Identifying Data Needed to Reduce Parameter Uncertainty in a Coupled Microbial Soil C and N Decomposition Model

Mustafa Saifuddin, Rose Z Abramoff, Eric A. Davidson, Michael C. Dietze, Adrien C. Finzi

HAL Id: insu-03721905

https://insu.hal.science/insu-03721905

Submitted on 13 Jul 2022

JGR Biogeosciences

Key Points:
- Model outputs are highly sensitive to certain enzyme rate parameters (e.g., activation energies), and less sensitive to other parameters
- Respiration data helped constrain enzyme activation energies and carbon use efficiency; dissolved organic CN data would be a high priority
- This approach can reduce parameter uncertainty in biogeochemical models by tailoring data collection to reduce model uncertainty

Correspondence to:
R. Z. Abramoff and M. Saifuddin,
rose.abramoff@lsce.ipsl.fr;
msaif@bu.edu

Identifying Data Needed to Reduce Parameter Uncertainty in a Coupled Microbial Soil C and N Decomposition Model

Mustafa Saifuddin[1], Rose Z. Abramoff[2] , Eric A. Davidson[3] , Michael C. Dietze[1], and Adrien C. Finzi[1]

[1]Boston University, Boston, MA, USA, [2]Laboratoire des Sciences du Climat et de l'Environnement, Gif sur Yvette, France, [3]University of Maryland Center for Environmental Science, Cambridge, MD, USA

Abstract Advancements in microbially explicit ecosystem models incorporate increasingly accurate representations of microbial physiology and enzyme-mediated depolymerization of soil organic matter in predicting biogeochemical responses to global change. However, a major challenge with model structural improvements is the requirement for additional parameters, which are often poorly constrained sources of uncertainty. Furthermore, it is often unclear how to best focus data collection efforts toward reducing model uncertainty. Here, we use Dual Arrhenius Michaelis-Menten Microbial Carbon and Nitrogen Physiology, a microbially mediated, coupled soil C and N cycling model, as a tool to explore the influence of microbial physiological and enzyme kinetic parameters on model estimates. We first quantify the potential for constraining model parameters using empirical measurements of soil respiration. We then use simulated data to identify which additional sources of data collection from the field would provide the greatest impact for constraining model estimates. We find that modeled soil C and N pools and fluxes are disproportionately sensitive to only a few parameters (e.g., activation energies and microbial *CUE*), while others exert less influence (e.g., Michaelis-Menten half-saturation constants). While some parameters can be constrained by the available data on heterotrophic respiration, the collection of additional data on dissolved organic C and N pools in the soil is identified as a high-priority data need. Improving our ability to model the interactions of soil microbial physiology, soil chemistry, enzyme activities, and environmental factors on C and N cycling will require closely considering model uncertainties and prioritizing future data collection opportunities based on their impact on model performance.

Plain Language Summary Soil biogeochemical models have increasingly realistic representations of microbial processes; however, model outputs are associated with multiple types of uncertainty. We quantify the uncertainty associated with specific parameters in Dual Arrhenius Michaelis-Menten Microbial Carbon and Nitrogen Physiology, a microbial coupled carbon and nitrogen cycle model. Reducing uncertainty in model outputs can be achieved through collecting additional data. We use existing and simulated data to identify which types of data would be most useful in reducing parameter uncertainty for this model, which may serve as an example for other microbial soil biogeochemical models.

1. Introduction

The cycling of carbon and nitrogen are closely coupled in terrestrial ecosystems. N is a required constituent of nucleic acids, proteins, enzymes and biomass, and reduced C provides the energy needed to fuel metabolic reactions. Soil microorganisms play a central role in linking these cycles through the decomposition of soil organic matter, followed by assimilation of C and N into microbial biomass, partitioning of these elements among metabolic pathways, and through the production of extracellular enzymes. Despite the activities of ubiquitous microbial decomposers, soils sequester vast quantities of organic matter (Friedlingstein et al., 2020; Jobbágy & Jackson, 2000; Stockmann et al., 2013). Identifying the complex drivers of changes in these stocks, including soil microbial physiology, soil chemistry, enzyme activities, and environmental factors, has remained a primary challenge in the field (Lehmann et al., 2020; Schmidt et al., 2011; Waring et al., 2020).

Soil biogeochemical models provide a useful framework for testing hypotheses on C and N cycling and predicting ecosystem responses to global change (Blankinship et al., 2018; Manzoni & Porporato, 2009). These types of models have increased in complexity over the past decades with growing information on linkages between soil C and N cycling, microbial physiology, enzyme kinetics, and their responses to temperature and moisture (e.g.,

Supervision: Rose Z. Abramoff, Eric A. Davidson, Michael C. Dietze, Adrien C. Finzi
Validation: Mustafa Saifuddin, Rose Z. Abramoff, Adrien C. Finzi
Visualization: Mustafa Saifuddin
Writing – original draft: Mustafa Saifuddin
Writing – review & editing: Mustafa Saifuddin, Rose Z. Abramoff, Eric A. Davidson, Michael C. Dietze, Adrien C. Finzi

Manzoni & Porporato, 2009; Zhang et al., 2020, 2021). Advancements in the representation of N cycling in these models to account for the potential for N limitation has been shown to improve their capacity to predict C sequestration, which is often over-estimated in C-only models (Hungate, 2003; Thornton et al., 2007; Zaehle, 2013). An additional development in modeling C and N cycling has been the incorporation of more explicit representations of soil microbial physiology and the role of microbial groups in producing extracellular enzymes (e.g., Bouskill et al., 2012; Fatichi et al., 2019; Huang et al., 2018; Kyker-Snowden et al., 2020; Schimel & Weintraub, 2003; Sistla et al., 2014; Tang & Riley, 2015; Waring et al., 2013). These advancements have allowed models to elucidate drivers of biogeochemical responses to environmental change (Allison et al., 2010; Sistla et al., 2014; Wieder et al., 2014) and improve their capacity to predict the spatial distribution of soil C (Wieder et al., 2013).

With growing complexity, biogeochemical models are sometimes able to more realistically represent microbial C and N cycling; however, the addition of new model compartments and parameters also has the potential for increasing uncertainty (Manzoni & Porporato, 2009; Marschmann et al., 2019; Shi et al., 2018). Several recent model comparison studies demonstrate divergent model predictions from closely related soil C cycle models. For example, even when given identical forcing data, models disagree on their response to temperature and plant inputs, due to a combination of differences in model forms and parameterization (Wieder et al., 2017). While such divergences are often demonstrated across model intercomparison efforts, it is often not immediately apparent how to best tailor data collection efforts toward reducing model uncertainty, highlighting the need for closer integration between model development and data collection (Sulman et al., 2018; Weintraub et al., 2019; Xie et al., 2020; Zhou et al., 2021).

Soil biogeochemical models are subject to a number of sources of uncertainty, including model structural uncertainty (Ajami & Gu, 2010; Myrgiotis et al., 2018). For example, microbial processes were previously often represented by a single microbial biomass pool, while more recent models have shifted to include multiple functional guilds or distinguish between different microbial functional types (Moorhead & Sinsabaugh, 2006; Sistla et al., 2014; Waring et al., 2013; Wieder et al., 2014). Similarly, the number and types of specific C and N pools represented across models also varies, with a shift toward representing measurable, rather than conceptual pools (Abramoff et al., 2021; Zhang et al., 2021). Identifying soil biogeochemical model structures that are not only accurate representations of soil microbial functions but also useful and parsimonious tools for making projections and predictions remains an active area of research, with many contrasting approaches to model structure (Fan et al., 2021; Kyker-Snowman et al., 2020; Sainte-Marie et al., 2021; Tang & Riley, 2020; Waring et al., 2020; Zhang et al., 2021) and several model comparison efforts exploring the implications of differences in structure (Georgiou et al., 2021; Sulman et al., 2018; Zhou et al., 2021).

A second source of uncertainty, which often receives less attention than model structural uncertainty, is related to parameters used in models (Luo & Schuur, 2019). Most soil biogeochemical models use fixed parameter values selected *ad hoc* or derived from site-specific literature measurements. While some model parameters correspond to values that can be directly measured with relative ease (e.g., stoichiometries of soil organic matter and plant litter) or which reflect physical constants (e.g., diffusion coefficients), parameters that describe enzyme kinetics and microbial physiology can be more challenging to directly quantify. Specific enzyme kinetics are often measured in lab assays, but their translation into models has been challenging due to a mismatch between the conceptual or simplified substrates represented in models and the diversity of highly specific extracellular enzymes, each with unique stoichiometries and sensitivities (Drake et al., 2013; Sinsabaugh et al., 2014, 2015). Microbial physiological parameters, such as carbon use efficiency or allocations toward extracellular enzyme production, are also particularly hard to quantify due to metabolic diversity across taxa and challenges with empirical measurements (Ballantyne & Billings, 2018; Dijkstra et al., 2015; Geyer et al., 2016; Hagerty et al., 2018; Saifuddin et al., 2019). Despite large uncertainties in these parameter choices, soil C and N cycle models rarely account for parameter uncertainty explicitly.

In addition to quantifying these distinct sources of uncertainty, identifying which types of data have the potential to reduce uncertainty in soil biogeochemical models is necessary to ensure that data collection efforts are synchronized with efforts to improve model performance. In complex soil biogeochemical models with several pools and dozens of parameters, it may not be readily evident which data sets will most directly reduce model uncertainty. The use of a Bayesian statistical framework has been demonstrated to allow for improved comparisons between soil biogeochemical models (Ajami & Gu, 2010; Xie et al., 2020), and the use of simulated data with a Bayesian approach has been used to identify pool sizes in a soil carbon model (Scharnagl et al., 2010). Here, we
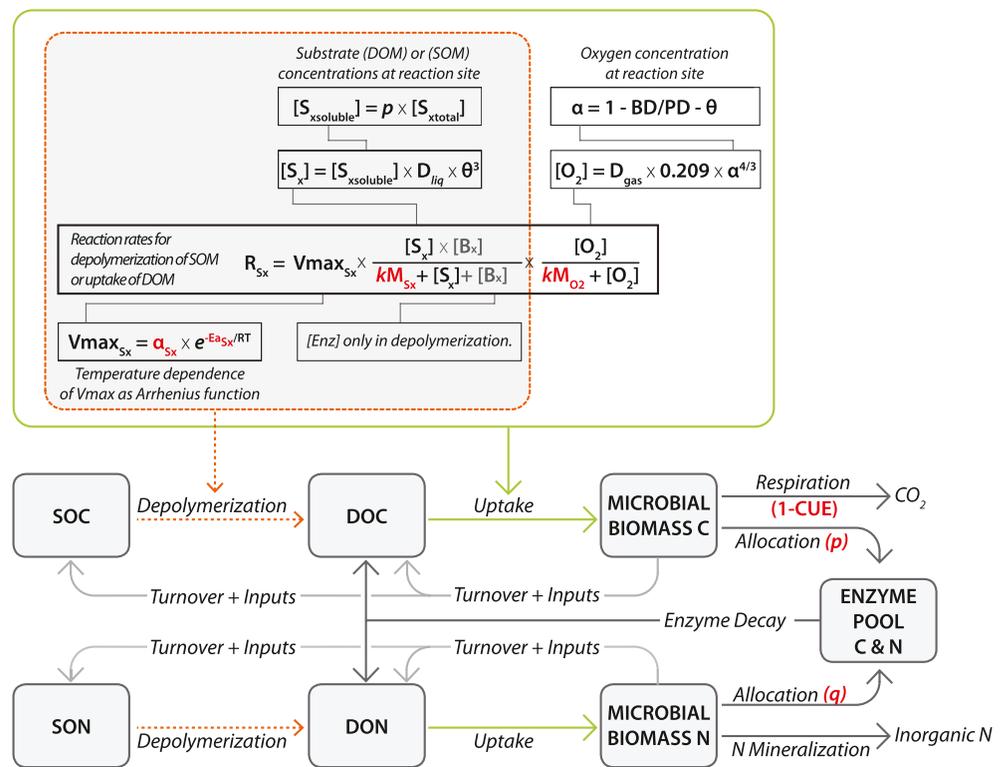
Figure 1. Major pools, fluxes, and equations in DAMM-MCNiP. Note. Dashed orange lines enclose equations used for depolymerization kinetics. Solid green lines enclose equations for uptake of DOCN (with the exception of [Enz], which only appears in depolymerization kinetics). Microbial parameters are shown above associated arrows. Parameters estimated in the present analysis are highlighted in red.

develop one approach to quantifying model parameter uncertainty using a combination of real and simulated data to inform future data collection efforts. Specifically, we use a microbially explicit model of soil C and N cycling, Dual Arrhenius Michaelis-Menten Microbial Carbon and Nitrogen Physiology (DAMM-MCNiP), to explore how parameters associated with enzyme kinetics, microbial uptake of C and N, and microbial physiology impact estimates of C and N cycling (Abramoff et al., 2017). Using a combination of sensitivity analyses and a Bayesian statistical approach, we explore the following questions:

Inherent Model Sensitivity: Which model parameters have the greatest impact on estimates of soil C and N cycling in this model?

