RESEARCH REVIEW

21st-century biogeochemical modeling: Challenges for Century-based models and where do we go from here?

Danielle Berardi[1,2] | Edward Brzostek[1,3] | Elena Blanc-Betes[4] | Brian Davison[5] | Evan H. DeLucia[1,4] | Melannie D. Hartman[1,6] | Jeffrey Kent[1,2] | William J. Parton[1,6] | Debasish Saha[7] | Tara W. Hudiburg[1,2]

[1]Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Champaign-Urbana, Urbana, IL, USA

[2]Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow, ID, USA

[3]Department of Biology, West Virginia University, Morgantown, WV, USA

[4]Institute for Sustainability, Energy and Environment, University of Illinois at Champaign-Urbana, Urbana, IL, USA

[5]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, USA

[6]Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, USA

[7]Biosystems, Engineering, and Soil Science, University of Tennessee, Knoxville, TN, USA

Correspondence
Tara W. Hudiburg, Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow, ID, USA.
Email: thudiburg@uidaho.edu

Abstract

21st-century modeling of greenhouse gas (GHG) emissions from bioenergy crops is necessary to quantify the extent to which bioenergy production can mitigate climate change. For over 30 years, the Century-based biogeochemical models have provided the preeminent framework for belowground carbon and nitrogen cycling in ecosystem and earth system models. While monthly Century and the daily time-step version of Century (DayCent) have advanced our ability to predict the sustainability of bioenergy crop production, new advances in feedstock generation, and our empirical understanding of sources and sinks of GHGs in soils call for a re-visitation of DayCent's core model structures. Here, we evaluate current challenges with modeling soil carbon dynamics, trace gas fluxes, and drought and age-related impacts on bioenergy crop productivity. We propose coupling a microbial process-based soil organic carbon and nitrogen model with DayCent to improve soil carbon dynamics. We describe recent improvements to DayCent for simulating unique plant structural and physiological attributes of perennial bioenergy grasses. Finally, we propose a method for using machine learning to identify key parameters for simulating N_2O emissions. Our efforts are focused on meeting the needs for modeling bioenergy crops; however, many updates reviewed and suggested to DayCent will be broadly applicable to other systems.

KEYWORDS
bioenergy, biogeochemical modeling, drought, N_2O, plant age dynamics, soil

1 | INTRODUCTION

Biogeochemical models are used to simulate how abiotic and biotic variables interact through time and across space to determine rates of biogeochemical fluxes. They provide a platform for scientists to evaluate how current and potential changes in climate, land use, disturbance regimes, or vegetation will impact greenhouse gas (GHG) budgets, carbon sequestration and storage, and water quality. As we consider bioenergy crops as alternatives to fossil fuels,

it is extremely valuable to use biogeochemical models in tandem with field experiments to make projections that determine which crops will be the most effective and sustainable and in which locations (Whitaker et al., 2018). Challenges we face include the fact that bioenergy crops have unique physiological traits (e.g., dynamic allocation, carbohydrate and nitrogen storage, and tolerance for low-quality soils), particularly in genetically modified forms (e.g., Energycane, Petro sorghum) that are not currently represented in models. Moreover, N$_2$O emissions, drought-response, and perennial crop age-related dynamics are poorly simulated by models despite their importance to making future predictions of bioenergy crop GHG balances. With mitigation strategies being developed for decarbonization of the atmosphere, it is becoming increasingly important for biogeochemical models to confidently project GHG emissions in the future for all ecosystem types, land uses, and climatic variability. This is especially true when evaluating the impact of potential bioenergy crops as they are genetically manipulated to maximize biofuel production.

In this review, we discuss the history, current status, and future directions of Century-based biogeochemical modeling in the context of better simulating GHG balances of bioenergy feedstock production, as well as general improvements to soil carbon dynamics that may be used more broadly across all ecosystems. This manuscript covers (a) the history of soil carbon modeling with Century-based biogeochemical models and their relevance in modern ecosystem modeling, (b) future direction of soil carbon modeling, (c) modifications necessary to accurately simulate bioenergy feedstock growth and functions, and (d) how machine learning may be used to refine modeling of N$_2$O fluxes despite the lack of high-resolution spatial and temporal data.

