Title
Toward more realistic projections of soil carbon dynamics by Earth system models

Permalink
https://escholarship.org/uc/item/1pw7g2r2

Journal
Global Biogeochemical Cycles, 30(1)

ISSN
0886-6236

Authors
Luo, Y
Ahlström, A
Allison, SD
et al.

Publication Date
2016

DOI
10.1002/2015GB005239

Peer reviewed
Explicitly representing soil microbial processes in Earth system models

William R. Wieder1, Steven D. Allison2,3, Eric A. Davidson4, Katerina Georgiou5,6, Oleksandra Hararuk7, Yujie He1,8, Francesca Hopkins1,9, Yiqi Luo10, Matthew Smith11, Benjamin Sulman12, Katherine Todd-Brown10,13, Ying-Ping Wang14, Jianyang Xia10,15, and Xiaofeng Xu16

1Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado, USA, 2Department of Ecology and Evolutionary Biology, University of California, Irvine, California, USA, 3Department of Earth System Science, University of California, Irvine, California, USA, 4Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, Maryland, USA, 5Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California, USA, 6Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA, 7Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada, 8Purdue University, West Lafayette, Indiana, USA, 9NASA Jet Propulsion Laboratory, Pasadena, California, USA, 10University of Oklahoma, Norman, Oklahoma, USA, 11Microsoft Research, Cambridge, UK, 12Indiana University, Bloomington, Indiana, USA, 13Pacific Northwest National Laboratory, Richland, Washington, USA, 14CSIRO Ocean and Atmosphere Flagship, Aspendale, Victoria, Australia, 15Tianlong National Forest Ecosystem Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China, 16Department of Biological Sciences, University of Texas at El Paso, Texas, USA

Abstract Microbes influence soil organic matter decomposition and the long-term stabilization of carbon (C) in soils. We contend that by revising the representation of microbial processes and their interactions with the physicochemical soil environment, Earth system models (ESMs) will make more realistic global C cycle projections. Explicit representation of microbial processes presents considerable challenges due to the scale at which these processes occur. Thus, applying microbial theory in ESMs requires a framework to link micro-scale process-level understanding and measurements to macro-scale models used to make decadal- to century-long projections. Here we review the diversity, advantages, and pitfalls of simulating soil biogeochemical cycles using microbial-explicit modeling approaches. We present a roadmap for how to begin building, applying, and evaluating reliable microbial-explicit model formulations that can be applied in ESMs. Drawing from experience with traditional decomposition models, we suggest the following: (1) guidelines for common model parameters and output that can facilitate future model intercomparisons; (2) development of benchmarking and model-data integration frameworks that can be used to effectively guide, inform, and evaluate model parameterizations with data from well-curated repositories; and (3) the application of scaling methods to integrate microbial-explicit soil biogeochemistry modules within ESMs. With contributions across scientific disciplines, we feel this roadmap can advance our fundamental understanding of soil biogeochemical dynamics and more realistically project likely soil C response to environmental change at global scales.

1. Introduction

Developing models that accurately simulate belowground processes is an important challenge for soil, environmental, and Earth sciences [Hinckley et al., 2014; Schmidt et al., 2011; Todd-Brown et al., 2013]. Although microbial processes fundamentally regulate the decomposition and stabilization of soil organic matter (SOM), our numerical representation of these processes has lagged behind our conceptual understanding of soil microbial and biogeochemical dynamics [Davidson et al., 2014]. Recent work, however, indicates that updating the assumptions underlying biogeochemical models can considerably shift the magnitude of projected soil carbon (C) in response to environmental perturbations [Allison et al., 2010; Sulman et al., 2014; Wieder et al., 2013, 2015]. Here we define “microbial models” as soil biogeochemical models that mathematically couple microbial biomass and C substrate pools. Significant efforts are still required to harmonize microbial models, observations [Sihi et al., 2015; Wang et al., 2014, 2015b], and theory, especially relating to physicochemical stabilization of SOM [Cotrufo et al., 2013; Grandy et al., 2009]. These findings highlight the need to refine our understanding of factors controlling soil C, the largest terrestrial C pool on Earth, and its representation in models. We acknowledge that representing microbial and mineral controls over soil biogeochemical processes is critical, although here we largely focus on biotic processes.
Currently, soil biogeochemical processes in Earth system models (ESMs) represent biological activity implicitly [Schimel, 2001], assuming that respired CO₂ is proportional to the soil C stock. These models were conceived to explain patterns from long-term incubations and parameterized to fit plot or field-scale stock and flux observations [Jenkinson and Rayner, 1977; Van Veen and Paul, 1981]. They typically bin SOM into one or more pools and describe the transfer of C between these pools and to the atmosphere using first-order linear decay rates that are modified by environmental scalars (e.g., ROTH-C and CENTURY) [Jenkinson et al., 1987; Parton et al., 1987]. These models have been critical in the development, understanding, and projection of soil biogeochemical processes at ecosystem and Earth system scales. Discrete SOM pools, however, are often not easy to relate to measurable quantities [Smith et al., 2002] and may not accurately represent heterogeneous SOM [Sierra et al., 2011]. These first-order models implicitly represent soil biology under the assumptions that (1) microbial dynamics occur faster than the long-term fluxes of interest, (2) biological responses to environmental and edaphic conditions are invariant across ecosystems and through time, and (3) microbial communities have functional equivalence allowing them to optimally process the available SOM [Bradford and Fierer, 2012]. As with any model, SOM predictions should be robust over the domain for which they were parameterized and where these simplifying assumptions hold. However, assumptions about biological homogeneity and functional equivalence are increasingly being called into question, especially when first-order models are applied in new spatiotemporal settings or forced to respond to novel ecoclimatological conditions [Hawkes and Keitt, 2015; Karhu et al., 2014; Keiser et al., 2011; Lawrence et al., 2009; Manzoni and Porporato, 2009]. Understanding the limitations of such predictions is critical when considering the accuracy of potential C cycle-climate feedbacks generated by ESMs.

When SOM models are applied at global scales, the empirical relationships on which the models are built must be extrapolated over space and through time to predict soil C dynamics across diverse physical, chemical, and biological conditions [Bradford and Fierer, 2012; Ise and Moorcroft, 2006]. Presently, this approach produces estimates of contemporary soil C storage that vary sixfold across models represented in the Coupled Model Intercomparison Project Phase 5 (CMIP5) [Todd-Brown et al., 2013], demonstrating the need for improvements in soil biogeochemical representations [e.g., Lawrence et al., 2009; Moorhead and Sinsabaugh, 2006]. A key motivation for greater process-level representation is to improve our ability to generate realistic projections from ESMs. For example, when the current generations of ESMs are applied in global change scenarios, they provide highly uncertain projections about the fate of soil C in response to environmental change [Todd-Brown et al., 2014]. Data integration offers a promising route for reducing some of the uncertainty generated by this class of models [Hararuk et al., 2014; Xia et al., 2013]. We believe that microbial-explicit models, once properly developed and integrated with ESMs, will provide better predictions and understanding of the terrestrial C cycle, especially when representing transient dynamics. We illustrate this point by highlighting three global change drivers that will alter soil biogeochemistry with a potential to change future soil C stocks: changing temperature, nitrogen (N) enrichment, and elevated CO₂.

