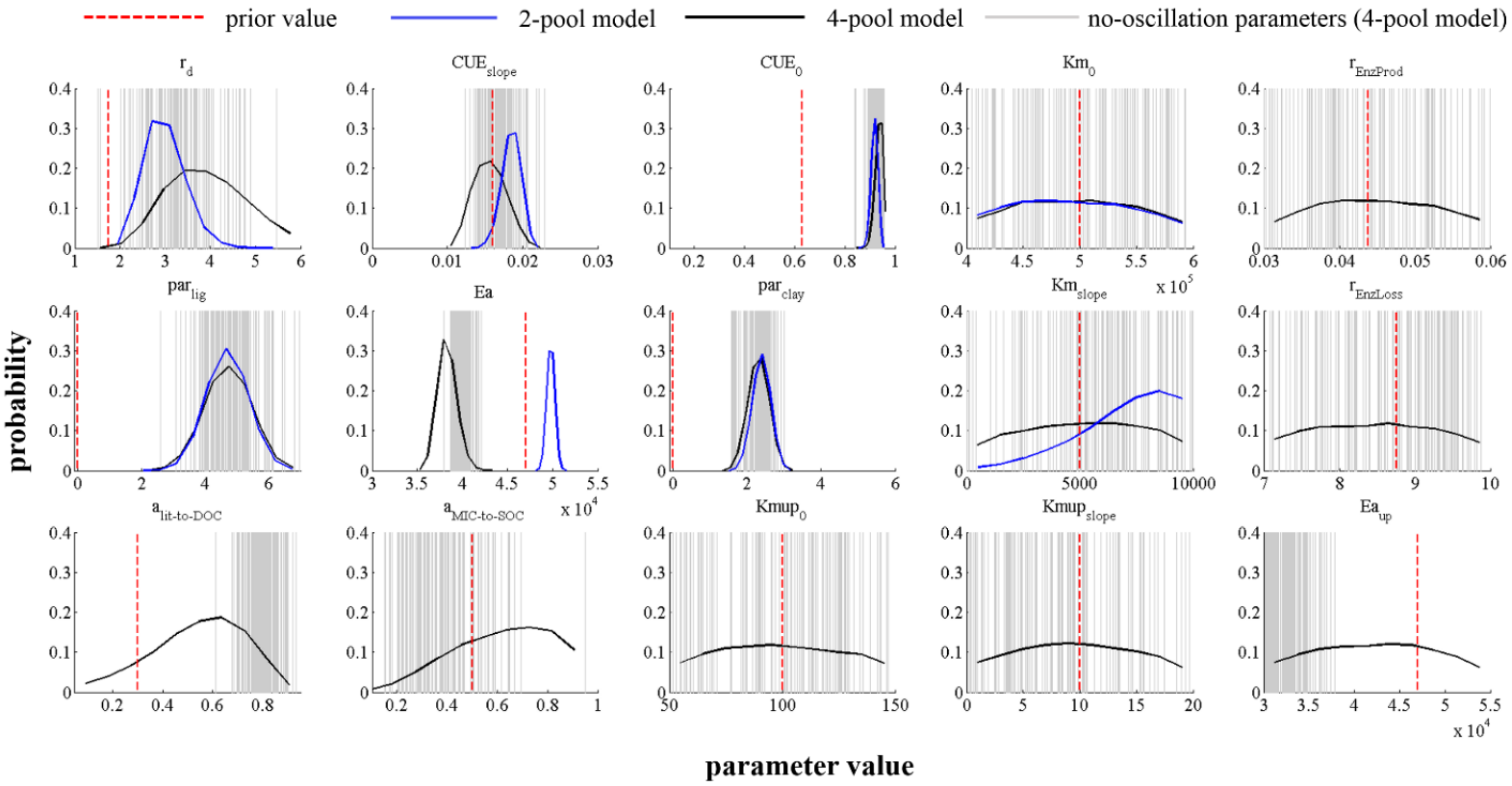


Figure 4.3 Model residuals for microbial biomass plotted in the environmental space: (a) 2-pool microbial model; (b) 4-pool microbial model. Circle diameters represent the relative magnitudes of the residuals.

4.3.2 Posterior parameter distributions and correlations

Many parameters were constrained by assimilating total SOC and microbial biomass C data into the two microbial models (Fig. 4.4). The most sensitive parameters to assimilation of the observed data were microbial death rate; degree of temperature dependency of microbial carbon use efficiency; baseline microbial carbon use efficiency; activation energy; lignin effect on the half saturation constant of SOC decomposition; and clay effect on the rate of the SOC decomposition. Unconstrained parameters were associated with substrate limitation of C pools' decomposition, temperature sensitivity of microbial DOC uptake, and dynamics of the enzyme pool, which indicated the need for assimilation of C flux data as according to the model equations 4.1-4.2 and 4.6-4.8 flux data would contribute additional information on temperature sensitivity and substrate limitation of microbial activity.

Figure 4.4 Posterior probability density functions of the microbial models' parameters



Lignin and clay effects on SOC decomposition, originally not included in the model, were significantly larger than 0 and converged to the same values in both microbial models (Fig. 4.4). No observations were available to evaluate the lignin regulation for half-saturation constant, however our estimates for par_{clay} were close to the range calculated from observations. Observed par_{clay} ranged from 1.94 (Wang *et al.*, 2003) to 3.02 (Müller & Höper, 2004), whereas the 95% confidence intervals (CI) for the 2-pool model were 1.99-2.8, and 1.86- 2.78 for the 4-pool model (Table S4.1). Temperature dependence of microbial carbon use efficiency was larger than 0 and was 0.016-0.02 $°C^{-1}$ for the 2-pool model, and 0.012-0.019$°C^{-1}$ for the 4-pool model, indicating that microbes were likely to adapt to rising temperatures, which would result in lower SOC loss than if carbon use efficiency were constant. The CUE adaptation rates estimated by the 4-pool model were close to the observed range of 0.01-0.014 $°C^{-1}$ (Devêvre & Horwáth, 2000, Frey *et al.*, 2013, Steinweg *et al.*, 2008).

Due to structural differences between the two microbial model formulations, some parameters shared by both models converged at different values altering SOC responses to changing temperatures (Figure 4.5 a, b). Lower temperature dependency of microbial use efficiency and lower activation energy of SOC decomposition in the 4-pool model than in the 2-pool model led to lower temperature sensitivities (Q10's) in the former. Unlike the conventional model which assumed Q10 was constant across space, both microbial models simulated spatially variable Q10's with higher values in the low-temperature regions and lower Q10's in the high-temperature regions – a widely reported pattern indicative of temperature acclimation of microbial activity (Chen & Tian, 2005, Luo *et al.*, 2001a, Peng *et al.*, 2009). In the microbial models the

acclimation was caused by substrate limitation and acclimating CUE (Fig. S4.4). High CUE facilitated increase in the microbial pool, and since respiration was proportional to microbial biomass, large microbial pool produced strong feedback to elevated temperature. Apart from temperature, microbial biomass was regulated by substrate: if there was insufficient amount of available substrate for the maximum potential biomass increase for the given temperature, the temperature sensitivity decreased (Fig. S.4.4), a phenomenon also observed in the field (Hartley *et al.*, 2007).

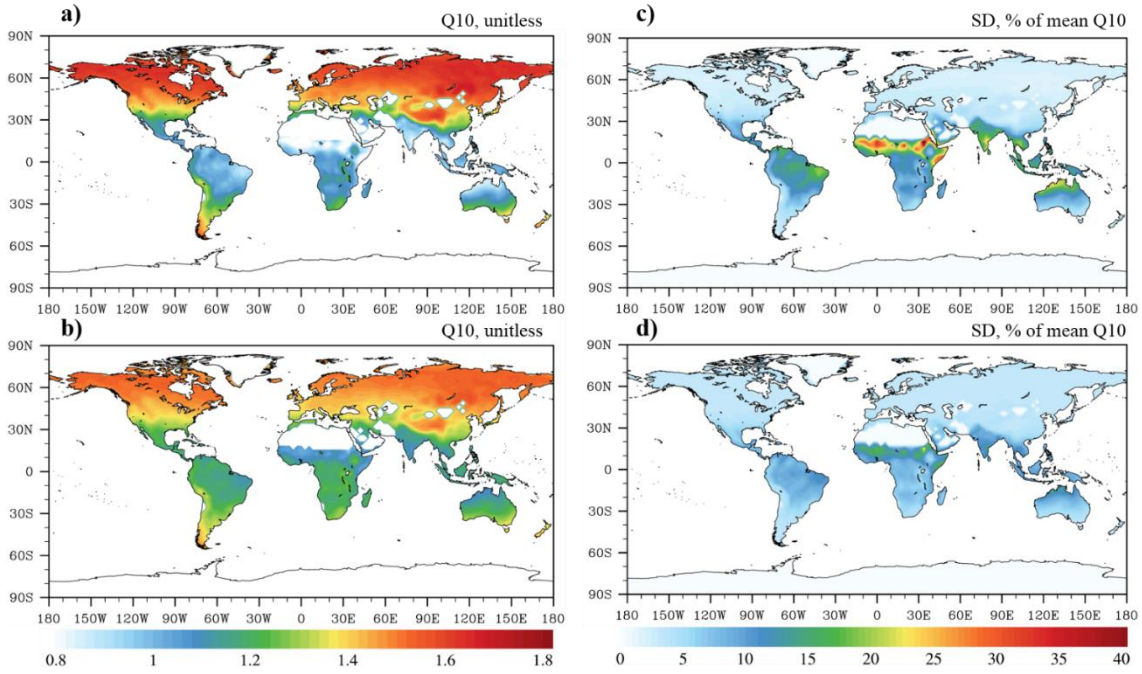


Figure 4.5 We calculated Q10's for the microbial models as ratios of SOC turnover rates at temperatures T+10°C to SOC turnover rates at temperatures T. Spatial distribution of the temperature sensitivities of SOC turnover rates and their uncertainties expressed as normalized standard deviations (SD) in the 2-pool microbial model (a, c) and 4-pool microbial model (b, d). Constrained Q10 for the conventional model was 1.86 with SD =4% of the mean, and was constant.