2 | THERE IS A RICH HISTORY OF CENTURY-BASED MODELS UTILIZING A SIMPLE BUT POWERFUL FRAMEWORK FOR PREDICTING SOIL CARBON POOLS

The Century model was initially used to predict the extent to which abiotic and biotic factors control carbon cycling in temperate grasslands (Parton, Schimel, Cole, & Ojima, 1987). Over time, a large number of contemporary earth system models (Table 1) have adopted the Century framework as the core prediction of carbon cycling in terrestrial ecosystems. It is difficult to find a study comparing the performance of soil carbon dynamics between models that does not include Century or DayCent, the daily time-step version of Century (Parton, Hartman, Ojima, & Schimel, 1998), among the ranks (Abramoff

et al., 2018; Sulman et al., 2018; Walker et al., 2014; Wieder, Bonan, & Allison, 2013; Ye, Bradford, Dacal, Maestre, & García-Palacios, 2019). Century runs on a monthly time-step and simulates carbon, nitrogen, phosphorous, and sulfur dynamics within soil and vegetation of a specified system. The soil organic matter (SOM) sub-model uses three major pools to model carbon flow through the soil relying on first-order kinetics (decay constants; k-values) and soil texture (Figure 1a). This system consists of discrete pools to represent SOM and litter. SOM is partitioned into active, slow, and passive pools, each with a different intrinsic rate of decomposition which is altered by abiotic factors (i.e., soil temperature and soil moisture; Parton, 1996). Most of the models reviewed in Table 1 share similar methods of simulating soil carbon dynamics using decay constants to drive decomposition and carbon transfer between pools. While some models, such as DayCent, CASA (Wang, Law, & Pak, 2010), and DNDC (Li, Frolking, & Frolking, 1992; Li, Frolking, & Harriss, 1994), include a microbial biomass pool, they do not use microbial biomass, microbial enzyme production, or microbial carbon use efficiency (CUE; the ratio of carbon assimilated that is converted to biomass vs. that which is respired) to drive decomposition rates. The microbial biomass pool simply serves as another carbon pool with a different rate of carbon loss and transfer.

While Century-like soil carbon dynamics have been the staple for many models over the past three decades (Table 1), there is a movement to improve how decomposition is represented in models by progressing beyond k-value driven first-order kinetic equations that lack the complexity necessary to capture smaller scale temporal dynamics that are important for balancing carbon budgets (Wieder et al., 2013). Recent model advances have shown that explicit representation of soil microbial physiology leads to divergent soil carbon flux trajectories not captured by models that rely on first-order kinetics. For example, microbial models have been successful in capturing ephemeral increases in decomposition due to warming (Allison, Wallenstein, & Bradford, 2010), re-wetting (Evans, Dieckmann, Franklin, & Kaiser, 2016), and root-priming (Sulman et al., 2017). Researchers argue that temperature sensitivities are too variable across space to work well with single site observational data and therefore do not allow models of decomposition to be truly mechanistic (Davidson, Janssens, & Luo, 2006). As such, earth system model development (CLM; Table 1) has begun to integrate microbial process-based reverse Michaelis–Menten kinetics (Sulman et al., 2019; Wieder et al., 2013) as well as incorporating vertical connectivity to the Century-based model of decomposition already present in CLM (Koven et al., 2013). However, microbial process-based mechanisms are not being incorporated into the CMIP6 version of CLM. To date, the only CMIP6 model that is incorporating microbial processes (i.e., microbial biomass has an influence on decomposition rates) is IPSL-CM6A-LR which uses the ORCHIDEE-PRIM biogeochemical model.

TABLE 1 Selected CMIP6 earth system models (ESM) and their land and biogeochemical/soil sub-models. Models not a part of CMIP6 are designated by "NA." Soil sub-model traits are included. Even though some have microbial pools, microbes do not influence the rate of decomposition, except for ORCHIDEE-PRIM and FUN-CORPSE. "Optimum" describes equations where temperature or moisture increases until a point and then decreases in decomposition. YASSO uses precipitation events rather than soil moisture to influence decomposition. When N (nitrogen) is "No," the model is carbon only; otherwise the model includes N dynamics