Bayesian Data Assimilation: Can the assimilation of existing data reduce parameter uncertainty in this model, and what is the impact of reducing parameter uncertainty on model estimates?

Simulated Data to Identify High-Priority Field Studies: Which specific types of data collection efforts have the potential to reduce parameter uncertainty in this model most effectively?

2. Methods

2.1. DAMM-MCNiP Model Description

The DAMM-MCNiP model represents coupled soil C and N cycling through soil organic stocks (SOCN), dissolved organic stocks (DOCN), microbial biomass, and extracellular enzyme pools based on soil temperature and soil moisture (Figure 1, Abramoff et al., 2017; Davidson et al., 2011; Finzi et al., 2015). The model utilizes a combination of Arrhenius and Michaelis-Menten kinetic equations to describe the depolymerization of SOCN, resulting in the production of DOCN. A similar set of Arrhenius and Michaelis-Menten kinetic equations describe the incorporation of DOCN into microbial biomass. A series of microbial physiological parameters determine how C and N are partitioned between microbial biomass, enzyme production, respiration, and N mineralization (Figure 1). The model captures seasonal patterns of heterotrophic respiration at trenched (root-free) plots in a

temperate hardwood forest (Abramoff et al., 2017). This model structure is similar to those of several microbially explicit biogeochemical models across ecosystem types, making inferences regarding parameter uncertainty and model structure potentially generalizable (Abramoff et al., 2018; Manzoni & Porporato, 2009; Schimel & Weintraub, 2003; Sistla et al., 2014; Waring et al., 2013).

In DAMM-MCNiP, extracellular enzymes produced by microbes depolymerize a fraction of the SOCN pool according to equilibrium chemistry approximation kinetics (Tang, 2015), which are an approximation of reaction kinetics, similar to Michaelis-Menten:

$$\text{Depolymerization rate} = Vmax_S \times \frac{[SOMavail] \times [\text{ENZ}]}{kM_{dep} + [SOMavail] + [\text{ENZ}]} \tag{1}$$

$SOMavail$, the fraction of the total SOCN pool available for depolymerization, is dependent on the cube of soil moisture (θ), which represents diffusional constraints of DOC substrates and extracellular enzymes in water films ($dLiq$), and the fraction of SOCN that is not physically or chemically occluded from enzyme-binding ($Frac$; Magill et al., 2000):

$$SOM_{avail} = SOM \times Frac \times dLiq \times \theta^3 \tag{2}$$

The maximum rate of depolymerization is determined by soil temperature according to the Arrhenius function:

$$Vmax_S = \alpha_{dep} \times e^{-\frac{Ea_{dep}}{RT}} \tag{3}$$

Thus, a total of three kinetic parameters (Km_{dep}, a_{dep}, and Ea_{dep}) are involved in specifying depolymerization rates based on temperature and the pool sizes of SOCN and enzymes.

Following depolymerization, SOCN enters the DOCN pool, which is available for uptake by microbial biomass (Figure 1). Uptake rate is determined by Michaelis-Menten kinetics, which are dependent on the concentration of DOCN and the concentration of O_2 based on soil moisture:

$$\text{Uptake rate} = \text{MICN} \times Vmax_D \times \frac{[\text{DOCN}]}{kM_{upt} + [\text{DOCN}]} \times \frac{[O_2]}{kM_{O2} + [O_2]} \tag{4}$$

The maximum rate of uptake is specified by the Arrhenius function:

$$Vmax_D = \alpha_{upt} \times e^{-\frac{Ea_{upt}}{RT}} \tag{5}$$

Thus, four parameters (Km_{upt}, Km_{o2}, a_{upt}, and Ea_{upt}) are involved in specifying rates of DOCN uptake based on DOCN pool sizes and soil O_2 concentration, which is dependent on soil moisture (Abramoff et al., 2017).

Following uptake, C and N are either retained within microbial biomass, allocated toward enzyme production, or lost through respiration and N mineralization. The model ranks the order of operation for these activities according to:

$$\text{respiration} > \text{enzyme production} > \text{microbial biomass production} > \text{overflow C} = \text{N min.}$$

Respiration is first calculated as a fixed fraction of C uptake ($1 - CUE$). Two separate parameters, p and q, determine maximum allocation of C and N toward enzyme production respectively, with actual allocations constrained by enzyme stoichiometry according to Liebig's law of the minimum (Liebig, 1842). If there remain excess C or N from this initial allocation that cannot be incorporated into enzymes due to stoichiometric demand, remaining C and N are maximally incorporated into microbial biomass. Finally, if there remain excess C or N that cannot be incorporated into biomass due to stoichiometric demand, it is lost as overflow respiration or N mineralization. While these assumptions regarding microbial physiology reflect a simplified representation of microbial resource allocation, they are in line with other microbially explicit biogeochemical models. Note that this model, parameterized with data from trenched (root-free) plots also lacks explicit representation of root and mycorrhizal uptake of N, which may lead to an overestimation of N availability for microbes relative to stands with intact roots. A more detailed description of the DAMM-MCNiP model and model code are available in Abramoff et al. (2017).

Table 1
DAMM-MCNiP Parameters, Units, Default Values and Definitions

Parameter	Units	Default value	Description
Ea_{dep}	kJ mol^{-1}	61.77	Activation energy for SOCN depolymerization
Km_{dep}	mg cm^{-3}	0.0025	Half-saturation constant for SOCN depolymerization
a_{dep}	mg SOCN cm^{-3} (mg Enz cm^{-3})$^{-1}$ hr^{-1}	1.0815×10^{11}	Pre-exponential constant for SOCN depolymerization
Ea_{upt}	kJ mol^{-1}	61.77	Activation energy for DOCN uptake
a_{upt}	mg DOCN cm^{-3} (mg biomass cm^{-3})$^{-1}$ hr^{-1}	1.0815×10^{11}	Pre-exponential constant for DOCN uptake
Km_{upt}	mg cm^{-3}	0.3	Half-saturation constant for DOCN uptake
Km_{O2}	cm^3 O$_2$ cm^{-3} air	0.121	Michaelis constant for O$_2$
CUE	mg mg^{-1}	0.31	Carbon use efficiency
p	-	0.5	Proportion of assimilated C allocated to enzyme production
q	-	0.5	Proportion of assimilated N allocated to enzyme production
CN_s	-	27.6	C:N of soil
CN_l	-	27.6	C:N of litter
CN_m	-	10	C:N of microbial biomass
CN_e	-	3	C:N of enzymes
BD	g cm^{-3}	0.8	Bulk density
PD	g cm^{-3}	2.52	Particle density
r_death	hr^{-1}	0.00015	Microbial turnover rate
r_EnzLoss	hr^{-1}	0.001	Enzyme turnover rate
MICtoSOCN	mg mg^{-1}	0.5	Fraction of dead microbial biomass allocated to SOCN
a	-	0.5	Proportion of enzyme pool acting on SOC pool (1 − a = proportion acting on SON pool)
Frac	g C cm^{-3}/g C cm^{-3}	0.000414	Fraction of unprotected SOCN, using soluble substrate estimated from Magill et al. (2000)
Sat	cm^3 H$_2$O cm^{-3} soil	1	Moisture saturation level
O$_{2airfrac}$	L O$_2$/L air	0.209	Volume fraction of O$_2$ air
D_{liq}	-	3.17	Diffusion coefficient for unprotected SOCN and DOCN in liquid
D_{gas}	-	1.67	Diffusion coefficient for O$_2$ in air

Note. First ten rows show parameters estimated in the present analysis.

2.2. Estimation of Parameter Sensitivity

DAMM-MCNiP requires a total of 25 parameters (Table 1). A subset of seven parameters are directly involved in the kinetic equations describing depolymerization and uptake, and three parameters are used to represent microbial physiology. The remaining 15 parameters are primarily associated with defining stoichiometries and soil physical properties. We assessed the sensitivity of model outputs to the 10 depolymerization, uptake, and physiological parameters. Each parameter was set to vary from 50% below to 50% above the default parameter value set in previous model publications (Table 1) in 10% increments (*i*) for a total of 11 parameter settings (*y*). The associated values for respiration, SOC, SON, DOC, DON, microbial biomass C, and microbial biomass N were calculated at each parameter setting (*x*). The sensitivity, or change in model outputs relative to changes in parameter values, was calculated over each pair of parameter settings (e.g., Pianosi et al., 2016):

$$\text{Sensitivity}_{i,x,y} = \frac{\text{Output}_{x+i} - \text{Output}_x}{\text{Output}_x} \times \left(\frac{\text{Parameter}_{y+i} - \text{Parameter}_y}{\text{Parameter}_y} \right)^{-1} \qquad (6)$$

This approach scales rates of change by both pool size and parameter size to allow for congruent comparisons across model outputs and parameter types, which can vary by orders of magnitude due to the variety of types of parameters and model outputs.

2.3. Estimation of Parameter Values Based on Data Assimilation

We utilized a Bayesian statistical framework to explore the potential for constraining parameter uncertainty through data assimilation (Dietze et al., 2014; Lu et al., 2017; Luo et al., 2009; Shi et al., 2015). This approach involves providing prior distributions describing the potential range of values for each parameter of interest. The likelihood of the observed data given particular parameter selections from within the prior distribution is then evaluated. This process is repeated manyfold to generate a posterior distribution describing how likely particular parameter values are given prior constraints and data.

As per Bayes' theorem (Box & Tiao, 1992), the posterior distribution $P(\theta|X)$ of the model parameters θ, given observational data X, can be calculated from the prior distribution $P(\theta)$ of the model parameters θ and the likelihood $P(X|\theta)$ as:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \qquad (7)$$

In this analysis, we utilized broad, uniform prior distributions throughout all simulations ($\pm 50\%$ of default parameter values) to explore a wide range of parameter options outside of those currently incorporated in the published model (Abramoff et al., 2017). Differential Evolution Markov Chain Monte Carlo (DEzs-MCMC, Ter Braak & Vrugt, 2008) simulations were performed to evaluate the likelihood across parameter space using the BayesianTools package (Hartig et al., 2018) in R (R core team, 2017). We used a normally distributed Likelihood function and performed simulations using three chains, each with 20,000 to 100,000 iterations as needed to achieve convergence based on Gelman-Ruben potential scale reduction factors (psrf < 1.3). We estimated all 10 parameters of interest simultaneously or estimated parameters of a given category (depolymerization, uptake, or physiology) while holding others fixed.

We first assimilated published, field measurements of heterotrophic respiration from trenched plots at the Little Prospect Hill tract of the Harvard Forest Long-Term Ecological Research Site (42.58°N, 72.188°W) in Petersham, Massachusetts, USA to constrain depolymerization, uptake and physiological parameters in DAMM-MCNiP. These observations were collected from a mixed hardwood forest dominated by *Quercus rubra* and *Acer rubrum*. Soils at these sites are classified as Canton fine sandy loam, Typic Distrochrepts. Mean annual precipitation at the site is 110 cm and mean annual temperature is 8°C. The respiration data used in this analysis were collected using automated soil respiration chambers monitoring CO_2 efflux from the soil at half-hourly increments across the growing season in 2009 (Davidson et al., 2011; Savage et al., 2008). Simultaneous, automated measurements of soil moisture and soil temperature were also recorded (Davidson et al., 2011; Savage et al., 2008). Additional details on the carbon budget of Harvard Forest placing these fluxes in context of carbon stocks in the forest are described in Finzi et al. (2020).