Out of 28 parameter pairs in the 2-pool model, only 5 pairs were strongly correlated (Table 4.1): microbial death rate was negatively correlated with activation energy of SOC decomposition, positively – with baseline carbon use efficiency, and negatively – with temperature dependency of carbon use efficiency. The correlated parameters regulated C influx and outflux in the microbial pool, and since we did not have any influx or outflux data to include into model calibration the parameters became correlated. Baseline CUE and CUE dependency on temperature were negatively correlated which was expected due to the CUE equation formulation (eqs. 4.3, 4.10). Activation energy of SOC decomposition was negatively correlated with baseline CUE, and since the two parameters had the opposing effects on regulating microbial C uptake, we attributed the correlation to lack of the observed data to separate the two processes.

Table 4.1 Parameter correlations in the 2-pool microbial model

Parameters	**Correlations**							
r_d,	1.00							
CUE_{slope}	**-0.75**	1.00						
CUE_0	**0.93**	**-0.61**	1.00					
Km_0	-0.07	0.08	-0.06	1.00				
Km_{slope}	0.33	-0.43	0.30	-0.04	1.00			
par_{lig}	-0.10	0.11	-0.07	-0.04	0.11	1.00		
Ea	**-0.63**	0.46	**-0.60**	-0.48	-0.27	-0.36	1.00	
par_{clay}	0.18	-0.23	0.16	0.01	-0.06	-0.19	-0.36	1.00
	r_d	CUE_{slope}	CUE_0	Km_0	Km_{slope}	par_{lig}	Ea	par_{clay}

Out of 105 parameter pairs in the 4-pool model, only 4 pairs were correlated (Table 4.2). As in the 2-pool model, microbial death rate was positively correlated with the baseline CUE and negatively – with the degree of temperature dependency of CUE; two latter parameters were also negatively correlated with each other. Activation energy of SOC decomposition was negatively correlated with the fraction of dead microbes transferred to SOC pool, which was also caused by assimilation of C pools' data and the absence of C flux data.

Table 4.2 Parameter correlations in the 4-pool microbial model

Parameters	Correlations														
r_d	1.00														
$r_{EnzProd}$	0.00	1.00													
$r_{EnzLoss}$	-0.01	-0.01	1.00												
$a_{Lit\text{-}to\text{-}DOC}$	-0.42	-0.08	0.04	1.00											
$a_{MIC\text{-}to\text{-}SOC}$	-0.09	-0.01	0.02	-0.40	1.00										
CUE_{slope}	**-0.89**	-0.01	0.01	0.47	0.09	1.00									
CUE_0	**0.93**	0.02	-0.01	-0.39	-0.09	**-0.81**	1.00								
$Kmup_{slope}$	0.00	-0.01	-0.02	0.02	-0.01	0.00	0.00	1.00							
$Kmup_0$	-0.02	-0.03	0.01	-0.01	0.01	0.02	-0.02	0.03	1.00						
Km_{slope}	0.01	0.02	0.02	0.00	0.00	0.01	0.00	0.01	-0.03	1.00					
Km_0	0.01	0.02	0.00	0.34	0.05	-0.02	0.00	0.02	0.00	-0.02	1.00				
par_{lig}	0.02	-0.02	-0.01	-0.07	0.01	-0.02	0.02	0.01	-0.01	-0.03	0.15	1.00			
Ea_{up}	-0.01	0.02	0.00	0.00	0.00	0.00	-0.01	0.00	0.01	-0.02	0.00	0.01	1.00		
Ea	-0.44	0.39	-0.22	0.49	**-0.65**	0.39	-0.41	0.00	0.00	-0.21	-0.05	-0.23	0.00	1.00	
par_{clay}	0.04	0.02	-0.01	0.15	0.04	-0.03	0.03	0.01	0.00	0.02	-0.04	-0.17	0.03	-0.14	1.00
	r_d	$r_{EnzProd}$	$r_{EnzLoss}$	$a_{Lit\text{-}to\text{-}DOC}$	$a_{MIC\text{-}to\text{-}SOC}$	CUE_{slope}	CUE_0	$Kmup_{slope}$	$Kmup_0$	Km_{slope}	Km_0	par_{lig}	Ea_{up}	Ea	par_{clay}

4.3.3 SOC feedbacks to climate change

To illustrate the differences between soil C feedbacks between conventional and microbial models, we ran the models forward using RCP8.5 climate change scenario. The initial soil C pool sizes differed between conventional and microbial models (Fig. 4.6 a, b, c): calibrated microbial models predicted a higher global soil C content than a calibrated conventional model (by 180-200 Pg C), and the predictions of all soil C cycle model formulations fell within the 95% CI of the observed SOC content [890–1660 PgC (Todd-Brown *et al.*, 2013b)]. Because the initial pools were different among the models, we illustrated soil C sensitivities to climate change as cumulative relative soil C changes. In microbial models soil C was more sensitive to climate change than in the conventional model: 2-pool and 4-pool microbial models simulated 8% and 11% cumulative SOC loss and conventional model simulated a 2.5% SOC loss after 95 years of climate change.

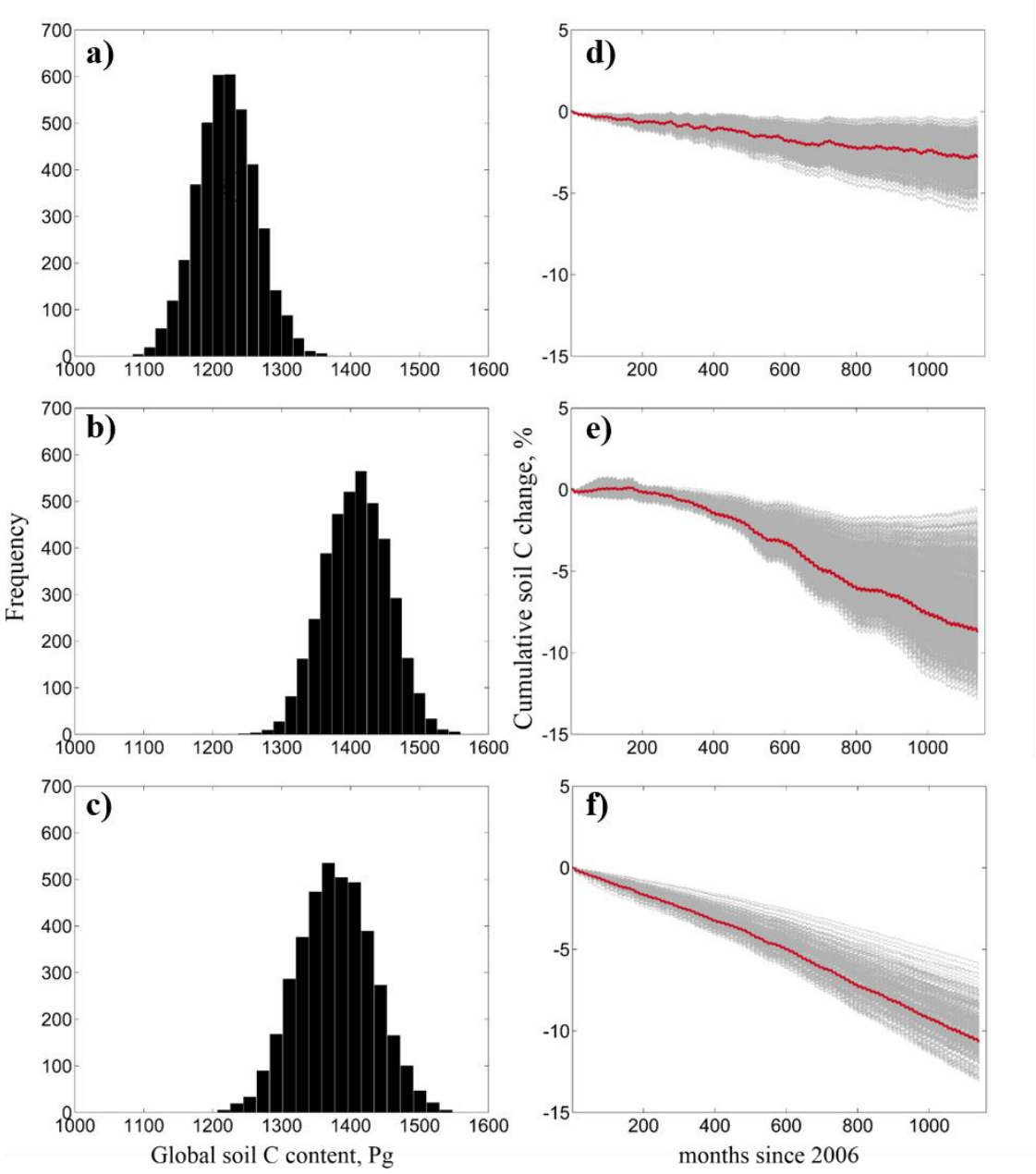


Figure 4.6 Global frequency distributions of total SOC pools produced by a calibrated CENTURY-type model [**a**, from Hararuk et al., (2014)], 2-pool and 4-pool microbial models (**c** and **d** respectively), and the cumulative SOC changes under RCP8.5 scenario (**b-d**, red lines are maximum likelihood cumulative changes, and gray lines are sample runs, representing uncertainty)

Under RCP8.5 scenario after 95 years of climate change mean annual global temperatures increased by 3.5°C (Keppel-Aleks *et al.*, 2013), and annual global soil C input increased by 20% (Fig. S4.5). As we mentioned earlier, the SOC residence times in the microbial models were regulated not only by temperatures, but also by SOC input. Temperature increase along with the increase in SOC input stimulated microbial biomass growth, and therefore increased SOC decomposition. In the conventional model fresh C input resulted in soil C accumulation, and only increase in temperatures caused SOC loss. Increases in both temperature and SOC input led to a higher relative SOC loss in the microbial models than in the conventional model, despite the lower Q10's of the former.

Among the three models the 2-pool microbial model had the highest uncertainty in SOC feedbacks to RCP8.5 scenario (Fig. 4.6e), ranging from 1.5% to 12.5% soil C loss after 95 years of climate change. This uncertainty did not result from higher uncertainties in the 2-pool model parameters than in the 4-pool model parameters (Fig. 4.4). Although convergent, microbial models often have oscillatory dynamics, with the period of oscillations dependent on parameter values (Wang *et al.*, 2013). To test whether our models had oscillatory dynamics, we calculated eigenvalues of the microbial models' Jacobians and derived the periods of oscillations for parameter chains as in Wang et al. (2013). For the 2-pool model the average global period of oscillations was 57 years (Fig. 4.7a) with the range of around 38 years and about 140 years required to damp the oscillations. Low ratio of oscillation period to convergence time was likely the cause of the large range of cumulative SOC feedbacks as the predictions were falling on different points of the oscillations.