The response of decomposition to temperature change is critical in anticipating the response of SOM stocks over the next hundred years. In first-order models, fixed parameters (e.g., Q₁₀) describe the temperature sensitivity of SOM decomposition [Davidson and Janssens, 2006]. Use of these environmental scalars is ubiquitous across ecosystem and Earth system models [Sierra et al., 2012a; Todd-Brown et al., 2013; Xia et al., 2013] and critical in determining steady state and projected changes in soil C pools [Exbrayat et al., 2014; Todd-Brown et al., 2014; Wieder et al., 2014a]. However, experimental evidence demonstrates that both soil microbial physiology and community composition change with warming [Bradford, 2013; Crowther and Bradford, 2013; Frey et al., 2008, 2013; Hagerty et al., 2014; Hall et al., 2010; Hawkes and Keitt, 2015; Karhu et al., 2014; Treseder et al., 2012], potentially leading to thermal acclimation of soil respiration rates [Bradford et al., 2008, 2010]. Physiological acclimation or community shifts under environmental perturbations may alter soil functioning in ways that challenge the assumptions and underlying structure of first-order models. This suggests that where the rate- or fate-controlling step in SOM turnover is microbial [Schimel and Sack, 2012], more mechanistic representations of soil microbial activity are necessary to accurately represent soil biogeochemical response to projected warming. The response of the physicochemical soil environment to changes in temperature is also important. For example, observation of increased substrate availability in early phases of soil warming experiments suggests that the rate at which soil C substrates become available for microbial uptake has a different temperature sensitivity than
microbial metabolism [Hopkins et al., 2012]. Although biotic and abiotic dynamics strongly influence model behavior [Sierra et al., 2011], current ESMs typically do not distinguish between these distinct biogeochemical processes.

Future nitrogen enrichment and atmospheric CO₂ levels will alter soil microbial community dynamics and biogeochemical function, as demonstrated from the laboratory to ecosystem scale. Nitrogen enrichment typically increases plant productivity and likely favors aboveground plant C allocation [Elser et al., 2007; LeBauer and Treseder, 2008; Liu and Greaver, 2010; Reynolds and D’Antonio, 1996; Xia and Wan, 2008]. Belowground responses to N enrichment include shifts in microbial community composition that often cause declines in fungal biomass and oxidative enzyme activity, while favoring microbes that are characterized by copiotrophic growth strategies that have higher growth rates and nutritional requirements [Fierer et al., 2007, 2012; Frey et al., 2004; Gallo et al., 2004; Ramirez et al., 2012; Saiya-Cork et al., 2002; Waldrop et al., 2004]. Collectively, changes in soil microbial activity and composition following N enrichment influence decomposition rates [Fog, 1988; Hobbie, 2008; Hobbie et al., 2012; Knorr et al., 2005], elicit declines in microbial biomass, and result in modest to negligible changes in soil C storage [Janssens et al., 2010; Liu and Greaver, 2010; Lu et al., 2011; Treseder, 2008]. Notably, Neff and others [2002] demonstrate the complex interactions of long-term N enrichment on different soil C pools, concluding that existing models were unable to capture key components of N effects on soil C storage [see also Sierra et al., 2012b]. Preliminary simulations, however, suggest that explicitly considering microbial physiology and functional composition in a simplified form may offer a path forward [Wieder et al., 2015].

Similarly, elevated CO₂ may increase plant productivity and shift plant C allocation strategies [Drake et al., 2011; Norby et al., 2004; Phillips et al., 2011], although these responses interact with N availability [Norby and Zak, 2011; Reich et al., 2006]. N constraints can result in negligible changes in soil C stocks or in some cases soil C losses [Hungate et al., 2009; Phillips et al., 2012; van Groenigen et al., 2014]. Again, first-order models generally are unable to capture observed responses from elevated CO₂ manipulations [Zaabale et al., 2014] because they project linear increases in soil C storage in response to increasing plant productivity, unless decomposition rates are modified by changes in litter inputs [van Groenigen et al., 2014; Wutzler and Reichstein, 2008]. In contrast, microbial-explicit models can represent observed priming effects by capturing the functional response of microbial communities to increased substrate availability [Drake et al., 2013; Sulman et al., 2014].

Explicitly considering the role of biology and the physicochemical soil environment could provide new insight into potential mechanisms responsible for ecosystem responses to diverse perturbations. We acknowledge that a growing body of literature documents the importance of abiotic factors, and biotic-abiotic interactions, in stabilizing SOM [Cotrufo et al., 2013; Dungait et al., 2012; Kleber et al., 2011; Miltner et al., 2012], but here we broadly focus on representing the microbial response to perturbations. For example, explicitly representing microbial physiological response to higher temperature can reduce the positive feedback between warming and simulated soil C decomposition through several pathways, including reductions in microbial growth efficiency (MGE) [Allison et al., 2010; Wieder et al., 2013], increases in microbial turnover [Hagerty et al., 2014], or shifts in microbe-mineral interactions [Tang and Riley, 2014]. Similarly, N enrichment can drive shifts in soil microbial communities in ways that affect soil biogeochemical function and C storage [Fierer et al., 2012; Gallo et al., 2004; Liu and Greaver, 2010]. While we have evidence that these microbial responses do occur, we have little knowledge of how important they are quantitatively as factors affecting soil C stocks [Neff et al., 2002]. We contend that mechanistic microbial models can better simulate soil C responses to environmental change, thus improving the accuracy of soil C projections—provided they can overcome other challenges that are reviewed here. Such efforts, however, remain in their infancy. To advance this field of research we survey the diversity of structures and applications of microbial-explicit models and discuss some of the particular drawbacks and challenges of these approaches. We also present a roadmap to facilitate more efficient, productive development and evaluation of microbial-explicit models. Specifically, we describe the modeling tools that need to be developed and suggest data sets that can be used to validate and parameterize microbial-explicit soil biogeochemical models for ESMs. The long-term goal of this work is to use theory, data, and models to explain and replicate SOM responses to environmental change, thus improving the accuracy of C cycle projections from ESMs and advancing our understanding of soil microbial and biogeochemical responses to environmental change.
2. Survey of Current Microbial-Explicit Models (Including Drawbacks and Challenges)

2.1. Model Structures

We briefly review various decomposition model structures; however, more detailed analysis and history of model development can be found in Manzoni and Porporato [2009]. First-order biogeochemical models assume that soil carbon decomposition depends linearly on the size of the soil carbon pool \((C)\) and a rate constant \((k)\) according to

\[
\frac{dC}{dt} = l - kC,
\]

where \(l\) is carbon input to a soil C pool. The decay rate constant \(k\) varies with soil environmental conditions, such as soil temperature, moisture, and texture [Parton et al., 1988; Randerson et al., 1996], but is independent of soil C and microbial biomass. Some first-order models include nutrients with nonlinear functions that capture the costs and benefits of nutrient acquisition [Rastetter et al., 2001, 2005; Wang et al., 2007; Houlton et al., 2008], but generally apply a first-order approach to soil C decomposition (equation (1)). Although a microbial role in C decomposition has long been recognized [Burns, 1982; Parnas, 1975; Sinsabaugh, 1994; Tenney and Waksman, 1929], microbial feedbacks are implied but not explicitly represented by first-order biogeochemical models, including all of the CMIP5 models analyzed in the latest report by the International Panel on Climate Change [Ciais et al., 2013; Todd-Brown et al., 2013].