Several of the model pools and fluxes in DAMM-MCNiP represent mechanistic processes upstream of respiration which currently lack comparable, high-resolution observational data (Figure 1). To identify which of these data sets would be most useful for future collection and to explore how additional data constraints might impact parameter estimation, we simulated data for SOCN, DOCN, and microbial biomass pools and respiration rates using the default parameter values set in previous model publications and added normally distributed observation error. Simulated data paralleled the frequency of the observational respiration data, with hourly estimates over the course of a single growing season. We then used this simulated data to estimate parameters and compare resulting posterior parameter distributions with the known parameter values used to simulate data. The model was fit to simulated data under five scenarios ((a) all pool size data available, (b) only SOCN pool size data available, (c) only DOCN pool size data available, (d) only microbial biomass data available, and (e) only respiration data available). We calculated the percent difference between the *maximum a posteriori* (MAP) parameter estimate

and the parameter value used in simulation to measure the accuracy of parameter estimation in each data availability scenario. Scenarios with low percent differences indicate that the data assimilation resulted in reconstruction of the true parameter value with high accuracy, while large percent differences between the MAP and the initial parameter value indicate that the data assimilation did not provide sufficient information for constraining parameter estimation. We measured the precision of parameter estimates by comparing the range of the posterior distribution (2.5%–97.5% interval) to the range of the prior distribution for each parameter:

$$\text{Uncertainty Reduction} = \left(1 - \frac{\text{Posterior}_{97.5} - \text{Posterior}_{2.5}}{\text{Prior}_{max} - \text{Prior}_{min}} \right) \times 100 \tag{8}$$

A reduction in the range of the posterior distribution relative to the prior distribution indicates that assimilation of the provided data has provided some constraint on the range of parameter values. In contrast, no change to the posterior distribution relative to the prior distribution would indicate no reduction in uncertainty from assimilation of the provided data.

2.4. Model Projections With Updated Parameter Estimates

For parameters that could be constrained through assimilation of respiration data, we assessed the potential impacts of new parameter estimates on long-term model projections by comparing model outputs using posterior parameter estimates to model outputs estimated using default parameters. An annual seasonal cycle of temperature, moisture, and litterfall inputs was repeated for 100 years. Annual totals were calculated for respiration rates and annual means were calculated for pool sizes of SOC, SON, DOC, DON, microbial biomass C, and microbial biomass N. Annual estimates at median parameter settings and their uncertainties based on 2.5% and 97.5% posterior parameter intervals were compared to annual estimates at default parameter settings and their uncertainties based on the range of prior distributions.
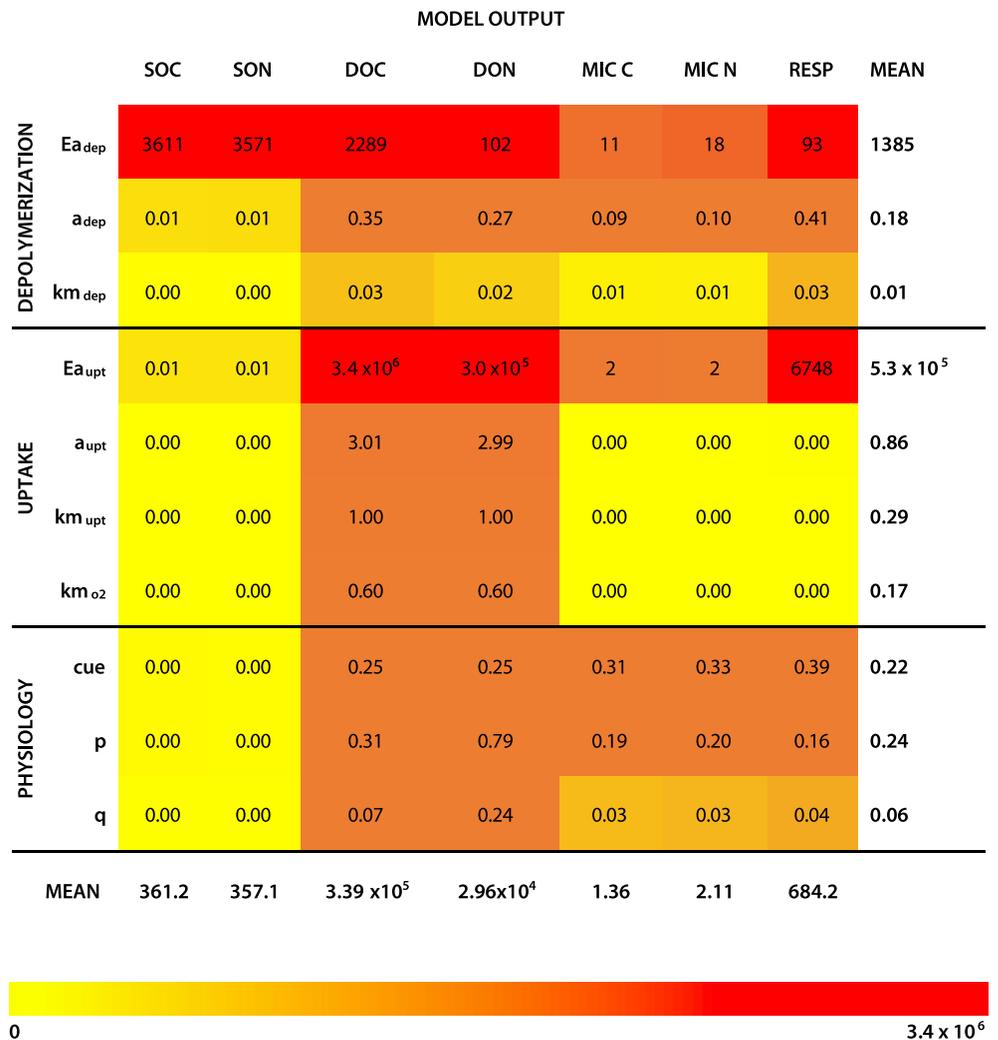
3. Results

3.1. Parameter Sensitivity

Model outputs are disproportionately sensitive to a few select parameters and largely insensitive to others (Figure 2). Calculated sensitivities, which reflect scaled changes in model outputs relative to scaled changes in model parameters from ±50% of their default values (Equation 6), ranged from 0 to 3.4×10^6. All model outputs are most sensitive to either the activation energy of depolymerization (Ea_{dep}) or the activation energy of uptake (Ea_{upt}). In terms of output variables, the DOCN pool sizes and respiration rates show the highest mean sensitivity to variation in individual parameters, while microbial biomass pools have the lowest mean sensitivity. SOCN pool sizes are almost exclusively sensitive to Ea_{dep}, with sensitivities below 0.02 for all other parameters. The low sensitivity of SOCN pools to most kinetic and physiological parameters is amplified by the fact that only a small fraction of total SOCN is available for enzymatic activity due to chemical and physical protection (Magill et al., 2000). DOCN pool sizes show relatively high sensitivities for all parameters except Km_{dep}, which shows the lowest sensitivity across all model outputs. Microbial biomass pools are most sensitive to Ea_{dep}, Ea_{upt}, CUE, and p. Respiration was most sensitive to Ea_{upt}, due to near-zero rates of respiration resulting when Ea_{upt} is raised above approximately 95 kJ mol^{-1}. In contrast, respiration rates are nearly insensitive to the other uptake parameters.

3.2. Results of Assimilating Observed Respiration Data

Assimilating direct observations of seasonal heterotrophic respiration data from Harvard Forest resulted in large reductions in parameter uncertainty for Ea_{dep}, Ea_{upt}, and CUE (Figures 3 and 4a, Table 2). Although respiration is sensitive to a_{dep}, this parameter does not show a reduction in uncertainty when provided with respiration data. The parameters used to define allocations of C and N to enzyme production (p and q) show reductions in uncertainty, but are highly skewed toward the maximum values allowed in the prior specification. Thus, assimilating respiration data alone allowed for large reductions in parameter uncertainty for some of the most sensitive parameters (e.g., Ea_{dep}, Ea_{upt}, and CUE), while most other parameters require additional data constraints as discussed below.

MODEL OUTPUT

		SOC	SON	DOC	DON	MIC C	MIC N	RESP	MEAN
DEPOLYMERIZATION	Ea_{dep}	3611	3571	2289	102	11	18	93	1385
	a_{dep}	0.01	0.01	0.35	0.27	0.09	0.10	0.41	0.18
	km_{dep}	0.00	0.00	0.03	0.02	0.01	0.01	0.03	0.01
UPTAKE	Ea_{upt}	0.01	0.01	3.4×10^6	3.0×10^5	2	2	6748	5.3×10^5
	a_{upt}	0.00	0.00	3.01	2.99	0.00	0.00	0.00	0.86
	km_{upt}	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.29
	km_{o2}	0.00	0.00	0.60	0.60	0.00	0.00	0.00	0.17
PHYSIOLOGY	cue	0.00	0.00	0.25	0.25	0.31	0.33	0.39	0.22
	p	0.00	0.00	0.31	0.79	0.19	0.20	0.16	0.24
	q	0.00	0.00	0.07	0.24	0.03	0.03	0.04	0.06
MEAN		361.2	357.1	3.39×10^5	2.96×10^4	1.36	2.11	684.2	

0 3.4×10^6

Figure 2. Heatmap showing sensitivity of model outputs to parameters. Note. Sensitivities are calculated as proportional change in output (columns) relative to proportional change in parameter (rows; Equation 6). Final row shows mean sensitivity of model output across the 10 parameters of interest. Final column shows mean sensitivity across model outputs for each parameter individually. MIC C is microbial biomass C, MIC N is microbial biomass N, resp is respiration.

3.3. Results of Parameter Estimation With Simulated Data

Using simulated respiration data with known parameter values was useful for corroborating the patterns observed with field measurements of respiration and identifying the potential for constraining parameter estimates with a more complete simulated respiration data set (i.e., one with no missing timepoints in contrast to the observational data set). Estimating parameters following assimilation of simulated respiration data with known, default parameter values indicates that more complete respiration data alone can reliably reduce parameter uncertainty with high accuracy for Ea_{dep}, Ea_{upt}, *CUE,* and *q* (Figure 4b). However, the remaining six parameters showed low accuracy and limited reductions in uncertainty, reinforcing the finding that reducing uncertainty in these parameters would require additional data sources.