Soil/ecosystem model	ESM/land model	No. litter/soil/microbe pools	Temperature response	Moisture	N	Decomposition driver
RothC (Harper et al., 2018)	UKESM1/JULES	0/4/1	Q10	Optimum	No	First-order kinetics
Soil-submodel of CTEM (Arora, 2003)	CanESM5/CTEM	1/1/0	Q10	Optimum	No	First-order kinetics
Century + PRIM (Guenet, Moyano, Peylin, Ciais, & Janssens, 2016)	IPSL-CM6A-LR/ ORCHIDEE-PRIM	2/4/0	Arrhenius	Optimum	Yes	Optimized first-order kinetics + priming effect
Modified Century or CLM-CN (Lawrence et al., 2019)	CESM2/CLM5	3/3/0	Q10	Increasing	Yes	First-order kinetics
CEVSA (Wu et al., 2019)	BCC-CSM2_MR/AVIM2.0	2/6/0	Optimum	Optimum	Yes	First-order kinetics
ELMv1 (Zhu et al., 2019)	E3SM (previously ACME)	3/3/3	Q10	Optimum	Yes	First-order kinetics + Equilibrium chemistry approximation
YASSO (Goll et al., 2015)	MPI-ESM1.2-LR/JSBACH3.2	4/5/0	Optimum	Precipitation	No	First-order kinetics
DNDC (Li et al., 1992)	NA	3/3/1	Optimum	Optimum	Yes	First-order kinetics
CASA (Wang et al., 2010)	NA	2/3/0	Q10	Optimum	Yes	First-order kinetics
MIMICS or MIMICS-CN (Wieder, Grandy, Kallenbach, & Bonan, 2014)	NA	2/3/2	Q10	Optimum	Yes	Forward or Reverse Michaelis–Menten kinetics
FUN-CORPSE (Sulman et al., 2017)	NA	3/2/3	Q10	Optimum	Yes	Reverse Michaelis–Menten kinetics
DayCent (Parton et al., 1998)	NA	2/5/1	Variable Q10	Optimum	Yes	First-order kinetics

(a) Decomposition driven by first-order kinetics:

$$\frac{\Delta C_i}{\Delta t} = K_i \cdot M_d \cdot T_d \cdot C_i$$

(b) Decomposition driven by microbial biomass using reverse Michaelis–Menten kinetics:

$$D_{C,i} = V_{max,i}(T) \cdot \left(\frac{\theta}{\theta_{sat}}\right)^3 \left(1 - \frac{\theta}{\theta_{sat}}\right)^{2.5} \cdot C_{U,i} \frac{c_M/c_{U,i}}{\frac{c_M}{c_{U,i}}+k_M}$$
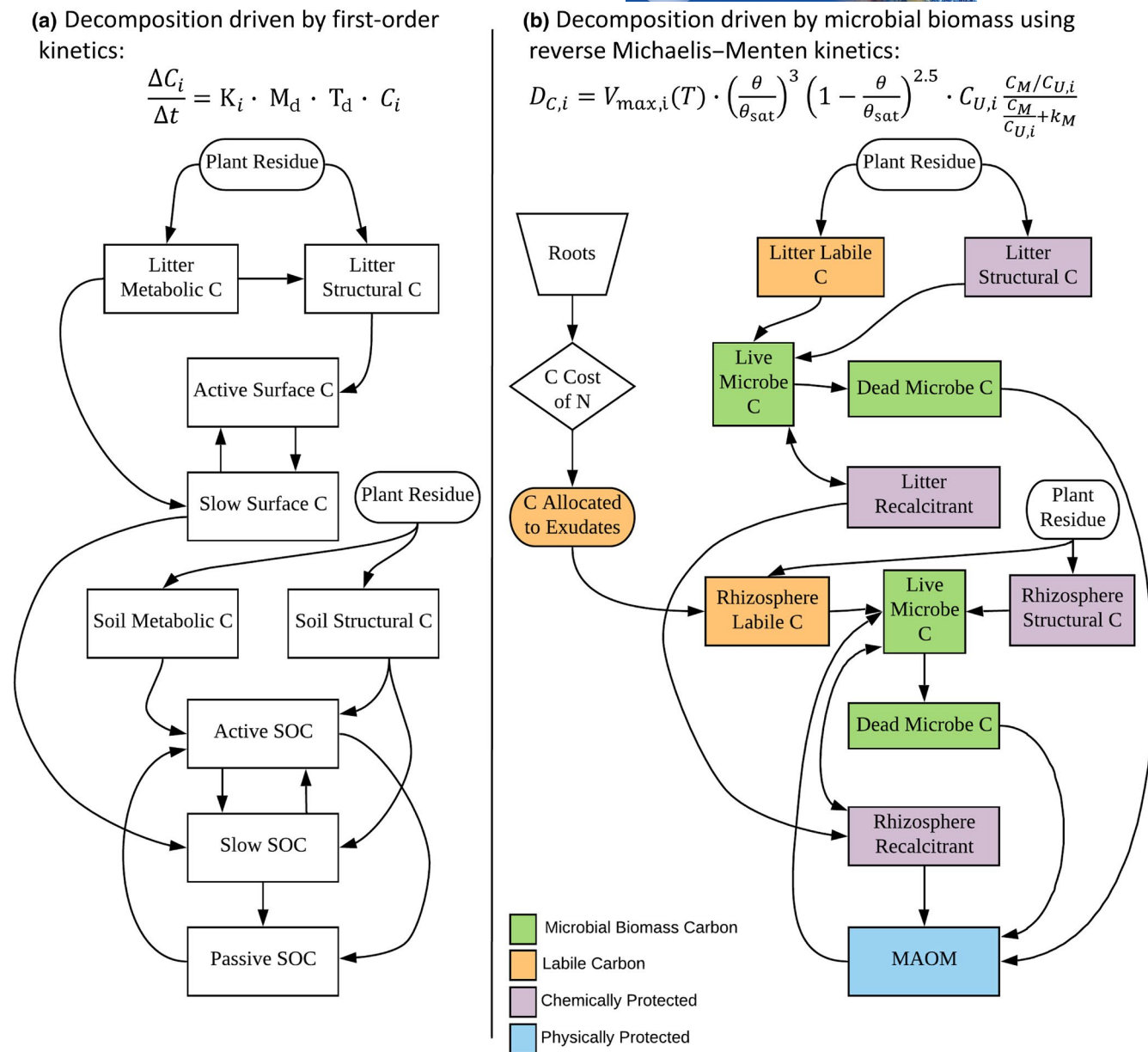