In contrast, nonlinear microbial models assume that decay rates depend on both substrate (soil C) and catalyst pool sizes (e.g., microbial biomass or enzyme pools). Soil carbon dynamics in nonlinear microbial models are often simulated as

\[
\frac{dC}{dt} = l - C \cdot f(C, E),
\]

where \(f(C, E)\) is a nonlinear function of substrate \((C)\) and catalyst (here soil enzyme, \(E\)) pool size. Various mathematical functions have been used for \(f\) [Schimel and Weintraub, 2003; Wutzler and Reichstein, 2008], and they can broadly be classified into three types: regular Michaelis-Menten, reverse Michaelis-Menten, or a combination of regular and reverse types. For simplicity, here we describe the regular and reverse types of Michaelis-Menten functions. Both of these models are rooted in kinetic rates observed in laboratory isolates [Monod, 1949].

The regular Michaelis-Menten function, as used in several recent publications [Allison et al., 2010; German et al., 2012; Wang et al., 2013, 2015a, 2015b; Wieder et al., 2013, 2014b], assumes that decay rate saturates as substrate concentration rises:

\[
\frac{dC}{dt} = l - \frac{V_{max} \cdot E \cdot C}{K_m + C} \quad \text{or} \quad f = \frac{V_{max} \cdot E}{K_m + C},
\]

where \(K_m\) is the half-saturation constant and \(V_{max}\) is the maximum reaction velocity per unit enzyme. \(K_m\) and \(V_{max}\) may depend on environmental conditions, and \(E\) is sometimes modeled as a function of soil microbial pool size. Rates become first-order on soil carbon when soil carbon concentration is very low, i.e., \(C << K_m\), and the catalyst pool size does not change in time. The regular Michaelis-Menten models assume that substrate availability is the rate limiting factor in decomposition. In some regular Michaelis-Menten models, equilibrium substrate pools are largely independent of input rates and dependent on microbial parameters such as \(K_m\) \(V_{max}\) microbial growth efficiency, and microbial turnover [German et al., 2012; Wang et al., 2014], although this finding may not be broadly applicable [see Allison et al., 2010].

Alternatively, reverse Michaelis-Menten kinetics assume that the soil C decay rate varies nonlinearly with enzyme pool size [Drake et al., 2013; Schimel and Weintraub, 2003; Sulman et al., 2014]. The kinetics are known as “reverse” because reaction rates saturate as enzyme concentrations increase rather than saturating as substrate concentrations increase as in the regular Michaelis-Menten expression. The reverse reaction is described as follows:

\[
\frac{dC}{dt} = l - \frac{V_{max} \cdot E \cdot C}{K_e + E} \quad \text{or} \quad f = \frac{V_{max} \cdot E}{K_e + E},
\]

where \(K_e\) is the half-saturation constant for enzyme binding. Rates become first order on soil C when enzyme concentration is very high, i.e., \(E >> K_e\) and \(f \approx V_{max}\). The assumption underlying reverse Michaelis-Menten
kinetics is that there are limited sites on which microbes (and their enzymes) can bind to substrates in the soil [Schimel and Weintraub, 2003]. Systems based on reverse kinetics may be more stable [Wang et al., 2015b] because increases in enzymes or microbial biomass do not necessarily increase decay rates and may ultimately approximate first-order models [Sihi et al., 2015]. Thus, forward and reverse Michaelis-Menten kinetics can provide equal estimates of SOM or litter decomposition [Moorhead and Sinsabaugh, 2006] because they represent the same fundamental process (competition among substrates vs. enzymes, respectively), but in the real world, both substrates and enzymes are variable.

Recent studies show that microbial-explicit and first-order models differ in their responses to environmental perturbations with potentially significant implications for C cycle projections under climate change. Nonlinear models simulated the observed response of soil CO2 respiration to drying-rewetting cycles better than first-order models which underestimated soil respiration rates [Lawrence et al., 2009]; similar models have been verified in field respiration responses to moisture pulses [Zhang et al., 2014]. Microbial-explicit and first-order models exhibit dramatically different soil C responses to simulated warming and elevated CO2, with microbial-explicit models exhibiting a high sensitivity to assumptions made regarding microbial physiological responses to perturbations [Allison et al., 2010; Sulman et al., 2014; Wieder et al., 2013]. Thus, a mounting body of work demonstrates that microbial-explicit models match observations at least as well as first-order models, if not better, but make different projections about the fate of that C under environmental change [Hararuk et al., 2015; He et al., 2014; Sulman et al., 2014; Wieder et al., 2014b, 2015]. While it is becoming clear that microbial-explicit and first-order models differ in their predictions, it is not clear that this is the result of higher degrees of model freedom or represents an actual gain from higher process resolution.

2.2. Challenges With Microbial-Explicit Models

As compared with first-order linear models, nonlinear microbial models may be preferable for simulating responses to perturbations. Yet, applying nonlinear soil microbial models at the global scale also presents significant challenges that should be resolved before they are incorporated into ESMs.

One challenge is that soil C models based on Michaelis-Menten kinetics can exhibit unrealistic temporal oscillations in response to small perturbations [Li et al., 2014; Sihi et al., 2015; Wang et al., 2014]. This may result from oversimplifying the spatial structure of soil and reflect scaling issues that arise from treating soil carbon and microbial pools as spatially and chemically homogenous. In reality, soil biota and substrates interact at micro scales; thus, simulated oscillations may weaken or disappear as micro-scale dynamics and parameter heterogeneity are integrated across the soil profile. Alternatively, oscillations could also be an artifact of simplifying the soil microbial community itself. Heterogeneity in the physiological traits present in the soil microbial community results in a distribution of parameters like microbial growth and extracellular enzyme production which in turn could mathematically damp oscillations seen in the system run with a single mean parameter value. Finally, top-down control in soil food webs [Crowther et al., 2015] could also regulate pulses of microbial growth and dampen the oscillatory responses.