Providing the model with simulated data on SOCN pool sizes only reduced uncertainty in Ea_{dep}, while other parameters showed mostly low reductions in uncertainty and inaccurate MAP values (Figure 4c). These observations are consistent with the strong sensitivity of SOCN pool sizes to Ea_{dep} (Figure 2). In contrast, assimilating data on DOCN pool sizes reduced parameter uncertainty for several parameters (Km_{O2}, Ea_{upt}, *q, CUE,* Ea_{dep}) with high accuracy (Figures 4d and 5). Microbial biomass data was primarily useful for reducing uncertainty in Ea_{dep} and the microbial physiological parameters *q* and *CUE* (Figure 4e).
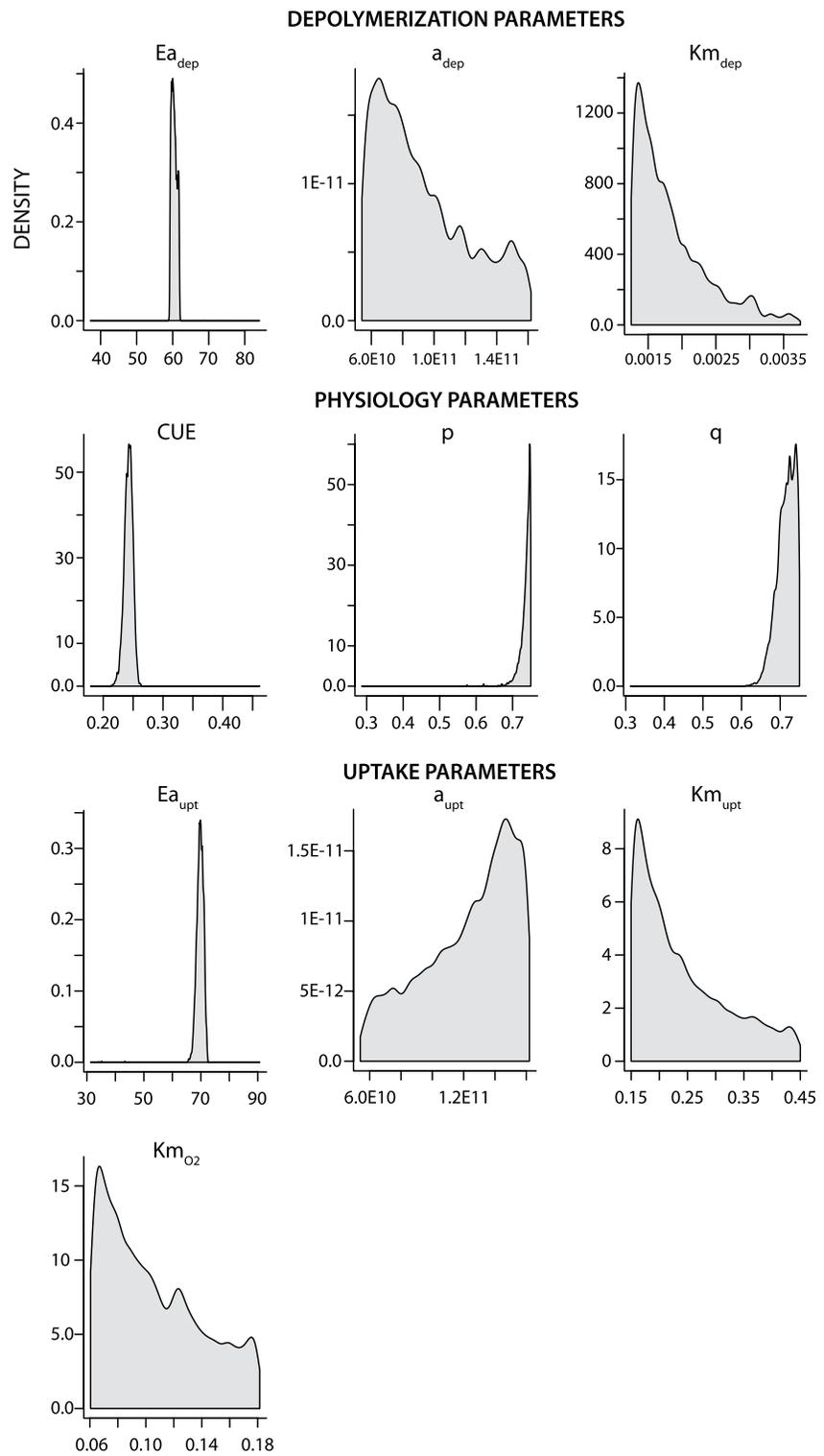
Figure 3. Marginal parameter uncertainties following incorporation of observed respiration data. Note. Prior distributions (not shown) are uniform spanning the full range of *x*-axes.
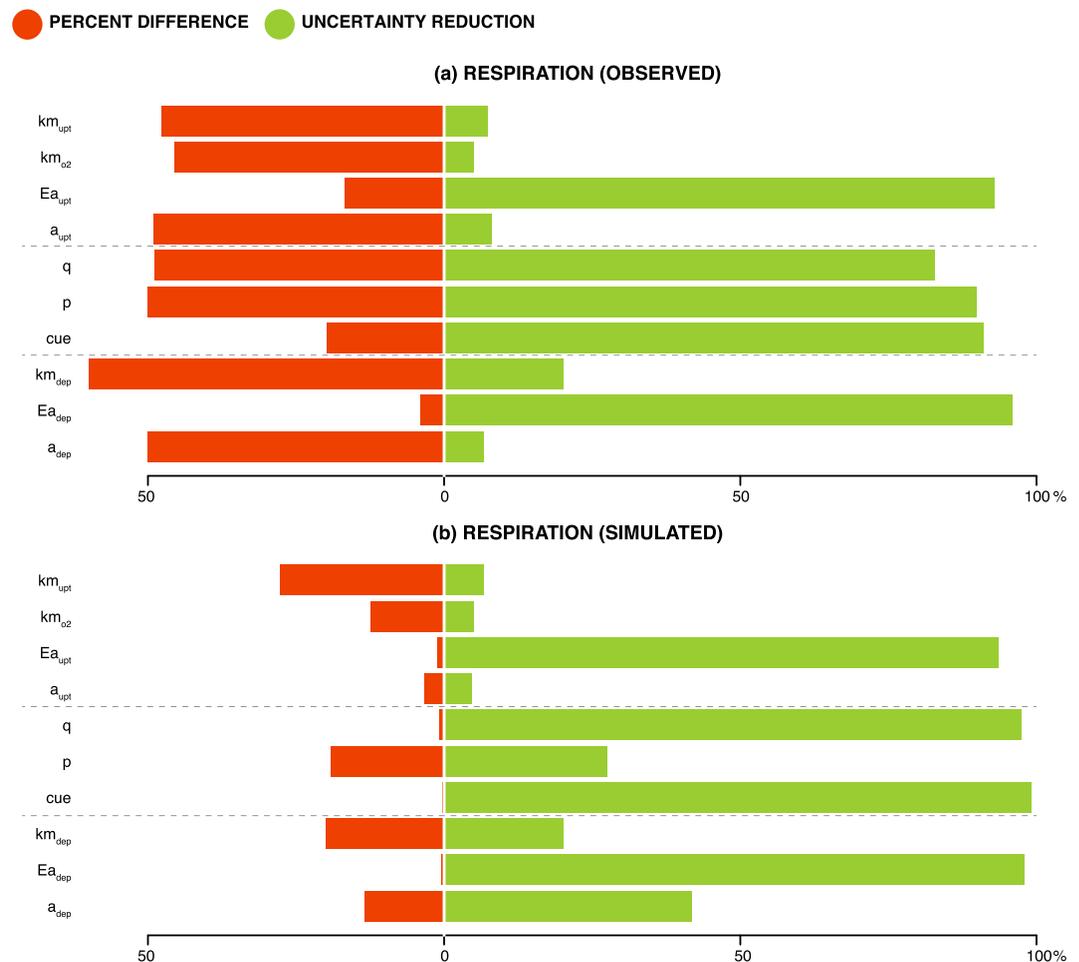
Figure 4. Uncertainty reductions and accuracies for parameters estimated provided different types of data. Note. Percent difference between maximum a posteriori (MAP) estimate and default parameter values (red) and uncertainty reduction (green) for parameters when provided with (a) observed data for respiration, or simulated data for (b) respiration, (c) SOC and SON, (d) DOC and DON, (e) Microbial Biomass C and N. Dashed lines separate between parameter types for uptake, physiology, and depolymerization. Uncertainty reductions approaching 100% indicate narrowing of posterior distribution relative to prior distribution based on data provided. Percent differences approaching 0% indicate MAP values close to the true parameter value used to simulate data.

The most useful data set for constraining all three depolymerization parameters and all three physiological parameters was DOCN data (Figures 4d and 5). Parameters associated with uptake were the most challenging to constrain, with Km_{upt} largely unidentifiable by any provided data. While reductions in uncertainty in uptake parameters were relatively small regardless of data type provided, DOCN showed the greatest overall uncertainty reduction and accuracy.

3.4. Results of Model Projections With Updated Parameter Values

Incorporating parameter estimates for Ea_{dep}, Ea_{upt}, and CUE based on assimilated respiration data results in major changes in estimates of specific pools and fluxes (Figure 6). For example, the posterior estimate for Ea_{dep} was 4% lower than the default parameter value (Table 2). Incorporating this relatively small decrease in parameter value resulted in large declines in SOCN stocks over time (Figure 6a), while other pools and fluxes equilibrate toward similar values. This reduction in Ea_{dep} represents a reduced barrier for SOCN depolymerization, which frees a larger fraction of the initial SOCN pool to be released as DOCN before eventually stabilizing. The increased availability of SOCN is temporarily associated with increases in DOCN and microbial biomass. Updating the parameter value for Ea_{upt} based on the posterior median estimate (a 17% increase in parameter value; Table 2)
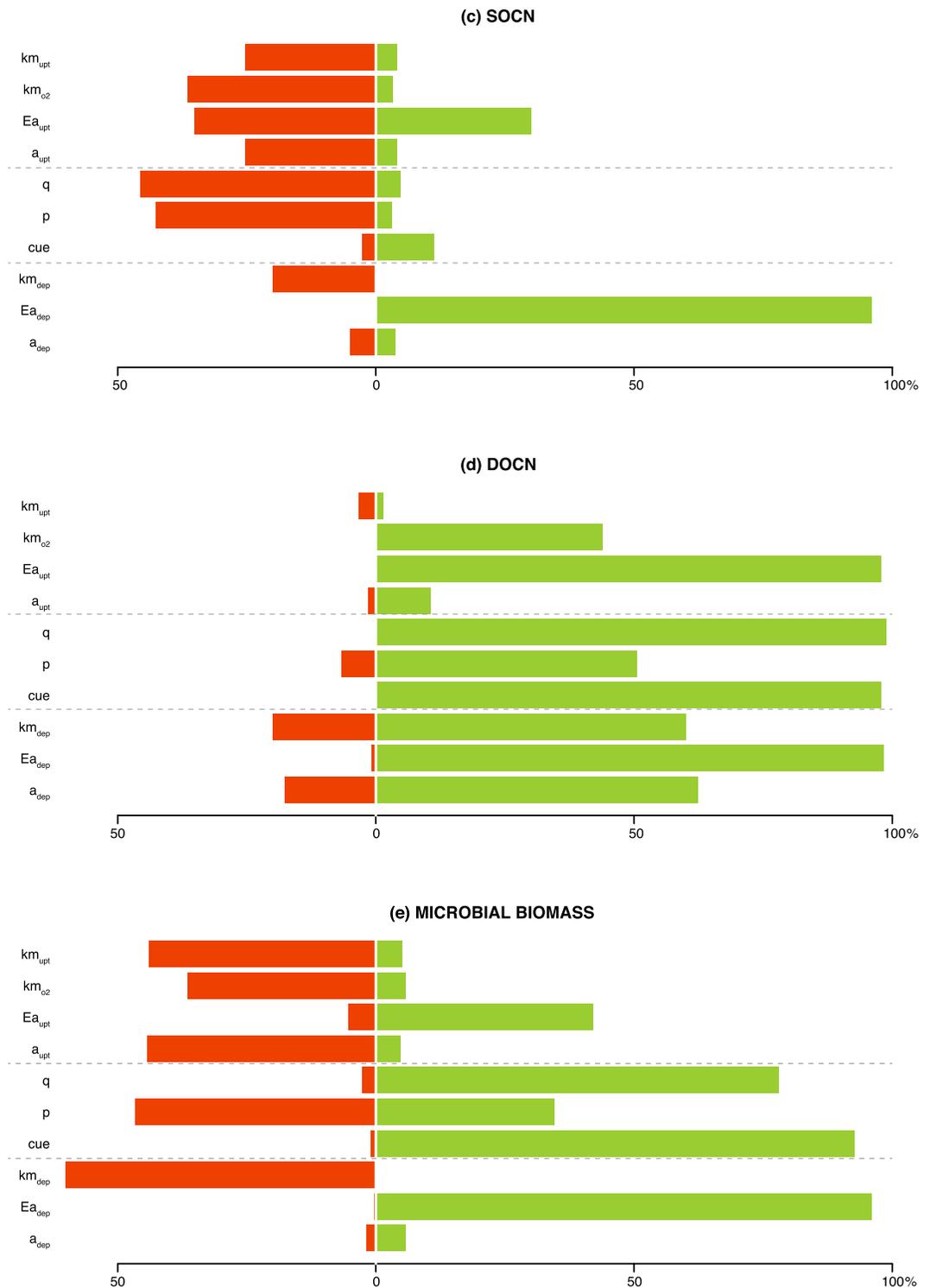
Figure 4. *Continued.*

results in large, sustained accumulations of DOCN (Figure 6b), with minimal impact on other models outputs. Updating *CUE* based on the posterior estimate (a 20% reduction in the parameter value; Table 2) results in sustained declines in microbial biomass, declining SOC stores due to reduced microbial inputs, and temporary increases in respiration, N mineralization, and DOC accumulation due to reduced uptake (Figure 6c).

Table 2
Parameter Estimates Given Field Observations of Heterotrophic Respiration

Parameter	Posterior 2.50%	Posterior median	Posterior 97.5%	Prior Lower	Prior Upper
a_{dep}	5.52×10^{10}	9.53×10^{10}	1.58×10^{11}	5.40×10^{10}	1.62×10^{11}
a_{upt}	6.07×10^{10}	1.20×10^{11}	1.60×10^{11}	5.40×10^{10}	1.62×10^{11}
CUE	3.21E-01	0.35	0.40	0.16	0.47
Ea_{dep}	59.3	60.6	61.9	30.9	92.7
Ea_{upt}	32.8	69.6	72.0	30.9	92.7
Km_{dep}	1.00×10^{-3}	2.00×10^{-3}	3.00×10^{-3}	1.25×10^{-3}	3.75×10^{-3}
Km_{o2}	6.30×10^{-2}	0.11	0.18	6.05×10^{-2}	0.18
Km_{upt}	0.15	0.23	0.42	0.15	0.45
p	0.30	0.63	0.74	0.25	0.75
q	0.53	0.70	0.75	0.25	0.75

Note. Parameter estimates for all 10 parameters estimated together using field observations of heterotrophic respiration (100,000 iteration MCMC with three chains and burn-in of 1,000).