FIGURE 1 DayCent soil sub-model and the proposed FUN-CORPSE integration to Century-based soil submodules. (a) In the current DayCent soil sub-model, each litter and soil pool have an empirically derived k-value (K_i) that drives the rate of decomposition along with a temperature (T_d) and moisture (M_d) effect. C_i represents the size of the carbon pool. (b) Suggested changes to Century-based soil submodules are (1) updating the decomposition function to include microbial biomass pools and use reverse Michaelis–Menten kinetics to drive decomposition in each pool, (2) redefining pool structure to align with microbial, chemical, and physical functions, and (3) use the carbon cost of nitrogen acquisition by plants described in the FUN model to determine carbon allocation from plants in the form of root exudates to the soil metabolic carbon pool. Carbon dioxide is lost as respiration when there is a transfer from a pool to one of the live microbial C pools

3 | NEW EMPIRICAL EVIDENCE IS BEING UTILIZED TO IMPROVE MODEL MECHANISMS AND CONSTRAIN MODEL PARAMETERS

The ability of models to accurately predict biogeochemical cycling of bioenergy systems ultimately relies on the quantity and quality of data available for parameterization and validation. When initially developed, Century was validated using

ecosystem carbon and nitrogen budgets with the goal of accurately simulating biomass yields and equilibrium soil carbon and nitrogen stocks across grassland ecosystems (Parton et al., 1987). Century used these budgets to distill first principle controls (climate, tissue chemistry, etc.) on plant productivity and SOM decomposition into simple yet powerful mechanistic representations of these complex processes. As technology progressed, particularly with the advent of eddy-covariance techniques to quantify ecosystem GHG budgets,

these new data streams were used to push the validation of Century and DayCent beyond ecosystem carbon and nitrogen pools to the fluxes that control them. Given the development and maturity of new empirical techniques, there is now the potential to delve even deeper and refine the mechanistic controls on these fluxes. For example, the ability to rapidly measure key leaf and photosynthetic traits (e.g., leaf mass per area, nitrogen content, V_{max}, etc.) of bioenergy feedstocks can improve plant functional type (PFT) parameterizations of such traits in DayCent. Belowground, the maturity of soil fractionation techniques has led to the development of soil decomposition models (MEMS; Cotrufo, Wallenstein, Boot, Denef, & Paul, 2013; Robertson et al., 2019) that simulate soil carbon fractions that are measurable in the field, thereby replacing the conceptual active, slow, and passive fractions used by Century-based models.