A second challenge is estimating and interpreting model parameters. In contrast to many microbial parameters, rate constants in first-order models have been estimated across broad spatial and temporal scales [Adair et al., 2008; Zhang et al., 2008]. Although using rate constants may be inappropriate for projecting soil biogeochemical response to global change scenarios [Bradford and Fierer, 2012; Sierra et al., 2012b; Todd-Brown et al., 2014], decay “constants” in first-order models could be adjusted as a function of environmental variables. To date, global microbial-explicit models apply parameter values for microbial kinetics that are largely based on laboratory studies [Sulman et al., 2014; Wieder et al., 2013, 2015]. While there are possible generalizations that could be made from theoretical ecology [Enquist et al., 2015], significant challenges remain in understanding how parameter values related to microbial physiology vary across different environmental and edaphic gradients. Some insight may be drawn from analogous approaches that are used aboveground, where leaf-level measurements are routinely applied in the photosynthesis parameterization of ESMs [Bonan et al., 2011, 2012]. For example, statistical distributions of some microbial parameters, such as $V_{\text{max}}$ and the turnover rates of exoenzymes or microbial biomass pools, can be calculated across ecosystems and soil types to facilitate model predictions at the global scale [German et al., 2012; Sinsabaugh et al., 2008]. Because soil is a highly heterogeneous environment, microbial parameters measured at specific sites over short time periods may be difficult to interpret at ecosystem to global scales. For instance, elevated microbial biomass in rhizosphere soil can generate hot spots for decomposition in microbial models [Sulman et al., 2014], and
pulses in C inputs can cause "hot moments" of elevated decomposition rates [Kuzyakov and Blagodatskaya, 2015]. Therefore, knowledge about root distributions and timing of C inputs may be required to scale up microbial parameters.

A third challenge is to analyze the consequences of interactions between C and nutrients in microbial models, particularly N and phosphorus (P). Early theoretical models included N [Schimel and Weintraub, 2003], but nutrient dynamics have not yet been incorporated into nonlinear microbial models at the global scale. At local to ecosystem scales, novel agent-based models have revealed the potential for community-level processes to alter nutrient dynamics and rates of litter decay [Kaiser et al., 2014; Moorhead and Sinsabaugh, 2006]. Using a microbial-explicit ecosystem model with flexible microbial C:N stoichiometry, Sistla and co-authors [2014] demonstrate how warming affects plant-soil feedbacks and ecosystem C storage via changes in SOM decomposition, N mineralization rates, and plant growth. In this study, assumptions about the ability of microbial communities to acclimate to warming temperatures strongly influenced ecosystem C balance. This work illustrates how greater mechanistic representation in models may help leverage empirical data to improve projections of ecosystem response to environmental change.

A fourth challenge is to quantify and parameterize potential microbial-mineral interactions. Microbial accessibility to substrates in mineral soils may determine the magnitude of microbial influence on SOM decomposition [Schimel and Schaeffer, 2012]. The C that is associated with mineral fractions and that is not available for decomposition may be irrelevant for short-term soil respiration measurements and may be ignored if CO₂ fluxes are the primary concern [Davidson et al., 2014]. In contrast, if long-term (decadal or longer) changes in C stocks are the main focus of model simulations, then microbial-mineral interactions and their responses to global change processes may be extremely important in determining future C stocks. Inputs of microbial residues and their potential stabilization on mineral surfaces and within soil aggregates determine how microbes may affect long-term SOM formation [Cotrufo et al., 2013; Dungait et al., 2012; Grandy and Neff, 2008]. Addressing the microbial contribution to these SOM decomposition and stabilization processes in heterogeneous soil environments and across edaphic gradients can both offer insights into our basic understanding of soil biogeochemical cycles and improve the accuracy of soil biogeochemical projections in a changing world, particularly for predictions on decadal timescales.

A final challenge is the need to subject nonlinear microbial models to rigorous benchmarking and model intercomparison analyses. Microbial models should be carefully validated against observational data before they are used to replace first-order linear models in Earth system simulations [Wang et al., 2014; Wieder et al., 2013]. Results from most litter decomposition and soil organic C (SOC) mineralization studies are consistent with first-order models [Adair et al., 2008; Bonan et al., 2013; Zhang et al., 2008], but similar tests for microbial models are scarce [Moorhead and Sinsabaugh, 2006; Wieder et al., 2014b, 2015]. Microbial-related processes may drive changes in SOC content over long periods of time that cannot be resolved in short-term mineralization measurements. Long-term soil manipulations [e.g., Lajtha et al., 2014; Metillo et al., 2011], accompanied by detailed measurements of microbial function, may be necessary to distinguish between alternative models of soil C response to changes in litter inputs or climate. Both first-order and microbial-explicit models could be improved and compared using such long-term data sets and statistical approaches such as Bayesian data assimilation [Hararuk and Luo, 2014; Tang and Zhuang, 2009; Wang et al., 2009; Xu et al., 2006]. The results of these analyses will be essential for improving soil C components of ESMs.

3. Roadmap

We outline the following roadmap to facilitate critical model development, parameterization, and evaluation activities needed to build mechanistic, microbial-explicit models that can both advance our fundamental understanding of soil biogeochemical dynamics and more realistically project soil C response to environmental change at Earth system scales. We stress that these activities represent a community-wide challenge that will require interdisciplinary collaborations and contributions from empirical and modeling communities. As such, given adequate support and community-wide engagement, we expect a 5–10 year timeframe to realize these aims—although we expect significant progress can be made on shorter timescales. This roadmap broadly identifies three objectives aimed at expediting the development, application, and evaluation of mechanistic models in ESMs with a focus on microbial-explicit approaches.
Table 1. Structural and Operational Characteristics of 13 Recently Published Microbial Models

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Number of Pools (and ODEs)</th>
<th>Biogeochemical Cycles</th>
<th>Temporal and Spatial Scales</th>
<th>Environmental Dependencies</th>
<th>Vertically Resolved?</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>C only</td>
<td>Hourly; cm⁻³ soil</td>
<td>T</td>
<td>No</td>
<td>Allison et al. [2010]</td>
</tr>
<tr>
<td>2</td>
<td>–</td>
<td>C only</td>
<td>Hourly; cm⁻³ soil</td>
<td>T, moisture, soil texture</td>
<td>No</td>
<td>Davidson et al. [2012]</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>C only</td>
<td>Hourly; cm⁻³ soil</td>
<td>T</td>
<td>No</td>
<td>German et al. [2012]</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>C only</td>
<td>Hourly; cm⁻³ soil</td>
<td>T, moisture, soil texture</td>
<td>Yes</td>
<td>He et al. [2014]</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>C, N</td>
<td>3 h; cm⁻³ litter</td>
<td>Litter chemistry</td>
<td>No</td>
<td>Kaiser et al. [2014]</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>C only</td>
<td>Hourly; m⁻³ soil</td>
<td>T, moisture, soil texture</td>
<td>No</td>
<td>Manzoni and Katul [2014]</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>C, N</td>
<td>Daily; mg⁻¹ soil</td>
<td>–</td>
<td>No</td>
<td>Moorhead and Sinsabaugh [2000]</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>C, N</td>
<td>Daily; g⁻¹ soil</td>
<td>T, litter chemistry</td>
<td>No</td>
<td>Schimel and Weintraub [2003]</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>C only</td>
<td>Hourly, daily, and annual; whole soil column</td>
<td>T, soil texture</td>
<td>No</td>
<td>Tang and Riley [2014]</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>C only</td>
<td>Hourly; mg⁻¹ soil</td>
<td>–</td>
<td>No</td>
<td>Wang et al. [2015a, 2015b]</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>C only</td>
<td>Hourly; cm⁻³ soil</td>
<td>T, soil texture, litter chemistry</td>
<td>No</td>
<td>Wieder et al. [2014b]</td>
</tr>
<tr>
<td>12</td>
<td>16</td>
<td>C, N</td>
<td>Daily; g⁻² soil</td>
<td>T, litter and SOM chemistry</td>
<td>No</td>
<td>Sistla et al. [2014]</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>C, N</td>
<td>Daily; g⁻¹ C</td>
<td>Litter chemistry</td>
<td>No</td>
<td>Moorhead and Sinsabaugh [2006]</td>
</tr>
</tbody>
</table>

*T = temperature, C = carbon, N = nitrogen, SOM = soil organic matter.