Updating parameter values based on the assimilation of available respiration data not only impacts projections of C and N stocks, but also their associated uncertainties. For example, the assimilation of respiration data to update Ea_{dep} resulted in a large reduction of the credible interval surrounding estimates of both C (Figure 7) and N (Figure 8) stocks and fluxes compared to intervals based on broad prior distributions around the default values.

4. Discussion

Modeling coupled soil microbial C and N cycling is critical for understanding the drivers of soil organic matter sequestration, losses of C and N through mineralization, and their responses to global change. In particular, incorporating explicit representations of microbial physiology and extracellular enzyme activities has been helpful for developing biogeochemical models that are more representative of the underlying processes (Abramoff et al., 2017; Allison et al., 2010; Fatichi et al., 2019; Kyker-Snowman et al., 2020; Schimel & Weintraub, 2003; Wieder et al., 2013). A major challenge with modeling these processes involves accounting for multiple sources of uncertainty, and few studies explicitly account for the potential impacts of parameter uncertainty in particular when reporting model outputs. Furthermore, efforts to reduce model uncertainty can be improved with closer integration of data collection efforts with an analysis of the impact of available data on model performance (Xie et al., 2020).

We used the soil biogeochemical model DAMM-MCNiP to explore how parameters associated with the enzyme kinetics of depolymerization, microbial uptake of C and N, and microbial metabolism impact estimates of C and N cycling. We then used a Bayesian statistical approach to constrain parameter estimates through data assimilation, and we identified specific parameters that could be constrained through existing respiration data alone. As several parameters required additional data constraints outside of the available respiration data, we used simulated data to identify which additional data sets would be most useful targets for future data collection to reduce model parameter uncertainty. Lastly, we updated model parameters based on data assimilation and explored the potential impacts of these shifts in parameter values on long-term model projections.

4.1. Which Model Parameters Have the Greatest Impact on DAMM-MCNiP Model Outputs?

DAMM-MCNiP requires a total of 25 parameters, with 10 parameters involved in depolymerization, uptake and physiology specifically (Table 1). We focused our analyses on these parameters as they are some of the most challenging parameters to measure directly, in contrast to soil biophysical parameters (e.g., bulk density, stoichiometries of soil organic matter, and plant litter) which are routinely measured. These parameters are associated with some of the most recent advancements in soil biogeochemical modeling, as they are related to the direct representation of enzyme-mediated decomposition and microbial metabolism (Allison et al., 2010; Manzoni & Porporato, 2009; Tang, 2015).
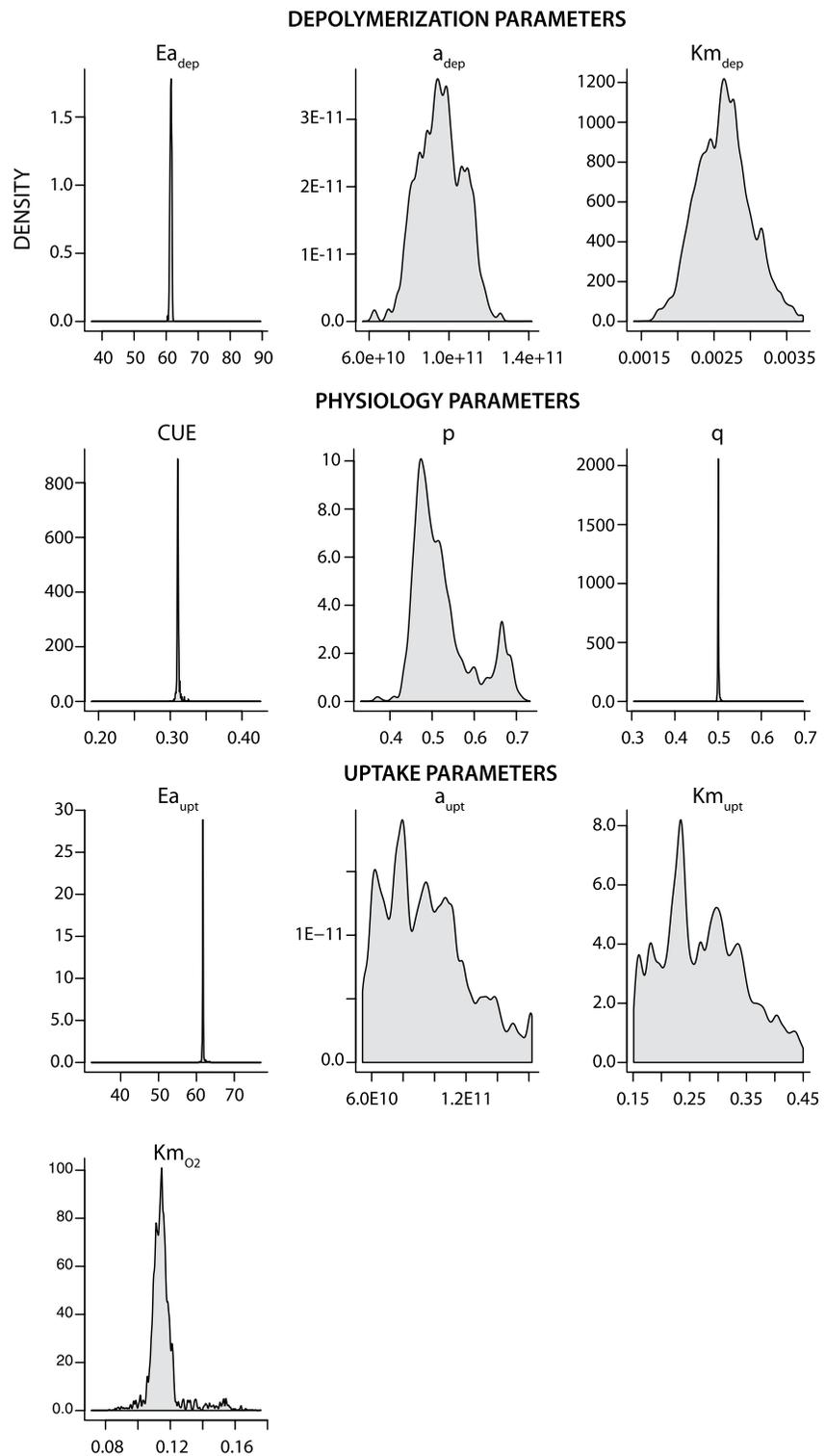
Figure 5. Note. Marginal parameter uncertainties following incorporation of simulated DOC data. Prior distributions (not shown) are uniform spanning the full range of *x*-axes.
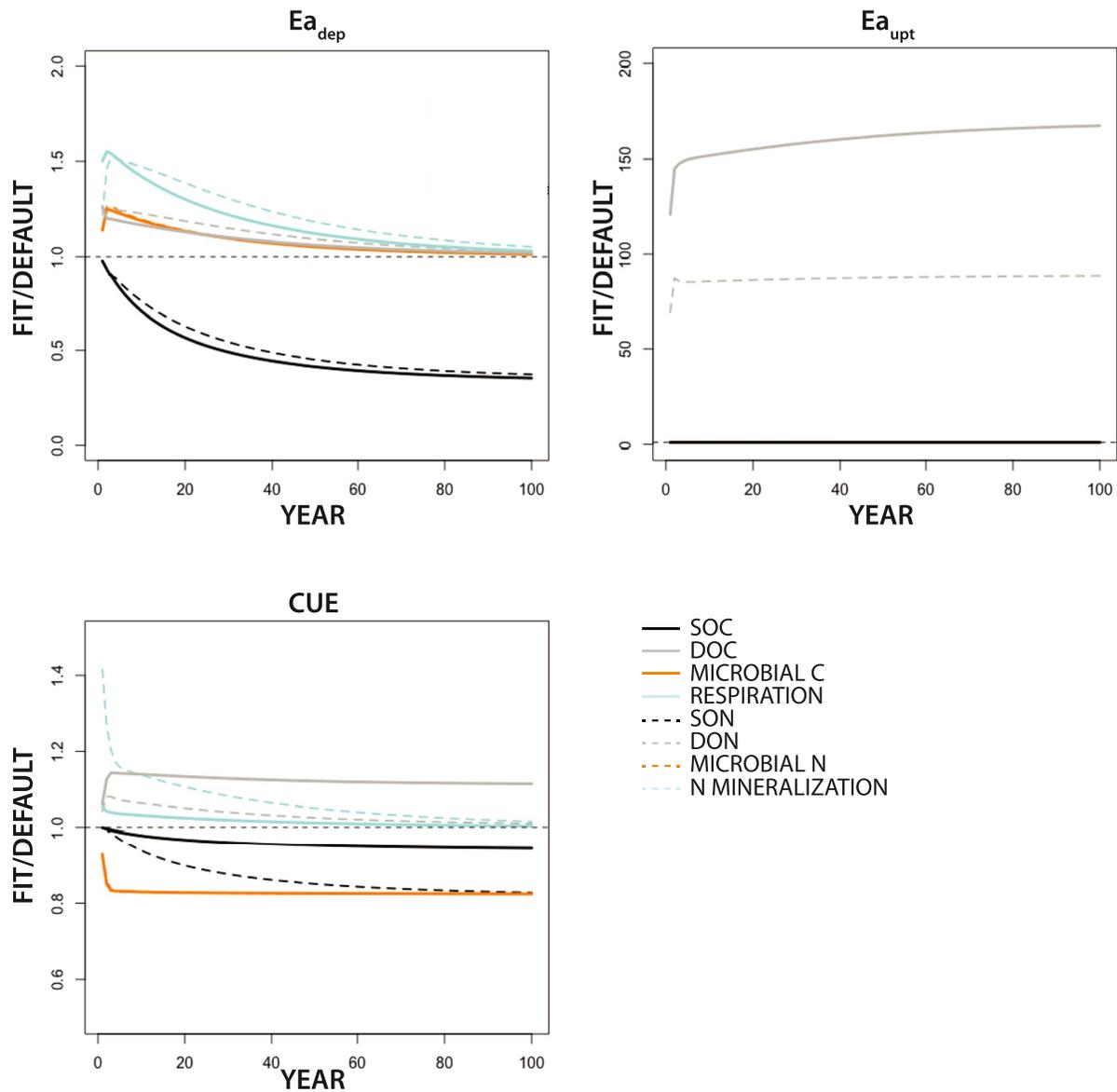
Figure 6. Modeled outputs using MAP parameter estimates. Note. Ratio of model outputs over 100 yr projections using MAP values following incorporation of observed respiration data (FIT) relative to model outputs based on default parameter value (DEFAULT) for (a) Ea_{dep}, (b) Ea_{upt}, and (c) *CUE*. Where the dotted line is not visible, it is plotted underneath the solid line.