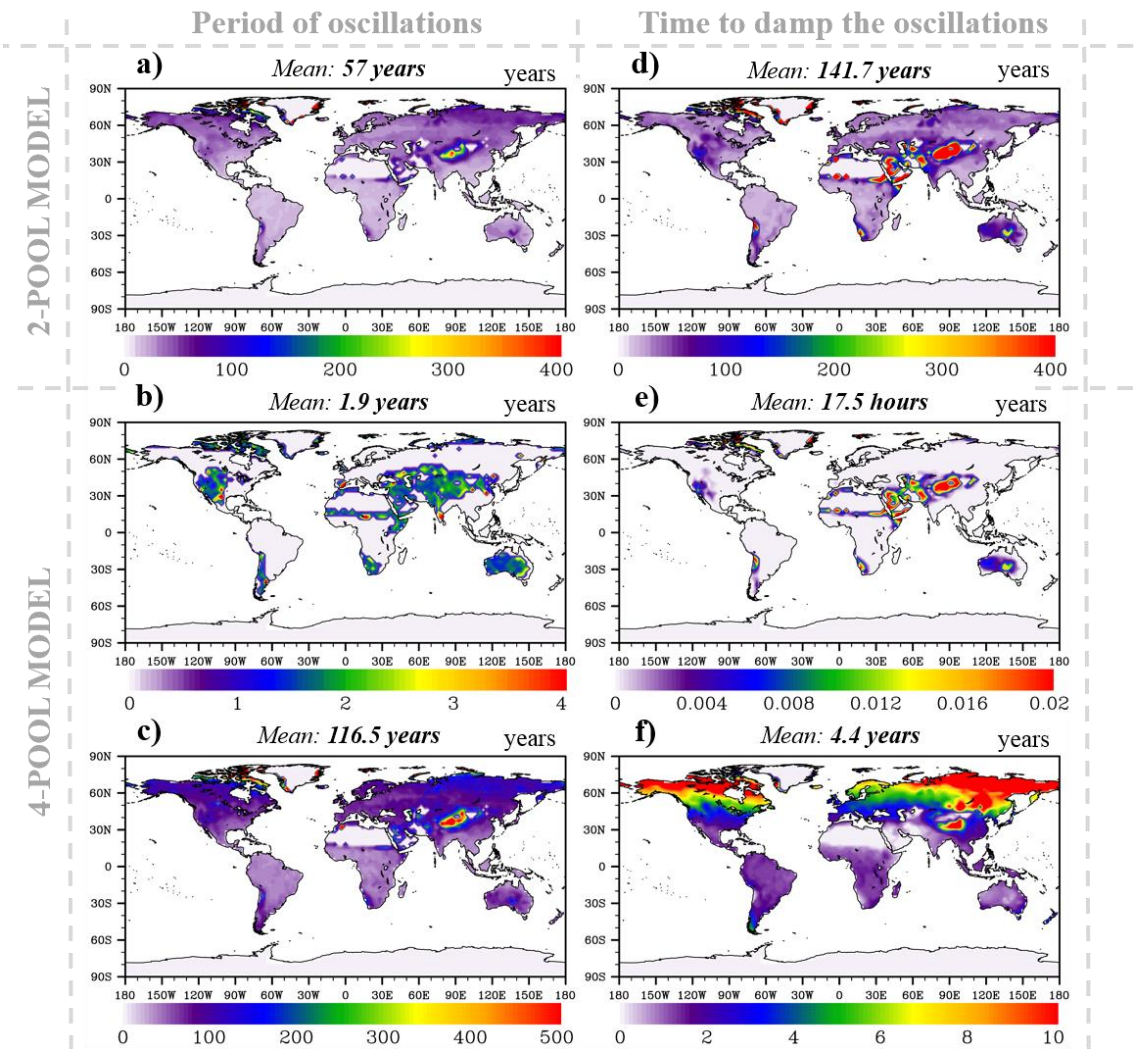


Figure 4.7 Spatial distribution of oscillations in the 2-pool (a), and 4-pool (b, c) models, and time required to damp the oscillations (d-f). 4-pool model had higher oscillation period to convergence time ratio than two pool model, which indicated that oscillations will be damped early in their evolution.

The uncertainty in soil C feedbacks in the 4-pool model was not as high as in the 2-pool model because the average periods of oscillations in the former were larger and convergence times were shorter: 1.9 years and 116.5 years, with the times required to damp the oscillations 17.5 hours and 4.4 years respectively (Fig. 4.7 b,e,c,f). Moreover,

in posterior parameter space for the 4-pool model there were parameter combinations that did not produce oscillations (Fig. 4.4, gray lines), and from the parameter sample density it was evident that there were thresholds in activation energies, temperature dependence of microbial carbon use efficiency, and partitioning of the C flux from litter to DOC after which the model started to have oscillatory dynamics. High ratios of oscillation period to their convergence times and existence of sample runs with no oscillations led to lower uncertainty in SOC feedbacks in the 4-pool model than in the 2-pool model.

4.4 Discussion

4.4.1 Progress in simulating SOC dynamics

Most carbon cycle models have been simulating SOC decomposition as a first-order decay process with the variability in the models' parameterization causing substantial uncertainties in SOC prediction (Todd-Brown *et al.*, 2013b). In the recent years studies have been focusing on reducing this uncertainty by applying data-model fusion techniques to constrain the model parameters (Hararuk et al., 2014, Ise & Moorcroft, 2006, Smith et al., 2013, Zhou et al., 2009). Still, data-constrained conventional models were not representing spatial SOC distribution well (Hararuk et al., 2014, Smith et al., 2013). Additionally, data-informed model parameters were often approaching unrealistic values (Hararuk et al., 2014), which indicated the potential flaws in the conventional model formulation.

Microbial models have shown potential to simulate global SOC dynamics better than conventional models (Wieder *et al.*, 2013), and our results illustrated that SOC

representation by microbial models was better than that of conventional models once both model formulations were calibrated against data. However, better representation of SOC distribution may not indicate that microbial models are better than conventional models as microbial models have been criticized for producing unrealistic SOC dynamics, such as oscillations in SOC pools (Wang *et al.*, 2013). While there are reports of oscillations in microbial biomass and respiration observed in the incubation studies (Cenciani *et al.*, 2008, Lloyd *et al.*, 1982), the phenomenon needs to be investigated further prior to being considered true for natural ecosystems.

In our study oscillations in the SOC pools were present in both 2-pool and 4-pool microbial models (Fig. 4.7), however, they were less prominent in the 4-pool model due to large oscillation periods and short convergence times. Moreover, many parameter combinations in the 4-pool model did not generate oscillations in the SOC pools (Fig. 4.4), therefore 4-pool microbial model did not have indications of unrealistic dynamics. Improved global SOC simulation and realistic soil C pool dynamics make a 4-pool microbial model a better-performing alternative to conventional models for simulating soil carbon dynamics.

4.4.2 Uncertainties in future SOC dynamics

Future climate change can be either mitigated or worsened by terrestrial feedbacks. For the past two decades soils have been slowing the rate of atmospheric CO_2 increase sequestering over 0.45 Pg C per year (Pan *et al.*, 2011), however it remains uncertain whether soils will continue to sequester C under changing temperatures and CO_2 concentrations. CMIP5 models are not unanimous in simulating

soil responses to the RCP8.5 scenario, ranging from a 7% loss of soil C after 95 years of climate change to a 22.6% gain in SOC (Todd-Brown *et al.*, 2013a). A data-constrained conventional model predicted a 1-6% loss in soil C after 95 years (Hararuk *et al.*, 2014), which was within the CMIP5 range. The calibrated microbial models, however, simulated much stronger negative feedbacks of soil C to RCP8.5 scenario than conventional models: a 2-13% loss (2-pool model) and 6-13% (4-pool model) loss in SOC.

In the conventional models SOC pools are positively affected by increasing C influx, and negatively – by increasing temperatures as they decrease SOC residence times (Xia *et al.*, 2013b). These relationships imply that temperature-induced negative SOC feedbacks to climate change can be diminished or even reversed by increasing C influx. Analysis of SOC feedbacks produced by CMIP5 models revealed that increase in soil C input had more weight in determining SOC feedbacks than increase in temperatures, revealing a C input increase threshold of around 30%, after which models simulated increase in SOC in response to climate change (Todd-Brown *et al.*, 2013a).

In microbial models C input decreased SOC residence times, which exacerbated the negative effect of increasing temperatures on SOC storage (Fig.S4.3). Because we used only CESM SOC input to drive the microbial model, the uncertainty in SOC feedbacks was caused by uncertainties in parameters or the presence of oscillations (in the 2-pool model). Among the parameters causing the largest uncertainties in SOC feedbacks temperature acclimation rate of CUE has received the most attention (Allison *et al.*, 2010, Wieder *et al.*, 2013). Our data-constrained CUE acclimation rates had more

narrow ranges than those explored in the literature, and therefore were unlikely to be the largest source of uncertainty.

Microbial turnover rates had wide posterior ranges (2.33-3.78 years^{-1} for the 2-pool model and 2.59-5.36 years^{-1} for the 4-pool model), and being the only parameter that regulated depletion of the residence time regulating microbial pool, was likely the cause of large uncertainties particularly in the 4-pool model. Another poorly-constrained parameter that caused substantial uncertainties in 4-pool model was $a_{lit\text{-}to\text{-}DOC}$ as the pool to which C influx was directed was reported to cause large uncertainties on the SOC feedbacks (Allison *et al.*, 2010).

4.4.3 Future improvements

Although the microbial models (particularly, the 4-pool model) demonstrate the best model performance in simulating global SOC distribution to date, there is much room for model improvement as 48% of variability in global total SOC and 70% of variability in global microbial biomass distribution remain unexplained. The nature of environmental limitation on microbial dynamics remains uncertain: CUE has been reported to decrease with increasing temperatures, however, under high temperatures the rate of acclimation tends to decrease rather than remain constant (Devêvre & Horwáth, 2000, Frey *et al.*, 2013, Wetterstedt & Ågren, 2011). Additionally, CUE may be dependent not only on temperature, but also on organic matter quality (Frey *et al.*, 2013, Sinsabaugh *et al.*, 2013). More studies are needed to investigate the controls over microbial death rate and partitioning of SOC input between DOC and SOC pools.