3.1. Guidelines for Common Model Parameters and Output

Model development and benchmarking (discussed in section 3.2) require some correspondence between empirical data and the parameters or outputs of models. Additionally, model intercomparison projects can be expedited if models simulate key state variables that correspond to observations. We surveyed 13 currently published microbial models that explicitly simulate microbial activity in SOC decomposition (Table 1). This survey was not intended to exhaustively describe different microbial-explicit models but instead broadly characterize the diversity of approaches that have been used in site-level or ecosystem simulations. The surveyed models are mostly formulated as systems of ordinary differential equations (with one exception), are run at hourly time steps, and represent a small spatial scale (cm³), with state variables representing different C pools; thus, the number of equations equals the number of state variables. Structurally, these microbial models usually consist of two or more SOC pools and separate pools for enzymes, microbial biomass, dissolved organic C (DOC), and sometimes a soluble C pool (Figure 1 and Table 2) (Allison et al., 2010; Kaiser et al., 2014; He et al., 2014). In contrast to microbial-implicit model formulations, the majority of the microbial models only consider C, with fewer models (models 5, 7, 8, 12, and 13) explicitly considering the effects of N on microbial and SOM dynamics; this is clearly an area for future development. We note that there is substantial overlap in the basic processes represented across models, which include microbial enzyme production, enzymatic SOC degradation, microbial assimilation of DOC substrate; Assimilation: microbial assimilation of DOC substrate; ENZ production: production of enzymes by microorganisms; ENZ deactivation: enzyme turnover and loss; Death: microbial death (turnover) rate.
Table 2. Summary of Common Features of 13 Recently Published Microbial Models

<table>
<thead>
<tr>
<th>Model Features</th>
<th>Description</th>
<th>Environmental Dependency/Regulator</th>
<th>Model Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant litter</td>
<td>Inputs of plant litter</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SOC enzymatic degradation</strong></td>
<td>Enzymatic decay of polymer</td>
<td>T, moisture, soil texture, litter chemistry</td>
<td></td>
</tr>
<tr>
<td><strong>DOC microbial assimilation</strong></td>
<td>Assimulation of DOC by microbes; magnitude controlled by T-dependent CUE, mineral adsorption may be involved</td>
<td>T, DOC, soil texture</td>
<td></td>
</tr>
<tr>
<td><strong>Enzyme production</strong></td>
<td>Carbon use efficiency (the fraction of C assimilated into microbial biomass)</td>
<td>T, substrate quality, microbial community composition</td>
<td></td>
</tr>
<tr>
<td>(or MGE, maintenance/growth)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Enzyme deactivation</strong></td>
<td>Mineral adsorption may be involved</td>
<td>T, ENZ, soil texture</td>
<td></td>
</tr>
<tr>
<td>(loss/turnover rate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Microbial death (turnover rate)</strong></td>
<td>Microbial biomass turnover</td>
<td>MIC</td>
<td>1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13</td>
</tr>
<tr>
<td><strong>Mineral surface</strong></td>
<td>Mineral surface binding of DOC and enzymes</td>
<td>T, soil texture</td>
<td>8, 9, 10, 11</td>
</tr>
<tr>
<td>adsorption/desorption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of DOC and ENZ</td>
<td>(to stabilize and deactivate, respectively)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diffusion of DOC and ENZ</strong></td>
<td>Diffusion of DOC and ENZ in soil column</td>
<td>T, moisture, soil texture</td>
<td>2, 4, 8, 9, 10</td>
</tr>
<tr>
<td><strong>Microbial functional groups</strong></td>
<td>Explicitly representing different metabolic activity of microbial functional groups, or flexible microbial stoichiometry</td>
<td>T, soil texture</td>
<td>5, 11, 12, 13</td>
</tr>
</tbody>
</table>

and the turnover of enzyme and microbial biomass (Figure 1 and Table 2). However, other mechanisms such as mineral adsorption/desorption, substrate/enzyme diffusion, and microbial dormancy are less commonly considered but will likely be the subject of future studies.

Structural differences among microbial-explicit models reflect uncertainties in model form and the underlying processes governing SOM decomposition and stabilization. Comparing and understanding the different theoretical and structural underpinnings of alternative microbial models presents a significant undertaking but one that can improve the representation of soil biogeochemical processes across scales. Although inclusion of microbial-explicit configurations in ESMs will likely increase the inter-model range of soil C projections in model intercomparison projects, we contend that this is a desirable outcome of improving process representation in terrestrial biogeochemical models [Knuth and Sedlacek, 2013]. An expanded range of inter-model variability may in fact represent a more accurate assessment of uncertainty in future carbon cycle projections. Specifically, models including microbial processes could be used in conjunction with first-order models to more comprehensively represent uncertainties in the terrestrial carbon cycle.

Microbial-explicit models represent an array of responses to environmental drivers including diffusion limitation, temperature, and nutrient availability. The representation of substrate and enzyme diffusion is typically represented as a function of volumetric soil moisture using empirical functions to account for solute diffusion rates (models 2 and 4). The temperature dependence of microbial kinetics is commonly represented by $Q_{10}$ or Arrhenius-type functions that determine maximum rates of substrate uptake and decomposition when substrate is not limiting, but lower rates are simulated when other environmental variables, such as water availability, limit substrate availability. Characterization of soil C decomposition temperature sensitivity is either an intrinsic or apparent response. For example, the Dual Arrhenius and Michaelis-Menten model (2) uses an Arrhenius function for the intrinsic temperature response of the enzyme, but because

10.1002/2015GB005188
substrate supply also affects the extent to which the Arrhenius function is expressed, the apparent temperature sensitivity becomes an emergent response. In addition to the temperature sensitivity of microbial kinetics, some models also consider potential temperature effects on MGE or microbial growth yields (models 3, 4, 5, 6, 8, and 9). Enzyme production, enzyme deactivation rates, and microbial death rates are commonly simulated across most models as a fixed proportion of microbial biomass, as a function of temperature, or as microbial stoichiometry (model 12). Most microbial models only consider temperature effects on microbial activity, growth, and turnover, with fewer representing soil moisture constraints (models 2 and 4). This gap clearly must be addressed, although it remains unclear whether soil moisture constrains microbial activity primarily through direct (via desiccation stress) or indirect (via its impact on diffusion of substrates or enzymes) mechanisms under different environmental conditions (e.g., for nitrification) [Stark and Firestone, 1995]. Future research should aim to develop relationships that describe the environmental controls on rates of microbial activity.