Prioritizing efforts to constrain model parameter uncertainty should be guided in part by an assessment of which parameters have the largest impacts on model outputs of interest. In DAMM-MCNiP, model outputs are disproportionately sensitive to the activation energies of depolymerization and uptake (Ea_{dep}, Ea_{upt}), while other parameters, including the half-saturation constant of depolymerization (Km_{dep}) show minimal impact on C and N cycling (Figure 2). For example, holding temperature and moisture constant, but varying Km_{dep} by ±50% of its default value only alters estimated rates of depolymerization by 3%. Furthermore, depolymerization acts only on a small fraction of the total SOCN pools, based on the available fraction of unprotected soil organic matter and diffusion dictated by soil moisture. Therefore, the large stocks of total SOCN are essentially insensitive to this parameter in this model. In contrast, varying Ea_{dep} by ±50% of its default value leads to variation in depolymerization rates over several orders of magnitude, allowing this parameter to have detectable impacts on SOCN pool sizes.
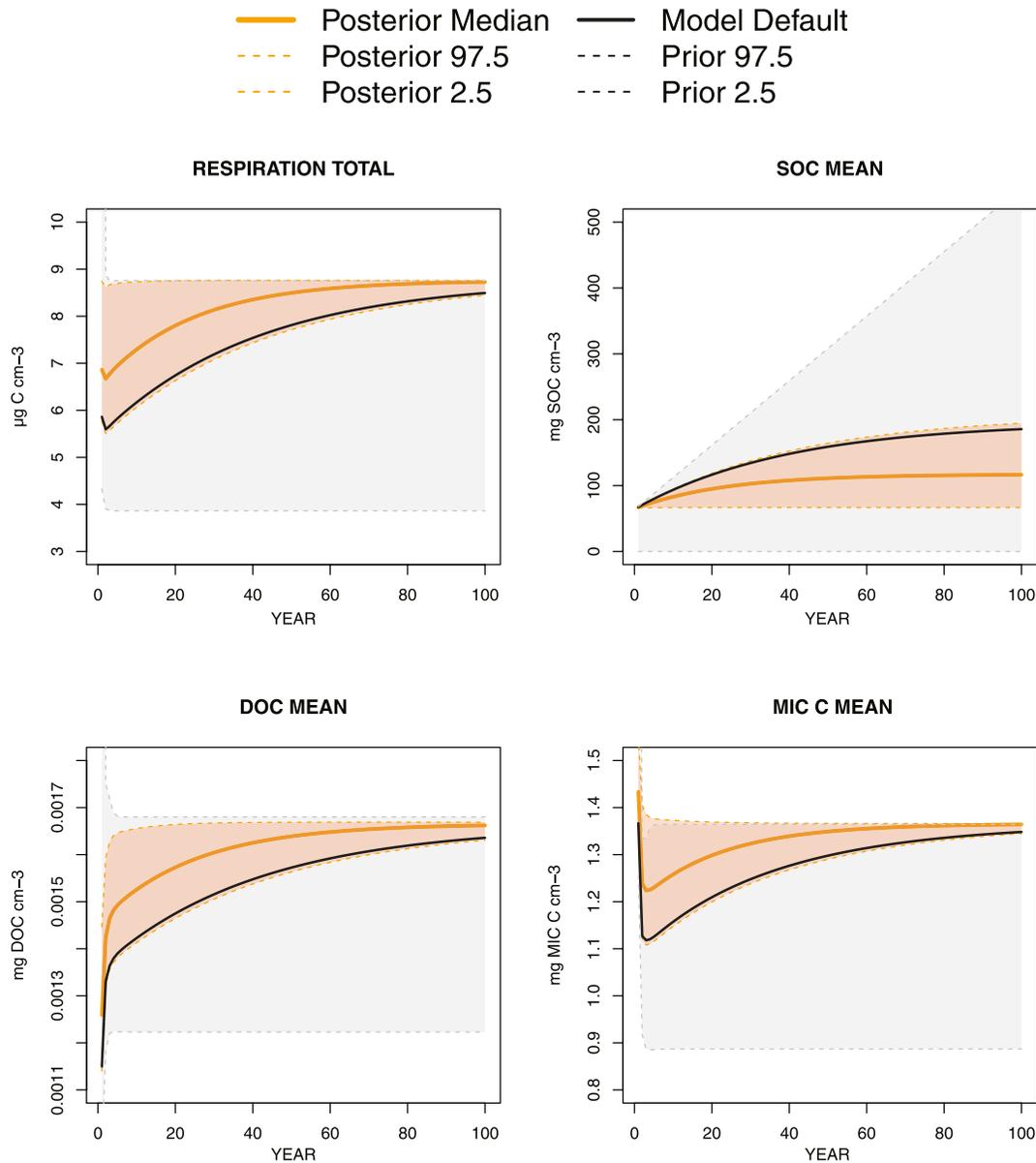
Figure 7. Modeled C outputs with prior and posterior estimates of Ea_{dep}. Note. Model estimates of (a) annual total respiration, (b) mean SOC, (c) mean DOC, and (d) mean microbial biomass C based on posterior median parameter estimates for Ea_{dep} (solid orange) and 2.5%–97.5% credible interval for posterior parameter estimate (orange interval). Solid black line shows model estimate based on default parameter setting. Gray interval shows range of estimates using prior parameter distribution.

In DAMM-MCNiP, activation energies (Ea_{dep}, Ea_{upt}) and the pre-exponential factors (a_{dep}, a_{upt}) are used to calculate the maximum reaction rates (*Vmax*) for depolymerization and uptake, which are both sensitive to temperature. In the context of global change, it is particularly important to constrain these enzyme kinetic parameters as they directly impact estimates of how SOCN stocks respond to warming (Allison et al., 2018). As *Vmax* increases exponentially in response to temperature, models parameterized according to this formulation predict positive feedbacks to warming. However, substrate supply can also limit reaction rates, resulting in observed rates that are below the predicted temperature-dependent *Vmax*. Indeed, if higher temperatures are accompanied by greater evapotranspiration and lower soil moisture, then the effect of substrate limitation (Equation 4) could offset the effect of temperature (Equation 5) on observed rates of decomposition. Climatic effects on plant inputs of soil C would also affect SOCN stocks, but are beyond the scope of this study. Collectively, these observations demonstrate the importance of constraining enzyme kinetic parameters, and articulating the contributions of uncertainties in *Ea*, *Km*, a_{dep}, and a_{upt} to driving predictions in SOCN responses to global change.
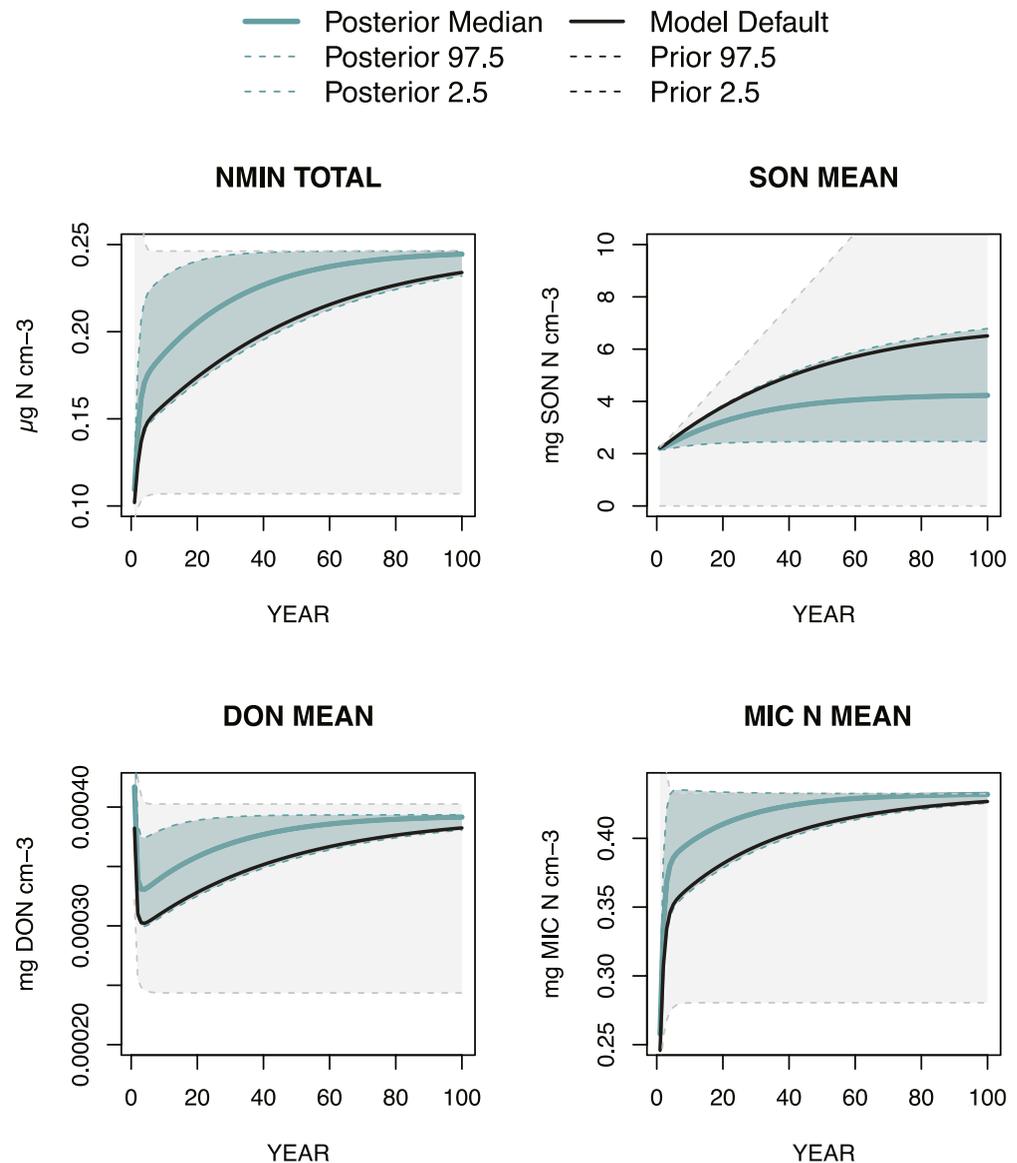
Figure 8. Modeled N outputs with prior and posterior estimates of Ea_{dep}. Note. Model estimates of (a) annual total N mineralization, (b) mean SON, (c) mean DON, and (d) mean microbial biomass N based on posterior median parameter estimates for Ea_{dep} (solid blue) and 2.5%–97.5% credible interval for posterior parameter estimate (blue interval). Solid black line shows model estimate based on default parameter setting. Gray interval shows range of estimates using prior parameter distribution.

4.2. Can the Assimilation of Existing Data Reduce Model Parameter Uncertainty, and What is the Impact of Reducing Parameter Uncertainty on Model Estimates?

In addition to analyzing the sensitivity of model outputs to variation in parameter settings, it is important to also pair these analyses with an assessment of parameter uncertainty. Focusing solely on parameter sensitivity can be misleading as highly sensitive, but tightly constrained parameters can potentially be smaller sources of uncertainty to model estimates than less-sensitive, but poorly constrained parameters (Dietze et al., 2014). Assimilating field data on heterotrophic respiration was primarily useful for constraining three parameters (Ea_{dep}, Ea_{upt}, CUE), while the remaining parameters were not reliably identifiable from respiration data alone (Figure 3). These parameters also had some of the largest impacts on modeled estimates of C and N. Thus, although respiration data

alone could not constrain all parameters, it was associated with uncertainty reductions in some of the most critical parameters to constrain.

Microbial physiological parameters like *CUE* are among the most challenging to directly measure and represent in biogeochemical models due to the diversity of microbial metabolic processes and variation across microbial taxa and substrate chemistry (Saifuddin et al., 2019; Sinsabaugh et al., 2013). The capacity to reduce uncertainty in *CUE* through the assimilation of heterotrophic respiration data may reflect the view of *CUE* as an emergent property of several intersecting microbial processes (Hagerty et al., 2018). However, reducing uncertainty in this individual parameter is itself not sufficient for determining whether or not the particular structural representation of *CUE* in this model and other similar ones is a valid way to simplify multiple sources of variation in emergent *CUE*, which may be dependent on constituent processes including microbial assimilation efficiency, biomass-specific respiration, and enzyme production (Hagerty et al., 2018).

Parameter estimates identified through data assimilation were sometimes very different from default parameter settings currently used in the published model. Even small adjustments to these parameter choices, justified by reductions in uncertainty, could have major consequences for model projections (Figure 6) and their associated uncertainties (Figures 7 and 8). For example, assimilation of the respiration data supported a small reduction in Ea_{dep} by 4%. This minimal change in a single parameter resulted in increased rates of depolymerization and reduced SOCN stocks by over 50% within 30 yr compared to the default parameter simulation (Figure 6a). Note that this is simply a reflection of model sensitivity to decomposition parameter selection rather than a projection of future global change scenarios, in which SOCN stocks would depend also on other changes which may covary with these parameters such as changes in rates of gross primary productivity, for example. Additional constraints outside of the observed respiration data would be helpful in further reducing uncertainty in parameter estimation for Ea_{dep} and other parameters. While respiration was the only direct observational data available to us, using simulated data allowed us to identify the particular types of data for future collection with the greatest potential for reducing parameter uncertainty.

4.3. What Specific Types of Data Have the Potential to Reduce Model Parameter Uncertainty Most Effectively in DAMM-MCNiP?