4 | NEW PARADIGMS OF SOM FORMATION, STABILIZATION, AND LOSS ARE NOT INCLUDED IN ECOSYSTEM AND EARTH SYSTEM MODELS

Despite the success of models that rely on decay constants or "*k*-values" to control the turnover rates of SOM pools, recognition of new paradigms describing the formation, stabilization, and loss of SOM may improve the ability of these models to predict the environmental impact of bioenergy crops (and other agricultural and native plants) as we transition to novel climatic conditions in the future. At the heart of these new theories is the concept that plant–microbial interactions govern the extent to which plant-derived inputs (i.e., root and leaf litter) form stable SOM (Sokol & Bradford, 2019). For bioenergy crops, there are two key plant traits that feedback on SOM persistence. The first is that there is a strong interaction between feedstock litter chemistry and microbial traits. The second is that differences between feedstocks in belowground carbon allocation to rhizosphere microbes can enhance the formation of stable SOM.

Feedstock litter chemistry drives keystone microbial traits that control the rate and pathway of SOM formation. For example, feedstocks like corn with low C:N ratio litters promote microbes with high CUE (carbon assimilated into biomass per unit carbon taken up) and fast turnover rates (Cotrufo et al., 2013; Zhu, Liang, Masters, Kantola, & DeLucia, 2018). The resulting pool of dead microbial products is thought to be preferentially sorbed to clay minerals in the soil making it physically protected from microbial attack (Schmidt et al., 2011). By contrast, plants with high C:N ratio litters like Miscanthus require greater enzyme investment for microbial decomposition, resulting in lower CUE and turnover rates. As a result, more SOM is chemically stabilized

as particulate organic matter because microbial decay is energetically limited (Castellano, Mueller, Olk, Sawyer, & Six, 2015). Why do these pathways matter for models? SOM that is chemically protected is highly susceptible to loss in a warmer world because rising temperatures lower the activation energy of microbial decay. As such, models that can mechanistically link microbial traits and activity with the distribution of SOM that is physically versus chemically protected are essential to projecting differences between feedstocks in SOM persistence.

To improve DayCent's ability to better represent plant–microbe interactions and to simulate soil carbon dynamics under future climate scenarios, we plan to integrate microbial explicit mechanisms of the coupled soil carbon and nitrogen cycling model, FUN-CORPSE (Figure 1; Sulman et al., 2017) into DayCent. The nitrogen portion of the model, FUN (Fixation and Uptake of Nitrogen), estimates the carbon cost of nitrogen to determine carbon allocation from plants to stimulate N cycling in the rhizosphere (Brzostek, Fisher, & Phillips, 2014; Fisher et al., 2010). The SOC model, CORPSE (Carbon, Organisms, Rhizosphere, and Protection in the Soil Environment), integrates microbial explicit controls of SOC transformations with feedbacks between live microbial biomass and decomposition to drive soil carbon efflux and stabilization (Sulman, Phillips, Oishi, Shevliakova, & Pacala, 2014). We select FUN-CORPSE as a guide to modifying DayCent's SOM cycling processes because, unlike many other microbial explicit SOM models, FUN-CORPSE simulates N and well and C cycling, which are both vital to DayCent's plant growth and trace gas models, and has been evaluated against measurements of rhizosphere fluxes and SOC (Sulman et al., 2017, 2018). While the mechanisms driving decomposition vary, much of the core structural components of the DayCent SOM cycling serve the similar roles as pools in microbial explicit models, such as FUN-CORPSE. For example, the metabolic litter pools in DayCent serve the same function as the labile carbon pools in FUN-CORPSE. The structural litter and slow SOC pools in DayCent are similar to the recalcitrant or chemically protected pools with varying levels of protection. DayCent's passive pool, defined as microbially derived inputs sorbed to soil clay particles, is representative of physically protected carbon. DayCent's active pool is its representation of live microbial biomass. While there are similarities between these pools, there is also a fundamental need to redefine and reroute flows between pools so they serve more mechanistic functions that align with microbial frameworks.

While maintaining a similar structure of the DayCent's SOM cycling sub-model (Figure 1a), we propose redefining pools to match advances in functional understanding in accordance with the MEMS framework (Cotrufo et al., 2013; Robertson et al., 2019). New pool definitions would align with measurable pools that serve a functional role similar to

those described by Lavallee, Soong, and Cotrufo (2020) to move beyond the conceptual active, slow, and passive framework in Century-based models.