Oscillatory dynamics observed in microbial models by Wang et al. (2013) and in this study is rarely observed in nature and needs to be further investigated, as if present in natural ecosystems it will cause large uncertainties in the future soil C feedbacks. If not present in the natural ecosystems oscillatory dynamics can be avoided in several ways. One option is to calibrate microbial models with an additional constraint so that parameters generating oscillations are discarded. Another option is to modify conventional models so that they mimic the processes in the microbial models, such as acclimation of carbon use efficiency to temperatures and dependence of SOC residence times on C input rates.

Still, microbial models may give the right answer for global SOC distribution for the wrong reason, therefore future improvements in soil C cycle modeling need not focus on microbial models exclusively. For instance, lower organic matter residence times in microbial models than in the conventional models in the regions with high C influx rates are explained by priming effect. However, lower than expected residence times may also be due to the activity of soil animals as it has been reported to increase organic matter turnover in the regions with high C influx (Wall *et al.*, 2008). In the wetland ecosystems, which are abundant in the northern regions, residence times are limited not only by temperatures, but also by low oxygen content (Davidson & Janssens, 2006), therefore including oxygen limitation of microbial activity in the conventional models may improve model performance.

4.5 Conclusion

We calibrated two global soil microbial models against the global observed total soil C and microbial biomass C datasets using a Bayesian MCMC approach. By adjusting model parameters we were able to substantially improve representation of global total SOC compared to a calibrated conventional model and non-calibrated conventional models. When propagated in the model run under RCP8.5 scenario posterior parameter uncertainties resulted in a 2-13 % SOC loss over 95 years, with maximum likelihoods of 8% (2-pool model) and 11% (4-pool model) loss which was larger than SOC loss projected by 11 CMIP5 models.

As we show in this study, data assimilation facilitates fairly accurate SOC stocks simulation by microbial models, often without a cost of unrealistic oscillatory dynamics. Unlike conventional models, microbial models simulated spatially variable temperature sensitivities of SOC decomposition with the variability pattern similar to the observed one. Assimilating data into microbial models also narrowed the global ranges for the acclimation rate of microbial CUE; established the data-informed range for soil clay content limitation on C uptake rate, as well as the range for substrate quality limitation of C uptake by microbes. More observations are needed to quantify the non-linearity of temperature effect on microbial CUE; quantify the substrate quality effect on microbial CUE; establish the controls for partitioning of the SOC input between fast and passive pools; and validate oscillations in the microbial biomass dynamics. These observations will help constrain C cycle feedbacks and provide reliable assessments of the climate change effects on terrestrial ecosystems, as well as facilitate development of the mitigation strategies for the climate change effects.

Table S4.1 Model parameter description, prior ranges, posterior ranges, and their convergence indices. MLE is maximum likelihood estimate, and G-R is Gelman-Rubin diagnostics

Parameter description	Symbol	Prior			MLE	G-R	Lower 95% bounds	Upper 95% bounds
		minimum	maximum	Value				
2-pool microbial model								
Microbial death rate, years^{-1}	r_d	1.00	6.00	4.38	2.80	1.00	2.33	3.78
Temperature dependency of carbon use efficiency, °C^{-1}	CUE_{slope}	0.00	0.03	0.02	0.018	1.00	0.016	0.02
Baseline carbon use efficiency	CUE_0	0.10	1.00	0.63	0.91	1.00	0.90	0.94
Baseline half-saturation constant for SOC decay, g C/m²	Km_0	400000	600000	500000	477799	1.00	407916	579353
Temperature dependency of half-saturation constant for SOC decay, g C/m²/°C	Km_{slope}	0	10000	5000	6681	1.00	2806	9619
Degree of lignin effect on Km_0	par_{lig}	-0.10	7.00	0.00	4.45	1.00	3.63	5.82
Activation energy of SOC decay, J/mol	Ea	30000	55000	47000	49626	1.00	49097	50570
Degree of clay effect on soil microbial respiration	par_{clay}	-0.10	6.00	0.00	2.30	1.00	1.99	2.80
4-pool microbial model								
Microbial death rate, years^{-1}	r_d	1.00	6.00	1.75	3.60	1.00	2.59	5.36
Rate of enzyme production, years^{-1}	$r_{EnzProd}$	0.03	0.06	0.04	0.04	1.00	0.03	0.06
Rate of enzyme loss, years^{-1}	$r_{EnzLoss}$	7.00	10.00	8.76	8.23	1.00	7.12	9.71
Fraction of C influx transferred from litter to DOC	$a_{lit\text{-}to\text{-}DOC}$	0.05	0.95	0.30	0.49	1.00	0.22	0.82
Fraction of dead microbes transferred to SOC	$a_{MIC\text{-}to\text{-}SOC}$	0.10	1.00	0.50	0.64	1.00	0.33	0.94
Temperature dependency of carbon use efficiency, °C^{-1}	CUE_{slope}	0.00	0.03	0.02	0.015	1.00	0.012	0.019
Baseline carbon use efficiency	CUE_0	0.10	1.00	0.63	0.93	1.00	0.91	0.96
Baseline half-saturation constant for DOC decay	$Kmup_0$	50	150	100	91	1.00	55	141
Temperature dependency of half-saturation constant for DOC decay, g C/m²/°C	$Kmup_{slope}$	0.00	20.00	10	8.15	1.00	0.99	18
Baseline half-saturation constant for SOC decay, g C/m²	Km_0	400000	600000	500000	481485	1.00	409794	580115
Temperature dependency of half-saturation constant for SOC decay, g C/m²/°C	Km_{slope}	0.00	10000	5000	4358	1.00	496	9068
Degree of lignin effect on Km_0	par_{lig}	-0.10	7.00	0.00	4.46	1.00	3.52	5.95
Activation energy of DOC decay, J/mol	Ea_{up}	30000	55000	47000	40267	1.00	31149	52434
Activation energy of SOC decay, J/mol	Ea	30000	55000	47000	37904	1.00	36726	40195
Degree of clay effect on soil microbial respiration	par_{clay}	-0.10	6.00	0.00	2.21	1.00	1.86	2.78

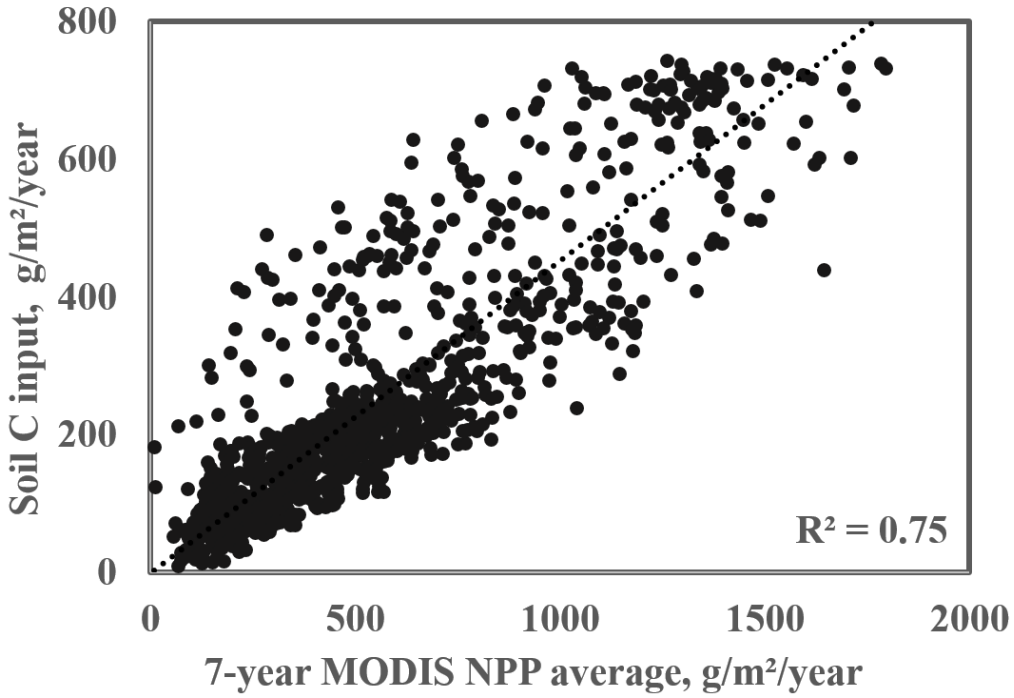


Figure S4.1 Comparison of CLM-CASA' soil C input with a 7-year average of MODIS data (Zhao & Running, 2010)