A minimum requirement for microbial models in the context of Earth system modeling is to accurately simulate microbial functional responses to environmental perturbations. Thus, experiments should be used to test model structures by measuring process responses to environmental drivers. A major challenge for experiments and models is to tease apart multiple processes that control decomposition rates, such as substrate supply and microbial activity, and the magnitude of their responses to environmental perturbations [Hopkins et al.], 2014]. Also, measurements are needed to parameterize key microbial functions on which models are based, such as maximum microbial uptake, C use efficiency, and turnover. While many experiments have examined the response of microbial respiration to temperature in laboratory microcosms [German et al., 2012; Stone et al., 2012], fewer include multiple measurements of microbial functions that drive models, such as microbial turnover and MGE [Frey et al., 2013; Hagerty et al., 2014]. Importantly, we emphasize that model parameters should be linked to measurable quantities.

Our overview of these models highlights a number of directions to explore, including the following: (1) heterogeneity in space and time caused by properties of the external environment such as edaphic conditions, soil moisture, and vegetation; (2) the diversity of microbial life history strategies, functional traits, and community dynamics; and (3) effects of coupled biogeochemical cycles. Given that a wide range of differences in parameters and assumed functional forms clearly exists between current microbial models and that the models themselves adopt relatively similar structure, it is an opportune time to conduct a range of formal model intercomparisons to assess the spatial and temporal scales over which differences in model structure and parameterization generate alternative predictions.

### 3.2. Development of Model-Data Integration and Benchmarking Frameworks

Microbial model development, parameterization, and evaluation should be guided by and informed with data that, in turn, can generate new hypotheses for experimental testing (Figure 2). Such tight integration between models and data will accelerate model development, guide experimental design, and, ultimately, improve our understanding of microbial regulation of soil biogeochemical processes. Model benchmarking is the process of assessing model performance against a set of standard data sets [Luo et al., 2012], while model-data integration is the incorporation of data into the process of model development and refinement (through parameter estimation and selection of structural forms) (Figure 2). Both techniques are useful to evaluate the correspondence between models and data and can also use the same data sets. Care must be taken to avoid over-fitting models to data, but this can be minimized by (1) the adoption of cross-validation methods, (2) the use of information theoretic metrics such as Akaike information criterion, (3) the use of multiple data sets on different system processes, (4) the collection of new independent data, and (5) controlling for spatial and temporal autocorrelation in the empirical data. Model benchmarking and model-data integration activities present two underutilized tools that need to be developed in the context of microbial-explicit soil biogeochemical models, both of which depend on the development and curation of appropriate data sets.

Initially, globally gridded data sets of soil C and microbial biomass estimates provide tractable benchmarks of microbial-explicit models at a scale relevant to global climate change projections. To date, global microbial models have been compared to a single globally gridded data set of total soil organic carbon [FAO et al., 2012; Wieder et al., 2014c]. While data sets like the Harmonized World Soil Database (HWSD) provide an initial benchmark for global C projections [Carvalhais et al., 2014; Todd-Brown et al., 2013], these static
estimates provide limited insight into the fate of soil C pools under environmental change, illustrating the need to draw on additional relevant data sets (Table 3). For example, T3 Hararuk and others [2015] used a gridded data set of microbial biomass [Xu et al., 2013] to benchmark their model, and other similar data products are available [Serna-Chavez et al., 2013]. Again, the utility of such metrics is debatable, given that methods for measuring microbial biomass provide little insight into their potential activity under various environmental conditions.

Observations spanning broad ecoclimatological gradients may provide an additional benchmarking tool. For example, data from the Long-term Intersite Decomposition Experiment Team (LIDET) [Adair et al., 2008; Harmon, 2013; Harmon et al., 2009; Parton et al., 2007] have been used to evaluate litter mass loss dynamics across continental-scale climate gradients in first-order and microbial-explicit models [Bonan et al., 2013; Xu et al., 2013; Bond-Lamberty and Thomson, 2010] and global analyses of soil microbial community composition [Fierer et al., 2012] (Tables 3 and 4).

Table 3. Steady State Observations of Belowground Pools, Fluxes, and Edaphic Characteristics that Can Be Used to Parameterize and Initialize Models.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Key Variables</th>
<th>Spatial, Time and Ecosystem Scales</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil stocks and properties</td>
<td>Harmonized World Soil Database (HWSDB)</td>
<td>Soil C, physical properties, top soil, and subsoil (0–1 m)</td>
<td>Global, 30 arc sec (0.008333°) resolution</td>
<td>FAO et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Northern Circumpolar Soil C Database (NCSCD)</td>
<td>Soil C to 3 m depth</td>
<td>Northern permafrost region</td>
<td>Hugelius et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>ISRIC-WISE and SoilsGrid1km</td>
<td>Soil C, physical properties at six depths</td>
<td>Global w/ 1 km resolution</td>
<td>Hengl et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>Isotopic measurements</td>
<td>Isotopic measurements</td>
<td>Lab incubations, field sites,</td>
<td>Amundson et al. (2003),</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and global estimates</td>
<td>Behrens et al. (2008),</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Torn et al. (2002)</td>
</tr>
<tr>
<td>Microbial biomass</td>
<td>Soil microbial abundance Biome extrapolated</td>
<td>MIC, soil C, physical properties</td>
<td>Global, 0.5° resolution</td>
<td>Serna-Chavez et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Biome N, and P</td>
<td>Microbial and soil C, N and P to 1 m depth</td>
<td>Global, 3422 data points from 14 biomes</td>
<td>Xu et al. (2013)</td>
</tr>
<tr>
<td>Litter decay</td>
<td>Long-term observations (LIDET, CIDET)</td>
<td>Litter quality, climate, litter mass loss, litter N</td>
<td>Multisite using standard substrates;</td>
<td>Harmon (2013),</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>up to 12 years</td>
<td>Moore et al. (2008)</td>
</tr>
<tr>
<td>Soil resp.</td>
<td>Database of soil resp. measurements (SRDB)</td>
<td>Soil resp., T, Q10, biome</td>
<td>&gt;800 studies and &gt;3300 records</td>
<td>Bond-Lamberty and Thomson</td>
</tr>
</tbody>
</table>

*Abbreviations: C = carbon, N = nitrogen, resp. = respiration, MGE = microbial growth efficiency (or carbon use efficiency), MIC = microbial biomass carbon.*
### Table 4. Belowground Response to Environmental Change Manipulations that Can Be Used to Parameterize and Evaluate Model Structure