Here, we rename the metabolic surface and rhizosphere litter pools as labile carbon pools to align better with contemporary terminology. They serve the same role as a category of plant residues with simple chemical structure entering the litter or rhizosphere layers. However, the labile pools will now be routed directly into live microbial pools. Recalcitrant plant material (e.g., lignin, cellulose, etc.) will still enter the litter or rhizosphere through pools categorized as structural carbon. Carbon from this pool will pass through the live microbial carbon pools. After microbial processing of structural carbon, this material will be transferred to the chemically protected, recalcitrant carbon pools. Recalcitrant carbon pools can may undergo further microbial decomposition or experience physical transfer of material from the litter recalcitrant pool to the rhizosphere recalcitrant pool. The recalcitrant carbon pools in the litter and rhizosphere will take the place of the slow carbon pools in Century-based models.

Carbon that enters the live microbial biomass pool will either go toward microbial biomass growth or will be lost as CO_2 to the atmosphere based on the CUE of the original pool. The live microbial biomass pool can transfer material to the dead microbial carbon pool as microbial necromass or biproducts. From there, carbon can be transferred abiotically to the Mineral Associated Organic Matter pool (MAOM); this will replace the conceptual passive soil carbon pool in Century-based models. There will be a carbon saturation level of the MAOM pool that will be based on the mineral soil composition (sand, silt, and clay content). Carbon, regardless of the source, must undergo microbial processing before it can be incorporated into the MAOM pool.

We are also integrating mechanisms and parameters from FUN-CORPSE that will incorporate more plant–soil and soil–microbial interactions (Figure 1b). We will replace DayCent's decomposition function with the microbial explicit, reverse Michaelis–Menten kinetics functions, and add the carbon cost of nitrogen function to determine carbon allocation as root exudates to the rhizosphere carbon pool. The new live microbial biomass pools for each, the litter and rhizosphere allow for the feedbacks between decomposition rate and microbial biomass growth and the transfer of microbial necromass to the MAOM pool. While here, we are describing how microbial processes may be incorporated specifically into DayCent, this may be applied similarly to other Century-based soil carbon models.

As microbial processes are incorporated into ecosystem models, the challenge will be to attain parameterization and validation data at larger spatial scales with varying microbial and plant communities. For example, to date, FUN-CORPSE has only been validated in Eastern USA deciduous forests. Microbial response has been a focus of many warming studies but changes in microbial CUE, microbial biomass, and enzyme activity vary in direction and duration of response (Li et al., 2019; Ye et al., 2019). While studies are addressing the mechanisms driving this variation (Alvarez et al., 2018), more measurements are necessary to validate models across larger spatial scales. Recent studies have found that CUE of microbial communities has a stronger correlation with mean annual temperature (MAT) rather than assay temperature (Sinsabaugh, Moorhead, Xu, & Litvak, 2017; Takriti et al., 2018). Recently, Ye et al. (2019) evaluated using MAT as a driver of CUE and microbial extracellular enzyme kinetics (V_{max}, the maximal activity, and k_m, the half-saturation constant) using three microbial-explicit models compared with an early version of the Century soil model (Parton, Stewart, & Cole, 1988). They found that there was a strong positive relationship between CUE and MAT and that the microbial models performed better than the first-order kinetics model. In both DayCent and FUN-CORPSE, each SOM pool has its own CUE, and while it can be tuned during model calibration, direct measurements are often not available. Until measurements of microbial traits become more widespread, developing and testing the use of empirical relationships with microbial traits, such as that between CUE and MAT, is a step toward more data-driven parameterization. As models progress toward including the more process-based microbial mechanisms of decomposition, it will be vital that we continue to expand on field-based studies that can evaluate microbial model parameterizations and results.