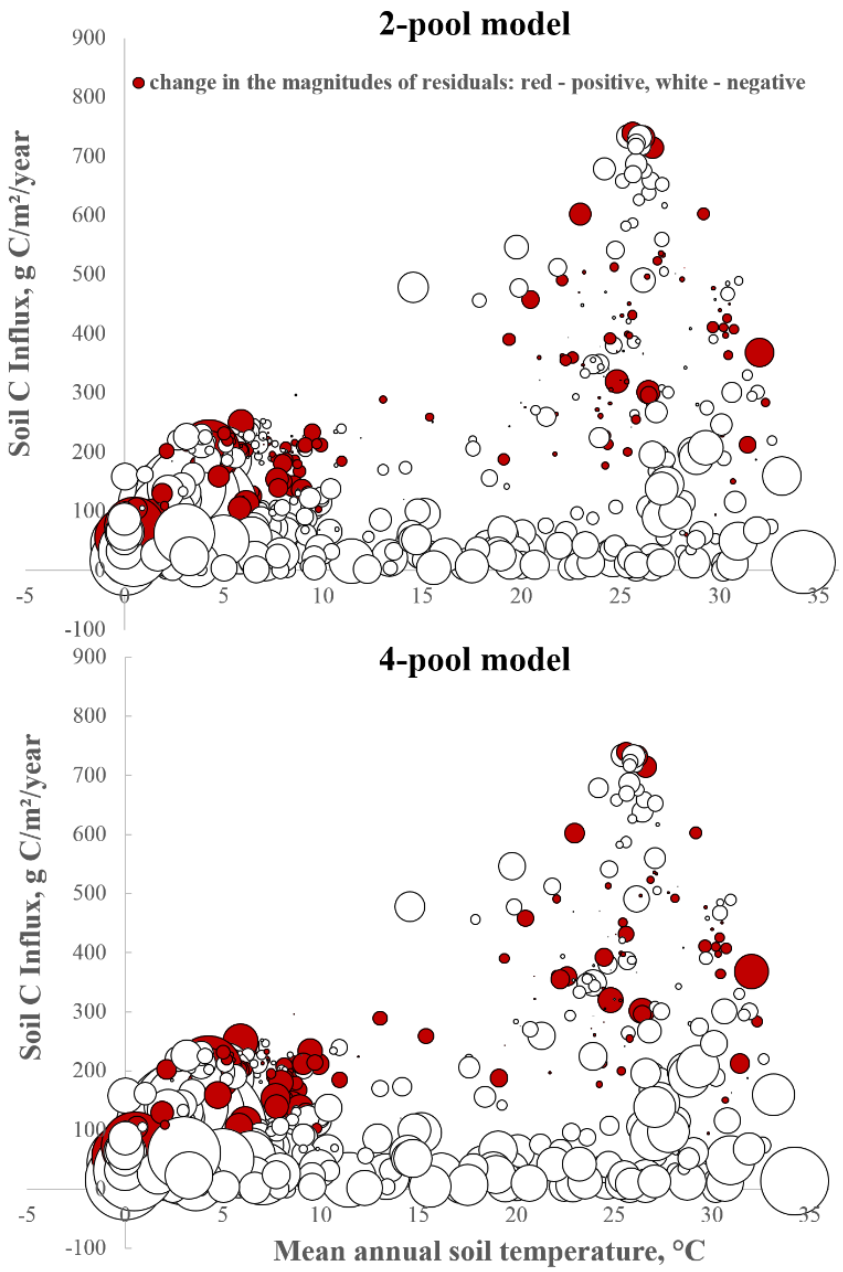


Figure S4.2 Differences in the magnitudes of the residuals between calibrated microbial models and conventional models in the environmental space. The circle diameters represent relative magnitudes of the residuals.

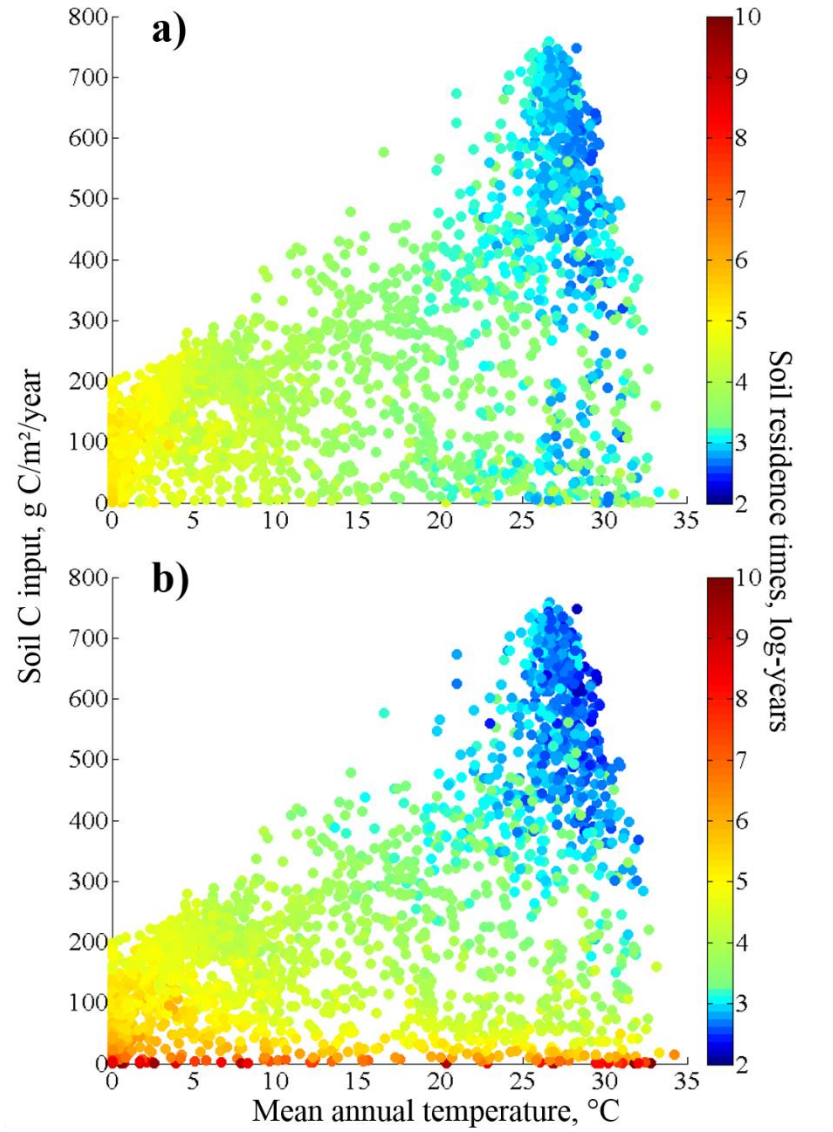


Figure S4.3 Distribution of soil C residence times in the environmental space simulated by a calibrated conventional model (a), and microbial models (b).

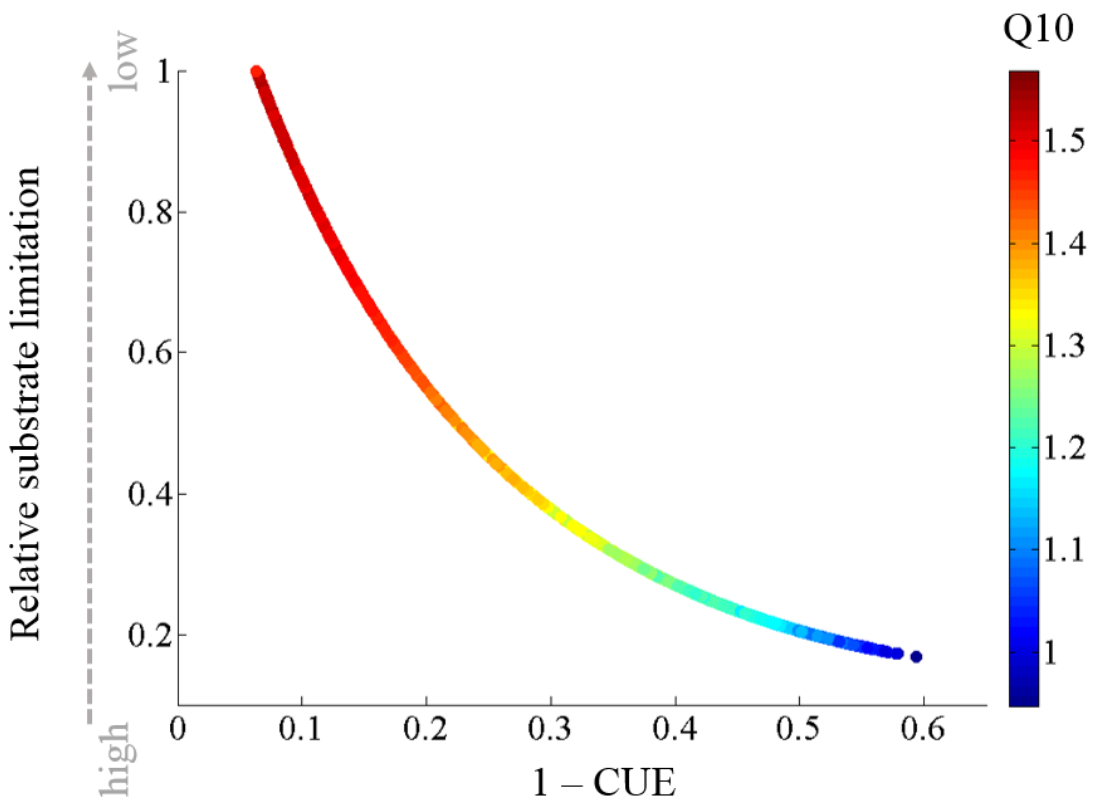


Figure S4.4 Change of temperature sensitivity of soil C residence time with respect to microbial carbon use efficiency and substrate limitation.

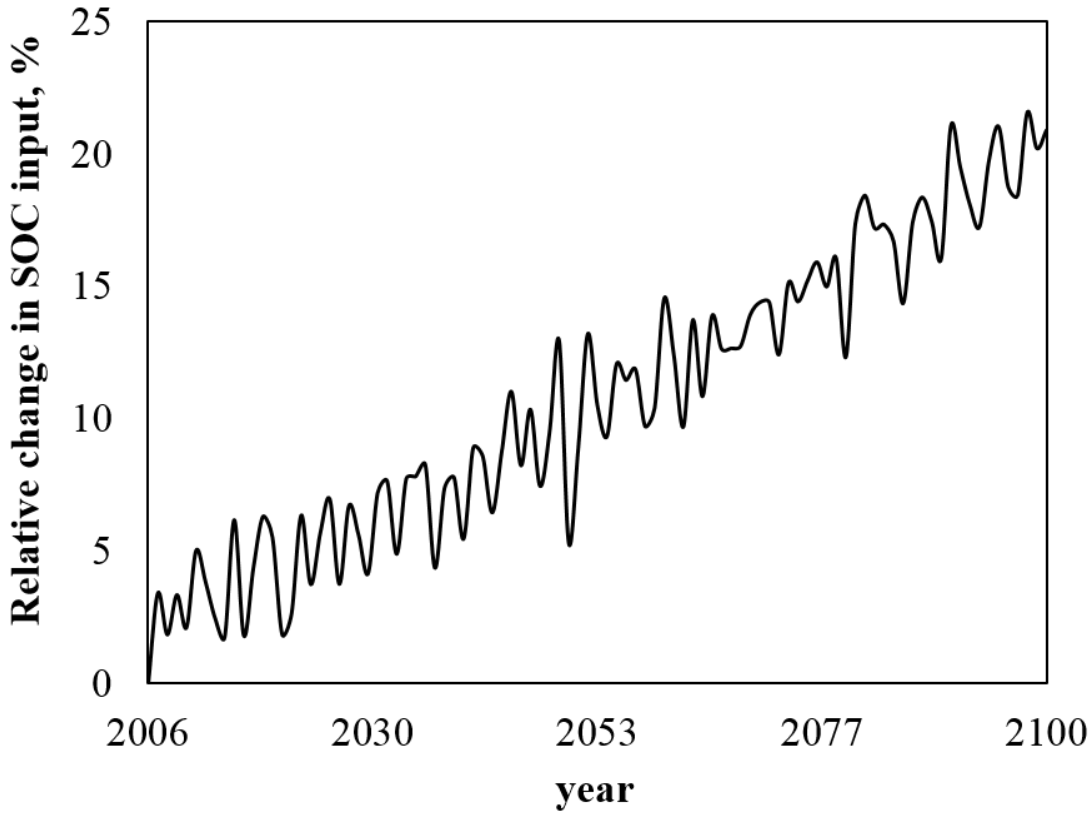


Figure S4.5 Relative change in SOC input in response to the RCP8.5 scenario simulated by CESM.