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Key Variables</th>
<th>Spatial, Time and Ecosystem Scales</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil warming</td>
<td>T sensitivity of resp. to warming C-N interactions with soil warming</td>
<td>C, N, resp., fine-root resp., and biomass V&lt;sub&gt;max&lt;/sub&gt;, K&lt;sub&gt;m&lt;/sub&gt;, Q&lt;sub&gt;10&lt;/sub&gt;</td>
<td>Field site (deciduous forest); 10+ yrs; 7+ yrs Lab study with cross-latitude samples 90 days lab study w/ soils along climate gradient 1 week lab study w/ mineral and organic soils</td>
<td>Melillo et al. [2002, 2011]; Butler et al. [2012]; German et al. [2012]; Karhu et al. [2014]; Hagerty et al. [2014]</td>
</tr>
<tr>
<td>Increased C inputs w/ and w/o warming</td>
<td>Microbial response to increased C inputs and nutrients Response of microbial growth rates in soil to elevated atmospheric CO&lt;sub&gt;2&lt;/sub&gt; Microbial response to increased C inputs and warming</td>
<td>Soil C, MIC, nutrients, labeled C input, and resp. Microbial growth rate, N fertilization, MIC, soil C, and resp. MIC, composition, CUE, and resp.</td>
<td>Lab study with soils from three CO&lt;sub&gt;2&lt;/sub&gt; enrichment sites Lab study with soils from decade-long Aspen FACE experiment; 6–31 days</td>
<td>Fontaine et al. [2004]; Blagodatskaya et al. [2010]; Hopkins et al. [2014]</td>
</tr>
<tr>
<td>Soil rewetting</td>
<td>Soil C and N response to soil rewetting frequency</td>
<td>Soil C and N stocks, dissolved organic and microbial C and N, resp.</td>
<td>Lab study with soils from California; 16 weeks</td>
<td>Miller et al. [2005]</td>
</tr>
<tr>
<td>Agricultural experiments</td>
<td>C inputs (quality/quantity) to soils, N amendments, tillage, crop rotations, etc.</td>
<td>Soil C and N stocks, soil resp., physical properties</td>
<td>Field sites across wide climate and soil texture gradients; years to multidecadal Data from field and lab studies across multiple ecosystems</td>
<td>Jenkinson and Rayner [1977]; Paul et al. [1996]; Paustian et al. [1992]; Janssens et al. [2010]; Liu and Greaver [2010], Lu et al. [2011]</td>
</tr>
<tr>
<td>N enrichment</td>
<td>Meta-analysis to ecosystem effects of N addition</td>
<td>Soil C, resp., litter input, MBC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Abbreviations: C = carbon, N = nitrogen, resp. = respiration, MGE = microbial growth efficiency (or carbon use efficiency), MIC = microbial biomass carbon.*

Just as isotopic information was critical for informing early SOM models [Jenkinson and Rayner, 1977; Van Veen and Paul, 1981; Tom et al., 1997], application of isotopic data should rapidly advance microbial-explicit models too. Numerous possibilities exist for using isotopic data to estimate biogeochemical fluxes and C stabilization [Bohn et al., 2012; Behrens et al., 2008; Kayler et al., 2011]. Isotope natural abundance can be used to validate process representations in models across space and time, such as using spatial extrapolations of 15N distributions [Anundson et al., 2003, as in Houlton et al., 2015] and 13C and 14C [Tom et al., 2002]. In particular, 14C data sets are a critical benchmark for correctly simulating depth distributions of soil C in models [Koven et al., 2013]. Isotope labeling, at scales from microbial metabolism [Dijkstra et al., 2011] to ecosystems [Kramer et al., 2010], also provides a powerful tool to trace pathways of C and other elements through the soil system. Isotopic measurements in global change experiments provide independent constraints for model representations of emergent feedbacks, such as those resulting from elevated CO<sub>2</sub>, warming, and N deposition [Hopkins et al., 2014; Hopkins et al., 2012; Sierra et al., 2012a, 2012b].

Experimental manipulations documenting soil biogeochemical responses to changes in temperature, inputs, nutrient availability, and moisture provide additional benchmarking data sets (Table 4). Although some data from long-term manipulations are available [e.g., Lajtha et al., 2014; Melillo et al., 2011], most manipulations typically only last for months to years, so caution is warranted when comparing to model projections that play out over decades to centuries in ESMs. Nonetheless, such data sets are likely to be important for testing microbial-explicit versus implicit models that diverge in predicted responses to perturbations. A wealth of literature and long-term observations from the agricultural sciences have been used to evaluate first-order models and can now provide valuable data sets to evaluate the long-term behavior of microbial-explicit models [Jenkinson and Rayner, 1977; Paul et al., 1996; Paustian et al., 1992]. These data include soil C and N response to changes in the chemical quality, quantity, and timing of litter inputs into soil systems across wide climatic and edaphic gradients (Table 3). Data from these types of experimental manipulations may present particularly...
valuable information for parameterizing and benchmarking models that are intended for applicability at wide spatial scales and for future predictions. Benchmarking models with observational data is particularly useful for assessing and refining models that operate on large spatial scales, such as the land surface components of ESMs, where there is greater emphasis on predictive power, and assessing the performance of multiple complex multicomponent models. For example, Carvalhais and others [2014] illustrate important biases of terrestrial C turnover times simulated within and among models represented in the CMIP5 archive [also see Koven et al., 2015]. Similarly, Todd-Brown and others [2013] evaluate present-day soil C pools simulated by models in the CMIP5 archive with observations from the Harmonized World Soil Database [FAO et al., 2012]. These types of analyses can be formally scored in a benchmarking system like the International Land Model Benchmarking Project [Luo et al., 2012]. Analogous approaches can and should be applied to microbial-explicit soil models. To develop benchmarking and model-data integration frameworks for microbial-explicit soil models we suggest developing and curating standardized data sets that include several key microbial variables (Table 3).

Model-data integration provides techniques to modify model parameter values, adjust initial values of model states, and generate uncertainty estimates based on goodness of fit between model predictions and data [Luo et al., 2011]. Model-data integration can be used to (1) reduce the projected uncertainty of models, (2) gain biological insights from the inferred parameter values, (3) quantify degrees of support for different model formulations, (4) infer unobserved properties and mechanisms, and (5) seek insights into model improvements by investigating discrepancies between model predictions and data [Hararuk et al., 2014; Ise and Moorcroft, 2006; Luo et al., 2011; Smith et al., 2013; Tang and Zhuang, 2009; Xu et al., 2006]. While these techniques are relatively common in other fields, there are relatively few applications to global biogeochemical models [see Rayner et al., 2005; Wang et al., 2009]. For example, Hararuk and others [2015] recently used global soil C databases to constrain parameter estimates in first-order and microbial-explicit models and found that the calibrated microbial models showed greater soil C losses than first-order models. We feel this is likely due to the lack of global data sets prior to date and a lack of recognition in the community of the value of formal model-data integration algorithms, both of which are expected to change in the near future.