5 | BIOENERGY CROP MODELING LACKS REPRESENTATION OF VEGETATION DYNAMICS AND LANDSCAPE HETEROGENEITY OVER TIME

The perennial growth form of many high-yielding bioenergy crops, including Miscanthus (*Miscanthus × giganteus*), switchgrass (*Panicum virgatum*), and sugarcane (*Saccharum officinarum*) pose a special challenge to modelers as endogenous (i.e., aging) and management factors, in addition to the environment, may affect their yields over long periods of time (Figure 2). For several years after planting, perennial bioenergy grasses produce relatively low yields (i.e., establishment effects; Figure 2b) and are more vulnerable to nutrient and water deficiencies than established stands. Because of high costs of crop establishment and for the construction of biorefineries, it is desirable that the yield of bioenergy crops remain high and stable over many years; however, this may not be the case. Sugarcane yields decline precipitously after only two or three cycles of vegetative growth—ratoon cycles (Smith, Inman-Bamber, & Thorburn, 2005) necessitating replanting. Generalizations about the yield stability
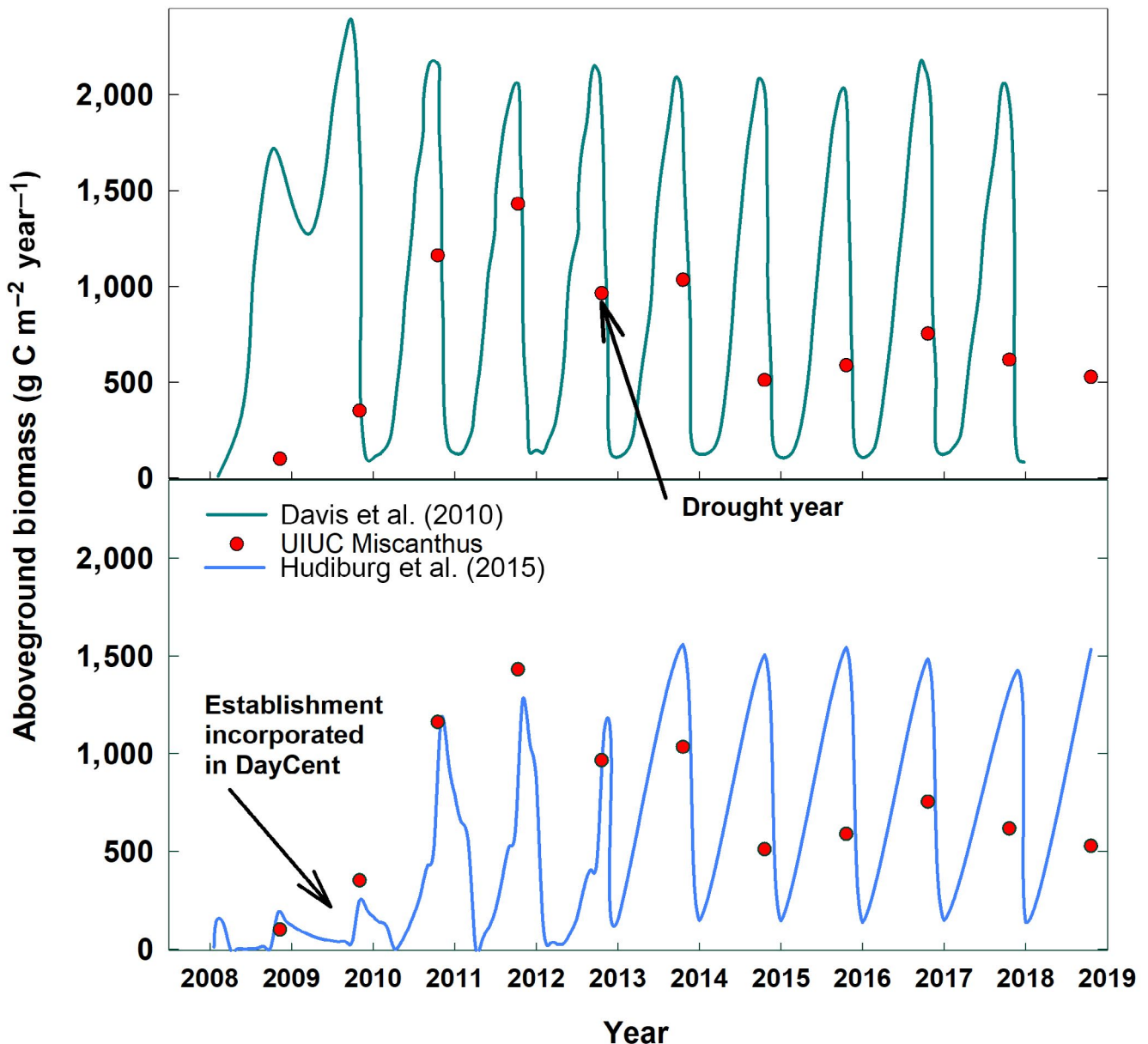
FIGURE 2 Modeled and observed aboveground biomass of Miscanthus from two studies at the University of Illinois Urbana Champaign Energy Farm. Davis et al. (2010) did not have any Miscanthus validation data at the time of the study. Hudiburg et al. (2015) had biomass data for validation and used daily site weather data through 2012 when the study was finished. In other words, observations at the time of these studies did not have either a drought or aging-induced decline in yield and in the case of Davis et al. (2010), an establishment phase. DayCent's crop model (primarily used for annual crops) was not designed to capture these types of dynamics because annual crops do not have multi-year establishment or multi-year responses to past events. These new observations have led to the development of a new plant sub-model in DayCent that is now being used for the perennial bioenergy crops

of Miscanthus and switchgrass are more difficult as few long-term data sets are available. After an initial establishment phase where yields increase for 1–5 years, Arundale et al. (2014) observed a gradual decline in yield for both species thereafter. This study was conducted on rich agricultural soils in the Midwest US without fertilization. After peak yields, reductions also were observed for these species

in other regions; however, observed yield reductions are not universal. For Miscanthus, it took longer (>15 years) for these reductions to become apparent in the Mediterranean (Alexopoulou et al., 2015).