Chapter 5

Conclusions and Implications

5.1 Conclusions

The studies conducted in this dissertation demonstrate that data assimilation provides many benefits when applied to global carbon cycle models. Calibration helped substantially improve the initially poorly-performing carbon cycle models: the performance of the model for litter turnover rates improved by 47% and performance of soil C cycle models improved by 13% and 23% once microbial biomass dynamics was included. Assimilating data helped identify models with flawed assumptions: unrealistic parameter values indicated that litter turnover rate was limited by litter lignin to nitrogen ratios rather than structural lignin content, and that CLM-CASA' was unlikely to fully capture soil C dynamics.

The data-constrained model for leaf litter turnover rates reversed the direction of feedbacks of global litter pool to changing climate, projecting future accumulation in leaf litter, increasing the ecosystem carbon storage capacity. However, the best-performing (microbial) soil carbon model projected the strongest negative feedback among 11 CMIP5 models. Lastly, posterior parameter distributions allowed to construct data-informed uncertainties for litter and soil C feedbacks to future climate change.

5.2 Implications for future work

Data assimilation studies in Chapters 2-4 reveal several implications for future modeling and observational studies. Functions representing turnover rates for litter and soil carbon pools in C cycle models often share parameter values. However, constrained temperature sensitivities, Q10's, in Chapters 2 and 3 significantly differ from each other

implying that temperature affects litter and soil C differently, therefore should be represented by different parameters and/or functions in carbon cycle models.

Chapter 4 illustrates that soil microbial models perform substantially better than conventional, CENTURY-type, models, and the oscillatory dynamics can be avoided if 4-pool rather than 2-pool microbial model is used. Temperature sensitivities in the microbial models differed from those in the conventional models: the Q10's in conventional models were constant across the globe, whereas microbial models simulated higher temperature sensitivities in the low temperature regions than those in the high-temperature regions – a pattern also observed in nature (Chen & Tian, 2005, Peng et al., 2009). This pattern can be reproduced in the conventional models by making partitioning of carbon among different pools (equivalent of carbon use efficiency in the microbial models) dependent on temperature. Future modeling studies should explore modifying conventional models to include features from the microbial models, such as variable temperature sensitivities, because conventional models do not have unrealistic properties such as oscillatory responses to small perturbations.

As illustrated in Chapter 3, the most uncertain regions for soil C predictions are located in the northern and tropical regions, therefore more effort should be put into studying the controls on C cycle in those regions. There is also high uncertainty in prediction of the leaf litter residence times in the northern regions, which may be caused by the effects of microbial community composition, or by inaccurate litter quality input data used in CESM. More observations of leaf litter quality are needed to construct a reliable dataset for use in models, and more research is needed to establish whether microbial community composition significantly affects litter turnover rates.

References

Aber JD, Melillo JM, Mcclaugherty CA (1990) Predicting long-term patterns of mass-loss, nitrogen dynamics, and soil organic matter formation from initial fine litter chemistry in temperate forest ecosystems. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **68**, 2201-2208.

Abramowitz G, Leuning R, Clark M, Pitman A (2008) Evaluating the Performance of Land Surface Models. *Journal of Climate*, **21**, 5468-5481.

Adair EC, Parton WJ, Del Grosso SJ *et al.* (2008) Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates. *Global Change Biology*, **14**, 2636-2660.

Allison SD, Wallenstein MD, Bradford MA (2010) Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience*, **3**, 336-340.

Austin AT, Ballaré CL (2010) Dual role of lignin in plant litter decomposition in terrestrial ecosystems. *Proceedings of the National Academy of Sciences*, **107**, 4618-4622.

Ayres E, Steltzer H, Simmons BL *et al.* (2009) Home-field advantage accelerates leaf litter decomposition in forests. *Soil Biology and Biochemistry*, **41**, 606-610.

Batjes N (1995) A homogenized soil data file for global environmental research: A subset of FAO, ISRIC and NCRS profiles.

Batjes NH (2009) Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Management*, **25**, 124-127.

Batjes NH (2014) Batjes, N. H. 1996. Total carbon and nitrogen in the soils of the world. European Journal of Soil Science, 47, 151–163. Reflections by N.H. Batjes. *European Journal of Soil Science,* **65**, 2-3.

Beare MH, Parmelee RW, Hendrix PF, Cheng W, Coleman DC, Crossley DA, Jr. (1992) Microbial and Faunal Interactions and Effects on Litter Nitrogen and Decomposition in Agroecosystems. *Ecological Monographs,* **62**, 569-591.

Bond-Lamberty B, Thomson A (2010) A global database of soil respiration data. *Biogeosciences*, **7**, 1915-1926.

Bond-Lamberty B, Thomson A (2012a) A Global Database of Soil Respiration Data, Version 2.0. *Data set. Available on-line [http://daac. ornl. gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA http://dx. doi. org/10.3334/ORNLDAAC/1070*.

Bond-Lamberty BP, Thomson AM (2012b) A Global Database of Soil Respiration Data. http://daac.ornl.gov, Oak Ridge National Laboratory Distributed Active Archive Center.

Carney KM, Hungate BA, Drake BG, Megonigal JP (2007) Altered soil microbial community at elevated CO2 leads to loss of soil carbon. *Proceedings of the National Academy of Sciences,* **104**, 4990-4995.

Cenciani K, Freitas SDS, Critter SaM, Airoldi C (2008) Microbial enzymatic activity and thermal effect in a tropical soil treated with organic materials. *Scientia Agricola*, **65**, 674-680.

Chen H, Tian H-Q (2005) Does a General Temperature-Dependent Q10 Model of Soil Respiration Exist at Biome and Global Scale? *Journal of Integrative Plant Biology,* **47**, 1288-1302.

Cotrufo MF, Ineson P (1996) Elevated CO2 reduces field decomposition rates of Betula pendula (Roth.) leaf litter. *Oecologia,* **106**, 525-530.

Cotrufo MF, Ineson P, Rowland AP (1994) Decomposition of tree leaf litters grown under elevated CO2: Effect of litter quality. *Plant and Soil,* **163**, 121-130.

Coûteaux M-M, Kurz C, Bottner P, Raschi A (1999) Influence of increased atmospheric CO2 concentration on quality of plant material and litter decomposition. *Tree Physiology,* **19**, 301-311.

Coûteaux M-M, Mousseau M, Célérier M-L, Bottner P (1991) Increased Atmospheric CO2 and Litter Quality: Decomposition of Sweet Chestnut Leaf Litter with Animal Food Webs of Different Complexities. *Oikos,* **61**, 54-64.

Cox PM, Betts RA, Jones CD, Spall SA, Totterdell IJ (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature,* **408**, 184-187.

Cusack DF, Chou WW, Yang WH, Harmon ME, Silver WL, The Lidet T (2009) Controls on long-term root and leaf litter decomposition in neotropical forests. *Global Change Biology,* **15**, 1339-1355.

Davidson EA, Janssens IA (2006) Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature,* **440**, 165-173.

Delire C, Foley JA, Thompson S (2003) Evaluating the carbon cycle of a coupled atmosphere-biosphere model. *Global Biogeochem. Cycles,* **17**, 1012.

Devêvre OC, Horwáth WR (2000) Decomposition of rice straw and microbial carbon use efficiency under different soil temperatures and moistures. *Soil Biology and Biochemistry*, **32**, 1773-1785.

Esser G, Aselmann I, Lieth H (1982) Modelling the carbon reservoir in the system compartment 'litter'. *Mitt. Geol. Paleontol. Inst. Univ. Hamburg*, **52**, 39-58.

Falkowski P, Scholes RJ, Boyle E *et al.* (2000) The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science,* **290**, 291-296.

Fang C, Smith P, Moncrieff JB, Smith JU (2005) Similar response of labile and resistant soil organic matter pools to changes in temperature. *Nature,* **433**, 57-59.

Fao (1995) Digital Soil Map of the World and Derived Soil Properties (Version 3.5).

Finzi AC, Schlesinger WH (2002) Species control variation in litter decomposition in a pine forest exposed to elevated CO2. *Global Change Biology*, **8**, 1217-1229.

Foley JA, Costa MH, Delire C, Ramankutty N, Snyder P (2003) Green surprise? How terrestrial ecosystems could affect earth's climate. *Frontiers in Ecology and the Environment,* **1**, 38-44.

Fontaine S, Bardoux G, Abbadie L, Mariotti A (2004) Carbon input to soil may decrease soil carbon content. *Ecology Letters,* **7**, 314-320.

Fontaine S, Barot S, Barre P, Bdioui N, Mary B, Rumpel C (2007) Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature*, **450**, 277-280.

Freibauer A, Rounsevell MDA, Smith P, Verhagen J (2004) Carbon sequestration in the agricultural soils of Europe. *Geoderma,* **122**, 1-23.

Frey SD, Lee J, Melillo JM, Six J (2013) The temperature response of soil microbial efficiency and its feedback to climate. *Nature Clim. Change*, **3**, 395-398.

Friedlingstein P, Cox P, Betts R *et al.* (2006) Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, **19**, 3337-3353.

Fung IY, Doney SC, Lindsay K, John J (2005) Evolution of carbon sinks in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 11201-11206.

Gelman A, Roberts G, Gilks W (1996) Efficient metropolis jumping hules. *Bayesian statistics*, **5**, 599-608.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat. Sci*, **7**, 457– 511.

German DP, Marcelo KRB, Stone MM, Allison SD (2012) The Michaelis–Menten kinetics of soil extracellular enzymes in response to temperature: a cross-latitudinal study. *Global Change Biology*, **18**, 1468-1479.

Gholz HL, Wedin DA, Smitherman SM, Harmon ME, Parton WJ (2000) Long-term dynamics of pine and hardwood litter in contrasting environments: toward a global model of decomposition. *Global Change Biology*, **6**, 751-765.

Group GSDT (2000) Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). [Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System)]. Data set., Oak Ridge, TN, USA, Oak Ridge National Laboratory Distributed Active Archive Center.

Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. *Bernoulli*, 223-242.

Hararuk O, Xia J, Luo Y (2014) Evaluation and Improvement of a Global Land Model against Soil Carbon Data Using a Bayesian MCMC Method. *Journal of Geophysical Research: Biogeosciences*, 2013JG002535.

Harmon ME, Silver WL, Fasth B *et al.* (2009) Long-term patterns of mass loss during the decomposition of leaf and fine root litter: an intersite comparison. *Global Change Biology,* **15**, 1320-1338.

Harmon R, Challenor P (1997) A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecological Modelling,* **101**, 41-59.

Hartley IP, Heinemeyer A, Ineson P (2007) Effects of three years of soil warming and shading on the rate of soil respiration: substrate availability and not thermal acclimation mediates observed response. *Global Change Biology,* **13**, 1761-1770.

Hobbie SE (1996) Temperature and Plant Species Control Over Litter Decomposition in Alaskan Tundra. *Ecological Monographs,* **66**, 503-522.

Hobbie SE, Schimel JP, Trumbore SE, Randerson JR (2000) Controls over carbon storage and turnover in high-latitude soils. *Global Change Biology*, **6**, 196-210.

Houghton JT, Ding Y, Griggs DJ *et al.* (2001) *Climate change 2001: the scientific basis*, Cambridge University Press Cambridge.

House JI, Colin Prentice I, Le Quéré C (2002) Maximum impacts of future reforestation or deforestation on atmospheric CO2. *Global Change Biology,* **8**, 1047-1052.

Ise T, Moorcroft PR (2006) The global-scale temperature and moisture dependencies of soil organic carbon decomposition: an analysis using a mechanistic decomposition model. *Biogeochemistry,* **80**, 217-231.

Jones C, Mcconnell C, Coleman K, Cox P, Falloon P, Jenkinson D, Powlson D (2005) Global climate change and soil carbon stocks; predictions from two contrasting models for the turnover of organic carbon in soil. *Global Change Biology,* **11**, 154-166.

Jones CD, Cox P, Huntingford C (2003) Uncertainty in climate–carbon-cycle projections associated with the sensitivity of soil respiration to temperature. *Tellus B,* **55**, 642-648.

Jones CD, Cox PM (2001) Constraints on the temperature sensitivity of global soil respiration from the observed interannual variability in atmospheric CO2. *Atmospheric Science Letters,* **2**, 166-172.

Kätterer T, Bolinder MA, Andrén O, Kirchmann H, Menichetti L (2011) Roots contribute more to refractory soil organic matter than above-ground crop residues, as revealed by a long-term field experiment. *Agriculture, Ecosystems & Environment,* **141**, 184-192.

Kätterer T, Reichstein M, Andrén O, Lomander A (1998) Temperature dependence of organic matter decomposition: a critical review using literature data analyzed with different models. *Biology and Fertility of Soils,* **27**, 258-262.

Keenan TF, Davidson E, Moffat AM, Munger W, Richardson AD (2012) Using model-data fusion to interpret past trends, and quantify uncertainties in future

projections, of terrestrial ecosystem carbon cycling. *Global Change Biology*, **18**, 2555-2569.

Keppel-Aleks G, Randerson JT, Lindsay K *et al.* (2013) Atmospheric Carbon Dioxide Variability in the Community Earth System Model: Evaluation and Transient Dynamics during the Twentieth and Twenty-First Centuries. *Journal of Climate*, **26**, 4447-4475.

Kirschbaum MUF (1995) The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biology and Biochemistry*, **27**, 753-760.

Kirschbaum MUF, Paul KI (2002) Modelling C and N dynamics in forest soils with a modified version of the CENTURY model. *Soil Biology and Biochemistry*, **34**, 341-354.

Köne AÇ, Büke T (2010) Forecasting of CO2 emissions from fuel combustion using trend analysis. *Renewable and Sustainable Energy Reviews*, **14**, 2906-2915.

Koven C, Friedlingstein P, Ciais P, Khvorostyanov D, Krinner G, Tarnocai C (2009) On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model. *Geophysical Research Letters*, **36**.

Koven CD, Ringeval B, Friedlingstein P *et al.* (2011) Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 14769-14774.

Krull ES, Baldock JA, Skjemstad JO (2003) Importance of mechanisms and processes of the stabilisation of soil organic matter for modelling carbon turnover. *Functional Plant Biology,* **30**, 207-222.

Kucharik CJ, Foley JA, Delire C *et al.* (2000) Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure. *Global Biogeochem. Cycles,* **14**, 795-825.

Kuppel S, Peylin P, Chevallier F, Bacour C, Maignan F, Richardson AD (2012) Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, **9**, 3757-3776.

Kuzyakov Y, Friedel JK, Stahr K (2000) Review of mechanisms and quantification of priming effects. *Soil Biology and Biochemistry,* **32**, 1485-1498.

Lawrence CR, Neff JC, Schimel JP (2009) Does adding microbial mechanisms of decomposition improve soil organic matter models? A comparison of four models using data from a pulsed rewetting experiment. *Soil Biology and Biochemistry*, **41**, 1923-1934.

Lawrence D, Oleson KW, Flanner MG *et al.* (2011a) Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems,* **3**, 27 pp.

Lawrence DM, Oleson KW, Flanner MG *et al.* (2011b) The CCSM4 Land Simulation, 1850–2005: Assessment of Surface Climate and New Capabilities. *Journal of Climate,* **25**, 2240-2260.

Lawrence PJ, Chase TN (2007) Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *Journal of Geophysical Research: Biogeosciences,* **112**, G01023.

Liski J, Palosuo T, Peltoniemi M, Sievänen R (2005) Carbon and decomposition model Yasso for forest soils. *Ecological Modelling,* **189**, 168-182.

Liu L, King JS, Giardina CP (2005) Effects of elevated concentrations of atmospheric CO2 and tropospheric O3 on leaf litter production and chemistry in trembling aspen and paper birch communities. *Tree Physiology,* **25**, 1511-1522.

Lloyd D, Edwards SW, Fry JC (1982) Temperature-compensated oscillations in respiration and cellular protein content in synchronous cultures of Acanthamoeba castellanii. *Proceedings of the National Academy of Sciences,* **79**, 3785-3788.

Luo Y (2007) Terrestrial carbon-cycle feedback to climate warming. In: *Annual Review of Ecology Evolution and Systematics.* pp 683-712. Palo Alto, Annual Reviews.

Luo Y, Hui D, Zhang D (2006) Elevated CO2 stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. *Ecology,* **87**, 53-63.

Luo Y, Mooney HA (1995) Stimulation of global photosynthetic carbon influx by an increase in atmospheric carbon dioxide concentration In: *Carbon dioxide and terrestrial ecosystems.* (ed Mooney GWKaHA) pp 381–397. San Diego, CA, USA, Academic Press.

Luo Y, Wan S, Hui D, Wallace LL (2001a) Acclimatization of soil respiration to warming in a tall grass prairie. *Nature,* **413**, 622-625.

Luo Y, Weng E (2011) Dynamic disequilibrium of the terrestrial carbon cycle under global change. *Trends in Ecology & Evolution,* **26**, 96-104.

Luo Y, Weng E, Wu X, Gao C, Zhou X, Zhang L (2009) Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecological Applications*, **19**, 571-574.

Luo Y, White LW, Canadell JG *et al.* (2003) Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach. *Global Biogeochem. Cycles*, **17**, 1021.

Luo Y, Wu L, Andrews JA, White L, Matamala R, V. R. Schäfer K, Schlesinger WH (2001b) Elevated CO2 differentiates ecosystem carbon processes: deconvolution analysis of Duke Forest FACE data. *Ecological Monographs,* **71**, 357-376.

Luo Y, Zhou X (2006) *Soil Respiration and the Environment*, New York, Academic Press.

Marshall L, Nott D, Sharma A (2004) A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research,* **40**, W02501.

Matthews E (1997) Global litter production, pools, and turnover times: Estimates from measurement data and regression models. *Journal of Geophysical Research: Atmospheres,* **102**, 18771-18800.

Melillo JM, Aber JD, Muratore JF (1982) Nitrogen and Lignin Control of Hardwood Leaf Litter Decomposition Dynamics. *Ecology*, **63**, 626.

Melillo JM, Steudler PA, Aber JD *et al.* (2002) Soil Warming and Carbon-Cycle Feedbacks to the Climate System. *Science,* **298**, 2173-2176.

Mosegaard K, Sambridge M (2002) Monte Carlo analysis of inverse problems. *Inverse Problems,* **18**, R29-R54.

Müller T, Höper H (2004) Soil organic matter turnover as a function of the soil clay content: consequences for model applications. *Soil Biology and Biochemistry,* **36**, 877-888.

Norby RJ, Cotrufo MF, Ineson P, O'neill EG, Canadell JG (2001) Elevated CO 2, litter chemistry, and decomposition: a synthesis. *Oecologia,* **127**, 153-165.

O'neill EG, Norby RJ (1996) 6 - Litter Quality and Decomposition Rates of Foliar Litter Produced under CO2 Enrichment. In: *Carbon dioxide and terrestrial ecosystems.* (eds Koch GW, Mooney HA) pp 87-103. San Diego, Academic Press.

Oechel WC, Vourlitis GL, Hastings SJ, Zulueta RC, Hinzman L, Kane D (2000) Acclimation of ecosystem CO2 exchange in the Alaskan Arctic in response to decadal climate warming. *Nature,* **406**, 978-981.

Oleson K, Niu G, Yang Z *et al.* (2007) CLM3. 5 documentation. *National Center for Atmospheric Research, Boulder, 34pp*.

Oleson KW, Dai Y, Bonan G *et al.* (2004) Technical description of the community land model (CLM). *NCAR Tech. Note NCAR/TN-461+ STR*, 173.

Oleson KW, Niu G, Yang Z *et al.* (2008) Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res,* **113**, G01021.

Olson JS (1963) ENERGY-STORAGE AND BALANCE OF PRODUCERS AND DECOMPOSERS IN ECOLOGICAL-SYSTEMS. *Ecology,* **44**, 322-&.

Olson R, Scurlock J, Prince S, Zheng D, Johnson K (2001) NPP multi-biome: NPP and driver data for ecosystem model-data intercomparison. *Data set. Available on-line [http://www. daac. ornl. gov] from the Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA*.