As respiration data alone could not be used to constrain all model parameters, we explored which potential future data sources would be most useful for reducing parameter uncertainty in DAMM-MCNiP. We found that identifying changes in DOCN pool sizes, particularly reflecting their availability at the site of microbial uptake, could reduce parameter uncertainty for most parameters (Figure 4d), while measuring SOCN pool sizes had limited utility (Figure 4c). DOCN pool sizes were also sensitive to variation in most parameters, in contrast to SOCN pool sizes which were relatively stable across parameter variation (Figure 2). We note that in the present version of DAMM-MCNiP, there is not a diffusivity function for DOCN to the enzyme reactive site as there is for the depolymerization step. Therefore, while bulk DOCN measurements alone may be useful, it is likely that modeling the diffusion of DOCN and hence DOCN concentrations at uptake sites is of equal, if not greater, importance to parameter estimation (Sihi et al., 2019).

DOCN pools are produced through depolymerization and consumed through uptake, placing them centrally in the model and making them directly responsive to both depolymerization and uptake parameters. Additionally, microbial physiological parameters indirectly impact DOCN pool sizes as the size of the microbial biomass pool impacts rates of uptake. Thus, due to their high connectedness and relatively small size compared to other pools and parameters in the model, DOCN data showed the greatest potential as a single source for constraining multiple parameters. While soil decomposition models differ in their specific representation of C and N pools, our analysis suggests that dynamic pools and fluxes that are located more centrally within model structures are most likely to constrain parameters and are a high-priority for data collection.

Classical decomposition models represent conceptual C pools with constant decay rates, which can be difficult to measure directly, in contrast to a growing trend of representing actual pools such as DOCN (Abramoff et al., 2018; Manzoni & Porporato, 2009; Robertson et al., 2019). DOCN is composed of amino acids, peptides, and other compounds which can be rapidly consumed by soil microorganisms as both C and N sources (Farrell et al., 2011, 2014; Finzi & Berthrong, 2005; Warren, 2014). A variety of methods exist for measuring this pool in the field, including *in situ* measurements through lysimetery or microdialysis (Currie et al., 1996; Inselsbacher

et al., 2011; Warren, 2014) making it possible to pair measurements of DOCN along with respiration measurements for constraining model parameters in the future. Despite these opportunities, collecting data on DOCN pools at a comparable temporal scale to the respiration data used in this study is likely to remain challenging.

4.4. Implications for Modeling Soil Biogeochemistry

DAMM-MCNiP combines the effects of temperature, soil moisture and N on C cycling through the SOCN-microbial system. Several soil biogeochemical models share similar structures to DAMM-MCNiP, featuring multiple distinct soil C pools with transfers among pools mediated by microbially produced enzymes according to Michaelis-Menten type kinetics (Huang et al., 2018; Manzoni & Porporato, 2009; Tang & Riley, 2015; Wang et al., 2013). Thus, our findings related to uptake and depolymerization kinetic parameters and microbial physiological parameters may have similar implications for models with shared core structures. However, it is critical to note that DAMM-MNCiP represents one of many potential model structural possibilities, and there exist many contrasting approaches to model structure that impact the function and interpretation of shared parameters (Fan et al., 2021; Kyker-Snowman et al., 2020; Sainte-Marie et al., 2021; Tang & Riley, 2020; Waring et al., 2020; Zhang et al., 2021).

Parameter estimation efforts in biogeochemical modeling have largely focused on aboveground processes, plant-related parameters, or simple decay rate constants for decomposition, while similar approaches in microbially explicit coupled C-N soil biogeochemical models are lagging (Luo et al., 2009). The assimilation of simulated data has previously been used to assess the identifiability of decay rate parameters in the two-pool Introductory Carbon Balance Model (ICBM; Luo et al., 2017) and initial pool sizes in the Rothamsted C model (Scharnagl et al., 2010). Both the ICBM and Rothhamsted C model lack a mechanistic representation of microbial processes, coupled C and N cycling, and enzyme-mediated depolymerization, making it necessary to extend these approaches to explore parameters in more recent soil biogeochemical models. Adopting a similar approach to the one explored here for DAMM-MCNiP will be critical to identifying the specific sources of parameter uncertainty and most promising opportunities for data collection to reduce parameter uncertainty in related soil biogeochemical models with different structures and parameters.

Current microbially explicit coupled C-N soil biogeochemical models predict widely divergent model projections in response to global change due to differences in structure and parameterization (Sulman et al., 2018; Wieder et al., 2017). A comparison of microbial biogeochemical models found that confronting conflicting models with a large synthesis of available data from field experiments on SOC responses to global change was unable to identify which models are most representative and also highlighted a consistent failure among models to predict certain empirical observations of increased SOC accumulation under warming (Sulman et al., 2018). These challenges indicate a need for more closely integrating data collection efforts with model development, both for model parameterization and model structural improvements.

Efforts on improving model representations of C and N cycling tend to focus on increasing model structural complexity to include more realistic representations of microbial physiology and enzyme-mediated decomposition. These advancements have been critical for understanding the direct controls of enzyme activities on soil organic matter storage and depolymerization, as well as the role of microbial physiology in regulating C and N cycling; however, it is equally important to consider model uncertainties associated with parameterization. The present study explored sources of parameter uncertainty and opportunities for tailoring data collection toward reducing this uncertainty for one specific model structure; however, it would be necessary to use a similar approach to compare results and opportunities across various soil biogeochemical model structures. Improving our ability to model the interactions of soil microbial physiology, soil chemistry, enzyme activities, and environmental factors on C and N cycling and their responses to global change will require quantifying model uncertainties and closely integrating future data collection with model needs.

Data Availability Statement

The DAMM-MCNiP code developed in this manuscript is archived and publicly accessible in a GitHub repository: https://github.com/rabramoff/DAMM-MCNiPv0. The C efflux measurements used in this study and related metadata can be accessed at the Harvard Forest Data Archive, a freely accessible online archive of measurements

made at the Harvard Forest Long Term Ecological Research Site. The URL for the archive is: http://harvardforest.fas.harvard.edu/harvard-forest-data-archive, and the ID for the data set is HF243-01.

References

Abramoff, R. Z., Davidson, E. A., & Finzi, A. C. (2017). A parsimonious modular approach to building a mechanistic belowground carbon and nitrogen model. *Journal of Geophysical Research: Biogeosciences*, *122*(9), 2418–2434. https://doi.org/10.1002/2017jg003796

Abramoff, R. Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Viscarra Rossel, R. A., et al. (2021). Improved global-scale predictions of soil carbon stocks with Millennial Version 2. *Soil Biology and Biochemistry*, *164*, 108466.

Abramoff, R. Z., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., et al. (2018). The Millennial model: In search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry*, *137*(1–2), 51–71. https://doi.org/10.1007/s10533-017-0409-7

Ajami, N. K., & Gu, C. (2010). Complexity in microbial metabolic processes in soil nitrogen modeling: A case for model averaging. *Stochastic Environmental Research and Risk Assessment*, *24*(6), 831–844. https://doi.org/10.1007/s00477-010-0381-4

Allison, S. D., Romero-Olivares, A. L., Lu, Y., Taylor, J. W., & Treseder, K. K. (2018). Temperature sensitivities of extracellular enzyme $Vmax$ and Km across thermal environments. *Global Change Biology*, *24*(7), 2884–2897. https://doi.org/10.1111/gcb.14045

Allison, S. D., Wallenstein, M. D., & Bradford, M. A. (2010). Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience*, *3*, 336-340. https://doi.org/10.1038/ngeo846

Ballantyne, F., IV, & Billings, S. A. (2018). Model formulation of microbial CO_2 production and efficiency can significantly influence short and long-term soil C projections. *The ISME Journal*, *12*(6), 1395–1403. https://doi.org/10.1038/s41396-018-0085-1

Blankinship, J. C., Berhe, A. A., Crow, S. E., Druhan, J. L., Heckman, K. A., Keiluweit, M., et al. (2018). Improving understanding of soil organic matter dynamics by triangulating theories, measurements, and models. *Biogeochemistry*, *140*(1), 1–13. https://doi.org/10.1007/s10533-018-0478-2

Bouskill, N. J., Tang, J., Riley, W. J., & Brodie, E. L. (2012). Trait-based representation of biological nitrification: Model development, testing, and predicted community composition. *Frontiers in Microbiology*, *3*. https://doi.org/10.3389/fmicb.2012.00364

Box, G. E. P., & Tiao, G. C. (1992). *Bayesian inference in statistical analysis*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118033197

Currie, W. S., Aber, J. D., McDowell, W. H., Boone, R. D., & Magill, A. H. (1996). Vertical transport of dissolved organic C and N under long-term N amendments in pine and hardwood forests. *Biogeochemistry*, *35*(3), 471–505. https://doi.org/10.1007/bf02183037

Davidson, E. A., Samanta, S., Caramori, S. S., & Savage, K. (2011). The Dual Arrhenius and Michaelis-Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, *18*(1), 371–384. https://doi.org/10.1111/j.1365-2486.2011.02546.x

Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., et al. (2014). A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 286–300. https://doi.org/10.1002/2013jg002392

Dijkstra, P., Salpas, E., Fairbanks, D., Miller, E. B., Hagerty, S. B., van Groenigen, K. J., et al. (2015). High carbon use efficiency in soil microbial communities is related to balanced growth, not storage compound synthesis. *Soil Biology and Biochemistry*, *89*, 35–43. https://doi.org/10.1016/j.soilbio.2015.06.021

Drake, J. E., Giasson, M.-A., Spiller, K. J., & Finzi, A. C. (2013). Seasonal plasticity in the temperature sensitivity of microbial activity in three temperate forest soils. *Ecosphere*, *4*(6), 1–21. https://doi.org/10.1890/es13-00020.1

Fan, X., Gao, D., Zhao, C., Wang, C., Qu, Y., Zhang, J., & Bai, E. (2021). Improved model simulation of soil carbon cycling by representing the microbially derived organic carbon pool. *ISME*, *15*, 2248–2263. https://doi.org/10.1038/s41396-021-00914-0

Farrell, M., Hill, P. W., Wanniarachchi, S. D., Farrar, J., Bardgett, R. D., & Jones, D. L. (2011). Rapid peptide metabolism: A major component of soil nitrogen cycling? *Global Biogeochemical Cycles*, *25*(3). https://doi.org/10.1029/2010gb003999

Farrell, M., Prendergast-Miller, M., Jones, D. L., Hill, P. W., & Condron, L. M. (2014). Soil microbial organic nitrogen uptake is regulated by carbon availability. *Soil Biology and Biochemistry*, *77*, 261–267. https://doi.org/10.1016/j.soilbio.2014.07.003

Fatichi, S., Manzoni, S., Or, D., & Paschalis, A. (2019). A mechanistic model of microbially mediated soil biogeochemical processes: A reality check. *Global Biogeochemical Cycles*, *33*(6), 620–648. https://doi.org/10.1029/2018gb006077

Finzi, A. C., Abramoff, R. Z., Spiller, K. S., Brzostek, E. R., Darby, B. A., Kramer, M. A., & Phillips, R. P. (2015). Rhizosphere processes are quantitatively important components of terrestrial carbon and nutrient cycles. *Global Change Biology*, *21*(5), 2082–2094. https://doi.org/10.1111/gcb.12816

Finzi, A. C., & Berthrong, S. T. (2005). The uptake of amino acids by microbes and trees in three cold-temperate forests. *Ecology*, *86*(12), 3345–3353. https://doi.org/10.1890/04-1460

Finzi, A. C., Giasson, M.-A., Barker Plotkin, A. A., Aber, J. D., Boose, E. R., Davidson, E. A., et al. (2020). Carbon budget of the Harvard Forest long-term ecological research site: Pattern, process, and response to global change. *Ecological Monographs*, *90*(4), e01423. https://doi.org/10.1002/ecm.1423

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global carbon budget 2020. *Earth System Science Data*, *12*(4), 3269–3340. https://doi.org/10.5194/essd-12-3269-2020

Georgiou, K., Malhotra, A., Wieder, W. R., Ennis, J. H., Hartman, M. D., Sulman, B. N., et al. (2021). Divergent controls of soil organic carbon between observations and process-based models. *Biogeochemistry*, *156*, 5–17. https://doi.org/10.1007/s10533-021-00819-2

Geyer, K. M., Kyker-Snowman, E., Grandy, A. S., & Frey, S. D. (2016). Microbial carbon use efficiency: Accounting for population, community, and ecosystem-scale controls over the fate of metabolized organic matter. *Biogeochemistry*, *127*(2–3), 173–188. https://doi.org/10.1007/s10533-016-0191-y

Hagerty, S. B., Allison, S. D., & Schimel, J. P. (2018). Evaluating soil microbial carbon use efficiency explicitly as a function of cellular processes: Implications for measurements and models. *Biogeochemistry*, *140*(3), 269–283. https://doi.org/10.1007/s10533-018-0489-z