3.3. Application of Scaling Tools to Bridge From Microbial to Global Scales

Appropriately scaling microbial dynamics from the micro scale, where decomposition and stabilization of SOM occur, to macro scales considered by ESMs is one of the greatest challenges for soil microbial modeling [Davidson et al., 2014]. There are roughly three key scales involved in connecting the nonlinear microbial rates that govern decomposition to ecosystem-scale rates of soil respiration. These include (1) the micro scale at which SOM decomposition and stabilization occur ($10^{-6}$ to $10^{-3}$ m), (2) the intermediate scale at which these processes are commonly measured ($10^{-2}$ to 1 m), and, finally, (3) macro scales that are relevant to soil management and climate change ($10^{0}$ to $10^{4}$ m) [Hinckley et al., 2014]. First-order soil C models were developed to capture simple C balance behaviors at the intermediate scale, focusing on litter decomposition or field experiments [Jenkinson et al., 1987; Olson, 1963; Parton et al., 1987]. However, as our fundamental understanding of soil processes has grown [Schmidt et al., 2011], a more mechanistic representation of microbial-explicit soil decomposition that offers a promising avenue of research to incorporate new microscale insights and measurements into macro-scale representations within ESMs.

Integrating new insights across scales presents significant challenges, although soil science is not alone in this quest, which is shared by other areas of modeling. For example, in physics and chemistry, there is sometimes a basic understanding of how equations for macroscopic phenomena (such as equations to model chemical diffusion) emerge from the underlying microscopic behaviors (in that case, the random movement of particles). We typically lack such formal derivations for ecological processes, although there are approaches that can be used to identify appropriate scaling tools in their absence. For example, significant efforts have been made to conduct formal scaling in model vegetation dynamics. The Ecosystem Demography [Moorcroft et al., 2001] and Perfect-Plasticity Approximation [Strigul et al., 2008] models apply individual-based forest gap dynamics to develop size and age-class properties of larger ecosystem-scale models that ultimately influence forest structure, productivity, and C storage [Fisher et al., 2015]. Applying similar approaches to deriving equations for microbial dynamics in soils could be a promising research direction.

Reactive transport models may provide an alternative framework to scale microbial processes across heterogeneous soil environments. The contaminant transport community has used reaction-diffusion models to simulate hydrologically driven contaminant diffusion in the subsurface soil, with particular emphasis on
the flow of dissolved compounds in the porous soil matrix and adsorption of compounds onto mineral surfaces [Xu et al., 2011]. However, recent efforts have introduced an explicit biological term to simulate the role of microbes in contaminant breakdown, and there has been success linking the site-specific parameters in large scale models to the prior microbial parameters [Scheibe et al., 2009]. Currently, similar models are being explored for more general C decomposition [Resat et al., 2012; Riley et al., 2014]; however, finding appropriate data sets to validate the fine resolution of these models presents significant challenges, in addition to scaling issues discussed, for incorporating these models at ecosystem and Earth system scales.

Explicitly scaling between micro and intermediate scales would link basic microbial metabolism and trade-offs (itself an active area of research) across microsite variations in substrate availability and environmental conditions to describe measurable soil C dynamics. The basic micro-scale mechanisms represented in microbial models are well supported by decades of biochemical and physiological studies [Monod, 1949; Sinsabaugh et al., 2014, 2015], but it remains challenging to predict how diverse enzymes and microbes interact with heterogeneous soil environments to generate emergent processes such as litter decomposition at the macro scale. Moreover, factors regulating substrate supply may be a dominant control at larger scales, especially in mineral soils [Schimel and Schaeffer, 2012], simplifying the representation of microbial physiology. First-order models have assumed micro-scale processes can be ignored; however, microbial models must contend with them because direct extrapolation of micro-scale dynamics to macro scales has been demonstrated to lead to unrealistic behaviors in soil properties [Wang et al., 2014], a problem shared by those modeling the dynamics of marine plankton populations [Franks, 2009]. To date, none of the global-scale microbial-explicit decomposition models directly account for this scaling in detail [Hararuk et al., 2015; Sulman et al., 2014; Wieder et al., 2013, 2015]. Making these models more realistic will likely require a nested approach whereby micro-scale mechanisms and heterogeneity (e.g., enzyme-substrate interactions) are represented explicitly in micro-scale models [Allison, 2012, 2014; Kaiser et al., 2014]. This could be done with 3-D models that represent microbial activity and substrate availability that are coupled into ecosystem-scale models.

Linking intermediate-scale measurements to societally and climatically relevant macro-scale projections is also critical to improve the accuracy of soil C models. Toward this end, empirical and statistical approaches (such as machine learning techniques) could provide valuable insights into emergent scaling relationships, such as between microscopic and the macroscopic scales, and how they vary in time and space. Advances in the availability of empirical data and modeling methodologies (described in the previous sections) will assist progress in this area. For example, improvements in scaling assumptions have been seen in the modeling of canopy photosynthesis at leaf, canopy, landscape, and global scales. In ESMs, leaf traits are now commonly scaled within plant canopies (e.g., sun-lit vs. shaded leaves) and across plant functional types (e.g., grassland vs. forest ecosystems), while still employing micro-scale enzyme kinetics [Farquhar et al., 1980] to simulate aboveground C balance and project its response to environmental change [Bonan et al., 2011, 2012, 2014]. Analogous developments are needed for modeling soil microbial dynamics across large spatial scales [Davidson et al., 2014]. Given the diversity of metabolic and physiological strategies used by microbes, significant simplifying assumptions will have to be made for scaling tools to be tractable in global models. In parallel with spatial scaling efforts, considerations of temporal dynamics are critical to linking short-term soil decomposition processes to decadal or longer projections in ESMs.

4. Conclusion

New technologies are expanding the scale of measurements and advancing the fundamental understanding of soil form and function. At micro scales, these include measurements of extracellular enzyme activities, genetic information on the composition and activity of soil microbial communities, and data on substrate-mineral interactions within heterogeneous soil matrices. Macro-scale technologies include satellite-derived soil moisture and geospatially extrapolated estimates of soil properties and C storage. New model structures and scaling approaches are needed to integrate these measurements and the insight they bring into models that are used to make C cycle-climate projections. We expect that within the next 5 years we will see the following: (1) better agreement on model parameters, output, and structural form (Tables 1 and 2); (2) a series of model perturbation experiments that use standardized forcings (similar to TRENDY simulations) [Sitch et al., 2015]; (3) synthesis and creation of formal benchmarking data sets (Table 3), which will lead to a (4) microbial soil model intercomparison project (MSMIP); and, finally, (5) studies scaling micro-scale microbial
mechanisms to macro-scale processes. Fully developing the theoretical understanding, observational data sets, and numerical tools to address the scaling issues outlined in section 3.3 is likely a decade or more away, but this approach offers promise in improving the representation of soil biogeochemical processes. These gains, however, will require sufficient intellectual engagement, (at least loosely) coordinated activities, and adequate financial support. We also note that these recommendations are not exclusive to microbial-explicit models and should complement a larger effort to improve the representation of soil biogeochemical processes in ecosystem and Earth system models. These activities should also broadly include soil scientists, ecologists, geoscientists, and mathematicians, as wider engagement across these scientific disciplines will more rapidly advance the field.

Acknowledgments
The National Center for Atmospheric Research is sponsored by the United States National Science Foundation and Department of Energy. This work was produced from a workshop that was financially supported by the National Science Foundation Research Coordination (RCN) grant DEB 0840964 and Department of Energy DE SC0008270. No data were used in producing this manuscript; material in the figures and table are properly cited and referred to in the reference list.

References