Pan Y, Birdsey RA, Fang J *et al.* (2011) A Large and Persistent Carbon Sink in the World's Forests. *Science*, **333**, 988-993.

Parton WJ, Hanson PJ, Swanston C, Torn M, Trumbore SE, Riley W, Kelly R (2010) ForCent model development and testing using the Enriched Background Isotope Study experiment. *Journal of Geophysical Research: Biogeosciences*, **115**, G04001.

Parton WJ, Schimel DS, Cole CV, Ojima DS (1987) Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal*, **51**, 1173-1179.

Parton WJ, Scurlock JMO, Ojima DS *et al.* (1993) Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochemical Cycles*, **7**, 785-809.

Paul S, Flessa H, Veldkamp E, López-Ulloa M (2008) Stabilization of recent soil carbon in the humid tropics following land use changes: evidence from aggregate fractionation and stable isotope analyses. *Biogeochemistry*, **87**, 247-263.

Peng S, Piao S, Wang T, Sun J, Shen Z (2009) Temperature sensitivity of soil respiration in different ecosystems in China. *Soil Biology and Biochemistry*, **41**, 1008-1014.

Peters GP, Marland G, Le Quere C, Boden T, Canadell JG, Raupach MR (2012) Rapid growth in CO2 emissions after the 2008-2009 global financial crisis. *Nature Clim. Change,* **2**, 2-4.

Post J, Hattermann FF, Krysanova V, Suckow F (2008) Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environmental Modelling & Software,* **23**, 125-138.

Post WM, Emanuel WR, Zinke PJ, Stangenberger AG (1982) Soil carbon pools and world life zones. *Nature,* **298**, 156-159.

Potter CS, Randerson JT, Field CB, Matson PA, Vitousek PM, Mooney HA, Klooster SA (1993) Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles,* **7**, 811-841.

Qian T, Dai A, Trenberth KE, Oleson KW (2006) Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations. *Journal of Hydrometeorology,* **7**, 953-975.

Raddatz TJ, Reick CH, Knorr W *et al.* (2007) Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate Dynamics,* **29**, 565-574.

Raich JW, Schlesinger WH (1992) The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B,* **44**, 81-99.

Randerson JT, Hoffman FM, Thornton PE *et al.* (2009a) Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models. *Global Change Biology*, doi: 10.1111/j.1365-2486.2009.01912.x.

Randerson JT, Hoffman FM, Thornton PE *et al.* (2009b) Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models. *Global Change Biology,* **15**, 2462-2484.

Rasse D, Rumpel C, Dignac M-F (2005) Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant and Soil,* **269**, 341-356.

Raupach MR, Marland G, Ciais P, Le Quéré C, Canadell JG, Klepper G, Field CB (2007) Global and regional drivers of accelerating CO2 emissions. *Proceedings of the National Academy of Sciences,* **104**, 10288-10293.

Rayner PJ, Scholze M, Knorr W, Kaminski T, Giering R, Widmann H (2005) Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochem. Cycles,* **19**, GB2026.

Roeckner E, Bäuml G, Bonventura L *et al.* (2003) The atmospheric general circulation model ECHAM5. PART I: Model description, Report 349, Max Planck Institute for Meteorology, Hamburg, Germany.

Rumpel C, Kögel-Knabner I (2011) Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. *Plant and Soil,* **338**, 143-158.

Rustad LR, Campbell J, Marion G *et al.* (2001) A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming. *Oecologia,* **126**, 543-562.

Schimel JP, Weintraub MN (2003) The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. *Soil Biology and Biochemistry,* **35**, 549-563.

Shaw MR, Harte J (2001) Control of litter decomposition in a subalpine meadow-sagebrush steppe ecotone under climate change. *Ecological Applications,* **11**, 1206-1223.

Sinsabaugh RL, Manzoni S, Moorhead DL, Richter A (2013) Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling. *Ecology Letters,* **16**, 930-939.

Smith MJ, Purves DW, Vanderwel MC, Lyutsarev V, Emmott S (2013) The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties. *Biogeosciences,* **10**, 583-606.

Smyth C, Trofymow J, Kurz W, Group CW (2009) Decreasing uncertainty in CBM-CFS3 estimates of forest soil C sources and sinks through use of long-term data from the Canadian Intersite Decomposition Experiment.

Steinweg JM, Plante AF, Conant RT, Paul EA, Tanaka DL (2008) Patterns of substrate utilization during long-term incubations at different temperatures. *Soil Biology and Biochemistry,* **40**, 2722-2728.

Stocker TF, Qin D, Plattner G-K *et al.* (2013) IPCC, 2013: Summary for Policy makers. Cambridge, United Kingdom and New York, NY, USA.

Stump LM, Binkley D (1993) Relationships between litter quality and nitrogen availability in Rocky Mountain forests. *Canadian Journal of Forest Research,* **23**, 492-502.

Svirezhev YM (2002) Simple spatially distributed model of the global carbon cycle and its dynamic properties. *Ecological Modelling,* **155**, 53-69.

Taylor BR, Parkinson D, Parsons WFJ (1989) Nitrogen and Lignin Content as Predictors of Litter Decay Rates: A Microcosm Test. *Ecology,* **70**, 97-104.

Taylor KE, Stouffer RJ, Meehl GA (2011) An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society,* **93**, 485-498.

Thum T, Räisänen P, Sevanto S *et al.* (2011) Soil carbon model alternatives for ECHAM5/JSBACH climate model: Evaluation and impacts on global carbon cycle estimates. *J. Geophys. Res.,* **116**, G02028.

Todd-Brown KEO, Randerson JT, Hopkins F *et al.* (2013a) Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences Discuss.,* **10**, 18969-19004.

Todd-Brown KEO, Randerson JT, Post WM, Hoffman FM, Tarnocai C, Schuur EaG, Allison SD (2012) Causes of variation in soil carbon predictions from CMIP5 Earth system models and comparison with observations. *Biogeosciences Discuss.,* **9**, 14437-14473.

Todd-Brown KEO, Randerson JT, Post WM, Hoffman FM, Tarnocai C, Schuur EaG, Allison SD (2013b) Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences,* **10**, 1717-1736.

Torn MS, Trumbore SE, Chadwick OA, Vitousek PM, Hendricks DM (1997) Mineral control of soil organic carbon storage and turnover. *Nature,* **389**, 170-173.

Vance ED, Chapin Iii FS (2001) Substrate limitations to microbial activity in taiga forest floors. *Soil Biology and Biochemistry,* **33**, 173-188.

Vogt KA, Grier CC, Vogt DJ (1986) Production, Turnover, and Nutrient Dynamics of Above- and Belowground Detritus of World Forests. In: *Advances in Ecological Research.* (eds Macfadyen A, Ford ED) pp 303-377. Academic Press.

Wall DH, Bradford MA, St. John MG *et al.* (2008) Global decomposition experiment shows soil animal impacts on decomposition are climate-dependent. *Global Change Biology*, **14**, 2661-2677.

Wang C, Han G, Jia Y, Feng X, Tian X (2012a) Insight into the temperature sensitivity of forest litter decomposition and soil enzymes in subtropical forest in China. *Journal of Plant Ecology*, **5**, 279-286.

Wang G, Post WM, Mayes MA (2012b) Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. *Ecological Applications*, **23**, 255-272.

Wang WJ, Dalal RC, Moody PW, Smith CJ (2003) Relationships of soil respiration to microbial biomass, substrate availability and clay content. *Soil Biology and Biochemistry*, **35**, 273-284.

Wang YP, Baldocchi D, Leuning RaY, Falge EVA, Vesala T (2007) Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. *Global Change Biology*, **13**, 652-670.

Wang YP, Chen BC, Wieder WR *et al.* (2013) Oscillatory behavior of two nonlinear microbial models of soil carbon decomposition. *Biogeosciences Discuss.*, **10**, 19661-19700.

Waring B, Weintraub S, Sinsabaugh R (2014) Ecoenzymatic stoichiometry of microbial nutrient acquisition in tropical soils. *Biogeochemistry*, **117**, 101-113.

Weng E, Luo Y (2011) Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. *Ecological Applications*, **21**, 1490-1505.

Wetterstedt JÅM, Ågren GI (2011) Quality or decomposer efficiency – which is most important in the temperature response of litter decomposition? A modelling study using the GLUE methodology. *Biogeosciences*, **8**, 477-487.

Wieder WR, Bonan GB, Allison SD (2013) Global soil carbon projections are improved by modelling microbial processes. *Nature Clim. Change*, **3**, 909-912.

Wieder WR, Cleveland CC, Townsend AR (2009) Controls over leaf litter decomposition in wet tropical forests. *Ecology*, **90**, 3333-3341.

Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, **11**, 89-105.

Wu X, Luo Y, Weng E, White L, Ma Y, Zhou X (2009) Conditional inversion to estimate parameters from eddy-flux observations. *Journal of Plant Ecology*, **2**, 55-68.

Xia J, Luo Y, Wang Y, Weng E, Hararuk O (2012) A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state. *Geosci. Model Dev.*, **5**, 1259-1271.

Xia J, Luo Y, Wang YP, Hararuk O (2013a) Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Global Change Biology*, n/a-n/a.

Xia J, Luo Y, Wang YP, Hararuk O (2013b) Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Global Change Biology, **In press***.

Xu T, White L, Hui D, Luo Y (2006) Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem. Cycles*, **20**, GB2007.

Xu X, Thornton PE, Post WM (2013) A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Global Ecology and Biogeography*, **22**, 737-749.

Zhang D, Hui D, Luo Y, Zhou G (2008) Rates of litter decomposition in terrestrial ecosystems: global patterns and controlling factors. *Journal of Plant Ecology,* **1**, 85-93.

Zhao M, Running SW (2010) Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009. *Science,* **329**, 940-943.

Zhou G, Guan L, Wei X *et al.* (2008) Factors influencing leaf litter decomposition: an intersite decomposition experiment across China. *Plant and Soil,* **311**, 61-72.

Zhou T, Luo Y (2008) Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States. *Global Biogeochemical Cycles*, **22**, n/a-n/a.

Zhou T, Shi P, Hui D, Luo Y (2009) Global pattern of temperature sensitivity of soil heterotrophic respiration (Q10) and its implications for carbon-climate feedback. *Journal of Geophysical Research: Biogeosciences*, **114**, n/a-n/a.

Ziehn T, Knorr W, Scholze M (2011) Investigating spatial differentiation of model parameters in a carbon cycle data assimilation system. *Global Biogeochem. Cycles*, **25**, GB2021.