Hartig, F., Minunno, F., Paul, S., Cameron, D., Ott, T., & Pichler, M. (2018). *BayesianTools Package*. Retrieved from https://github.com/florianhartig/BayesianTools

Huang, Y., Guenet, B., Ciais, P., Janssens, I. A., Soong, J. L., Wang, Y., et al. (2018). ORCHIMIC (v1.0), a microbe-mediated model for soil organic matter decomposition. *Geoscientific Model Development*, *11*(6), 2111–2138. https://doi.org/10.5194/gmd-11-2111-2018

Hungate, B. A. (2003). Atmospheric science: Nitrogen and climate change. *Science*, *302*(5650), 1512–1513. https://doi.org/10.1126/science.1091390

Inselsbacher, E., Öhlund, J., Jämtgård, S., Huss-Danell, K., & Näsholm, T. (2011). The potential of microdialysis to monitor organic and inorganic nitrogen compounds in soil. *Soil Biology and Biochemistry*, *43*(6), 1321–1332. https://doi.org/10.1016/j.soilbio.2011.03.003

Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, *10*(2), 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:tvdoso]2.0.co;2

Kyker-Snowman, E., Wieder, W. R., Frey, S. D., & Grandy, A. S. (2020). Stoichiometrically coupled carbon and nitrogen cycling in the MIcrobial-MIneral Carbon Stabilization model version 1.0 (MIMICS-CN v1.0). *Geoscientific Model Development*, *13*(9), 4413–4434. https://doi.org/10.5194/gmd-13-4413-2020

Lehmann, J., Hansel, C. M., Kaiser, C., Kleber, M., Maher, K., Manzoni, S., et al. (2020). Persistence of soil organic carbon caused by functional complexity. *Nature Geoscience*, *13*, 529–534. https://doi.org/10.1038/s41561-020-0612-3

Liebig, J. (1842). *Chemistry and its application to agriculture and physiology* (Third American edition). John Owen.

Lu, D., Ricciuto, D., Walker, A., Safta, C., & Munger, W. (2017). Bayesian calibration of terrestrial ecosystem models: A study of advanced Markov chain Monte Carlo methods. *Biogeosciences*, *14*(18), 4295–4314. https://doi.org/10.5194/bg-14-4295-2017

Luo, Y., & Schuur, E. A. G. (2019). Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. *Global Change Biology*, *26*(3), 1109–1117. https://doi.org/10.1111/gcb.14939

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*(3), 571–574. https://doi.org/10.1890/08-0561.1

Luo, Z., Wang, E., & Sun, O. J. (2017). Uncertain future soil carbon dynamics under global change predicted by models constrained by total carbon measurements. *Ecological Applications*, *27*(3), 1001–1009. https://doi.org/10.1002/eap.1504

Magill, A. H., Aber, J. D., Berntson, G. M., McDowell, W. H., Nadelhoffer, K. J., Melillo, J. M., & Steudler, P. (2000). Long-term nitrogen additions and nitrogen saturation in two temperate forests. *Ecosystems*, *3*(3), 238–253. https://doi.org/10.1007/s100210000023

Manzoni, S., & Porporato, A. (2009). Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry*, *41*(7), 1355–1379. https://doi.org/10.1016/j.soilbio.2009.02.031

Marschmann, G. L., Pagel, H., Kügler, P., & Streck, T. (2019). Equifinality, sloppiness, and emergent structures of mechanistic soil biogeochemical models. *Environmental Modelling & Software*, *122*, 104518. https://doi.org/10.1016/j.envsoft.2019.104518

Moorhead, D. L., & Sinsabaugh, R. L. (2006). A theoretical model of litter decay and microbial interaction. *Ecological Monographs*, *76*(2), 151–174. https://doi.org/10.1890/0012-9615(2006)076[0151:atmold]2.0.co;2

Myrgiotis, V., Rees, R. M., Topp, C. F. E., & Williams, M. (2018). A systematic approach to identifying key parameters and processes in agroecosystem models. *Ecological Modelling*, *368*, 344–356. https://doi.org/10.1016/j.ecolmodel.2017.12.009

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Robertson, A. D., Paustian, K., Ogle, S., Wallenstein, M. D., Lugato, E., & Cotrufo, M. F. (2019). Unifying soil organic matter formation and persistence frameworks: The MEMS model. *Biogeosciences*, *16*(6), 1225–1248. https://doi.org/10.5194/bg-16-1225-2019

Saifuddin, M., Bhatnagar, J. M., Segrè, D., & Finzi, A. C. (2019). Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11488-z

Sainte-Marie, J., Barrandon, M., Saint-André, L., Gelhaye, E., Martin, F., & Derrien, D. (2021). C-STABILITY an innovative modeling framework to leverage the continuous representation of organic matter. *Nature Communications*, *12*, 810. https://doi.org/10.1038/s41467-021-21079-6

Savage, K., Davidson, E. A., & Richardson, A. D. (2008). A conceptual and practical approach to data quality and analysis procedures for high-frequency soil respiration measurements. *Functional Ecology*, *22*(6), 1000–1007. https://doi.org/10.1111/j.1365-2435.2008.01414.x

Scharnagl, B., Vrugt, J. A., Vereecken, H., & Herbst, M. (2010). Information content of incubation experiments for inverse estimation of pools in the Rothamsted carbon model: A Bayesian perspective. *Biogeosciences*, *7*(2), 763–776. https://doi.org/10.5194/bg-7-763-2010

Schimel, J. P., & Weintraub, M. N. (2003). The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: A theoretical model. *Soil Biology and Biochemistry*, *35*(4), 549–563. https://doi.org/10.1016/s0038-0717(03)00015-4

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., et al. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, *478*(7367), 49–56. https://doi.org/10.1038/nature10386

Shi, Z., Crowell, S., Luo, Y., & Moore, B., III. (2018). Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04526-9

Shi, Z., Yang, Y., Zhou, X., Weng, E., Finzi, A. C., & Luo, Y. (2015). Inverse analysis of coupled carbon-nitrogen cycles against multiple data sets at ambient and elevated CO_2. *Journal of Plant Ecology*, *9*(3), 285–295. https://doi.org/10.1093/jpe/rtv059

Sihi, D., Davidson, E. A., Savage, K. E., & Liang, D. (2019). Simultaneous numerical representation of soil microsite production and consumption of carbon dioxide, methane, and nitrous oxide using probability distribution functions. *Global Change Biology*, *26*(1), 200–218. https://doi.org/10.1111/gcb.14855

Sinsabaugh, R. L., Belnap, J., Findlay, S. G., Shah, J. J. F., Hill, B. H., Kuehn, K. A., et al. (2014). Extracellular enzyme kinetics scale with resource availability. *Biogeochemistry*, *121*(2), 287–304. https://doi.org/10.1007/s10533-014-0030-y

Sinsabaugh, R. L., Manzoni, S., Moorhead, D. L., & Richter, A. (2013). Carbon use efficiency of microbial communities: Stoichiometry, methodology and modelling. *Ecology Letters*. https://doi.org/10.1111/ele.12113

Sinsabaugh, R. L., Shah, J. J. F., Findlay, S. G., Kuehn, K. A., & Moorhead, D. L. (2015). Scaling microbial biomass, metabolism and resource supply. *Biogeochemistry*, *122*(2–3), 175–190. https://doi.org/10.1007/s10533-014-0058-z

Sistla, S. A., Rastetter, E. B., & Schimel, J. P. (2014). Responses of a tundra system to warming using SCAMPS: A stoichiometrically coupled, acclimating microbe-plant-soil model. *Ecological Monographs*, *84*(1), 151–170. https://doi.org/10.1890/12-2119.1

Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., et al. (2013). The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture. *Ecosystems & Environment*, *164*, 80–99. https://doi.org/10.1016/j.agee.2012.10.001

Sulman, B. N., Moore, J. A. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., et al. (2018). Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry*, *141*(2), 109–123. https://doi.org/10.1007/s10533-018-0509-z

Tang, J. (2015). On the relationships between the Michaelis-Menten kinetics, reverse Michaelis-Menten kinetics, equilibrium chemistry approximation kinetics, and quadratic kinetics. *Geoscientific Model Development*, *8*(12), 3823–3835. https://doi.org/10.5194/gmd-8-3823-2015

Tang, J., & Riley, W. J. (2015). Weaker soil carbon-climate feedbacks resulting from microbial and abiotic interactions. *Nature Climate Change*, *5*(1), 56–60. https://doi.org/10.1038/nclimate2438

Tang, J., & Riley, W. J. (2020). Linear two-pool models are insufficient to infer soil organic matter decomposition temperature sensitivity from incubations. *Biogeochemistry*, *149*, 251–261. https://doi.org/10.1007/s10533-020-00678-3

Ter Braak, C. J. F., & Vrugt, J. A. (2008). Differential evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, *18*(4), 435–446. https://doi.org/10.1007/s11222-008-9104-9

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., & Mahowald, N. M. (2007). Influence of carbon-nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability. *Global Biogeochemical Cycles*, *21*(4). https://doi.org/10.1029/2006gb002868

Wang, G., Post, W. M., & Mayes, M. A. (2013). Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. *Ecological Applications*, *23*(1), 255–272. https://doi.org/10.1890/12-0681.1

Waring, B. G., Averill, C., & Hawkes, C. V. (2013). Differences in fungal and bacterial physiology alter soil carbon and nitrogen cycling: Insights from meta-analysis and theoretical models. *Ecology Letters*, *16*(7), 887–894. https://doi.org/10.1111/ele.12125

Waring, B. G., Sulman, B. N., Reed, S., Smith, A. P., Averill, C., Creamer, C. A., et al. (2020). From pools to flow: The PROMISE framework for new insights on soil carbon cycling in a changing world. *Global Change Biology*, *26*, 6631–6643. https://doi.org/10.1111/gcb.15365

Warren, C. R. (2014). Organic N molecules in the soil solution: What is known, what is unknown and the path forwards. *Plant and Soil*, *375*(1–2), 1–19. https://doi.org/10.1007/s11104-013-1939-y

Weintraub, S. R., Flores, A. N., Wieder, W. R., Sihi, D., Cagnarini, C., Gonçalves, D. R. P., et al. (2019). Leveraging environmental research and observation networks to advance soil carbon science. *Journal of Geophysical Research: Biogeosciences*, *124*(5), 1047–1055. https://doi.org/10.1029/2018jg004956

Wieder, W. R., Bonan, G. B., & Allison, S. D. (2013). Global soil carbon projections are improved by modeling microbial processes. *Nature Climate Change*, *3*(10), 909–912. https://doi.org/10.1038/nclimate1951

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., & Bonan, G. B. (2014). Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences*, *11*(14), 3899–3917. https://doi.org/10.5194/bg-11-3899-2014

Wieder, W. R., Hartman, M. D., Sulman, B. N., Wang, Y.-P., Koven, C. D., & Bonan, G. B. (2017). Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Change Biology*, *24*(4), 1563–1579. https://doi.org/10.1111/gcb.13979

Xie, H. W., Romero-Olivares, A. L., Guindani, M., & Allison, S. D. (2020). A Bayesian approach to evaluation of soil biogeochemical models. *Biogeosciences*, *17*(15), 4043–4057. https://doi.org/10.5194/bg-17-4043-2020

Zaehle, S. (2013). Terrestrial nitrogen-carbon cycle interactions at the global scale. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621), 20130125. https://doi.org/10.1098/rstb.2013.0125

Zhang, H., Goll, D. S., Wang, Y., Ciais, P., Wieder, W. R., Abramoff, R., et al. (2020). Microbial dynamics and soil physicochemical properties explain large-scale variations in soil organic carbon. *Global Change Biology*, *26*(4), 2668–2685. https://doi.org/10.1111/gcb.14994

Zhang, Y., Lavallee, J. M., Robertson, A. D., Even, R., Ogle, S. M., Paustian, K., & Cotrufo, M. F. (2021). Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model. *Biogeosciences*, *18*, 3147–3171. https://doi.org/10.5194/bg-18-3147-2021

Zhou, J., Chen, S., Yan, L., Wang, J., Jiang, M., Liang, J., et al. (2021). A comparison of linear conventional and nonlinear microbial models for simulating pulse dynamics of soil heterotrophic respiration in a semi-arid grassland. *Journal of Geophysical Research: Biogeosciences*, *126*(5). https://doi.org/10.1029/2020jg006120