Predicting age-related declines in yield remains a key challenge for models. Adding to this challenge, the empirical understanding of the mechanisms driving time-dependent

yield reductions is limited. Excessive tillage at planting followed by soil compaction contribute to the strong decline in sugarcane yield from 1 year to the next (Pankhurst et al., 2003). Nutrient management for Miscanthus and switchgrass is not well understood (Heaton, Dohleman, & Long, 2009). Particularly for Miscanthus, which can yield well in excess of 20 dry Mg ha^{-1} year^{-1} (Laurent, Pelzer, Loyce, & Makowski, 2015), removal of nitrogen and phosphorus from the soil as plant biomass is harvested can contribute to declining yields. Miscanthus more than switchgrass is sensitive to hard winter freezes which can damage overwintering rhizomes and reduce yields the following summer (Clifton-Brown & Lewandowski, 2000). Physiologically, the extraordinarily dense canopies developed by these crops can reduce photosynthetic efficiency by shading (Pignon, Jaiswal, McGrath, & Long, 2017). There is ongoing work that is using a staggered stand age to separate age dynamics from environmental effects on Miscanthus physiology and yield (Tejera et al., 2019). Studies such as this will provide insight to develop a more comprehensive and integrated understanding of mechanisms driving these changes, including the role of management practices, edaphic and biotic factors, and endogenous physiological controls of productivity. It is critical to continue incorporating emerging mechanistic improvements as they are identified in the literature.

In addition to an establishment phase for Miscanthus, a decline in yield was observed following the 2012 drought at the University of Illinois Energy Farm (Figure 2; red dots); this decline has continued. DayCent model predictions for Miscanthus yield at the Energy Farm prior to planting (Figure 2a; Davis et al., 2010) neither capture the establishment phase nor a decline in yield (to drought or aging). Model predictions were improved in later modeling studies (Hudiburg, Davis, Parton, & Delucia, 2015) by adding an establishment phase by modifying DayCent's plant growth potential parameter for Miscanthus. However, representing the establishment phase did not improve the ability of the model to capture the response to drought, or yield declines due to aging. The suggested path to addressing age related decline is described in the following section describing long-term plant dynamics in DayCent.

To improve our ability to model drought effects on yield, we must be able to incorporate the ways in which plants respond to drought which varies between feedstocks. Switchgrass pursues a strategy of drought avoidance by investing in root mass to extract deep soil water, whereas Miscanthus has been found to have a variety of physiological responses to drought. When established Miscanthus stands were subjected to an extreme drought in 2012, Joo, Zeri, Hussain, DeLucia, and Bernacchi (2017) observed increased ET driven by deep soil water use by Miscanthus compared to switchgrass and prairie plots. In contrast, another study found that Miscanthus had lower evapotranspiration in 2012 than either switchgrass,

corn or mixed prairie (Hamilton, Hussain, Bhardwaj, Basso, & Robertson, 2015). Other research has suggested a lack of sensitivity of Miscanthus to moderate drought, with water use and photosynthesis remaining high through early phases of scarcity and stomatal closure only upon onset of severe drought (Ings, Mur, Robson, & Bosch, 2013). DayCent crop parameters currently allow specification of root allocation responses to water scarcity. However, the rate of consumption of available water is dictated by leaf area and climatic factors that are not specific to crop species. If field results establish clear differences in the water use strategies of perennial bioenergy crops, it may be necessary to introduce a water consumption algorithm that accounts for species-specific responses to drought.

6 | RECENT DAYCENT DEVELOPMENTS SHOW IMPROVED SHORT- AND LONG-TERM DYNAMICS

To simulate the dynamics of fast-growing, highly productive biofuel plants which have different dynamics and traits than more traditional crops and grasses, we have developed a new bioenergy grass PFT with additional physiological parameters for the DayCent ecosystem model. Although the new bioenergy grass can represent both annual and perennial crops, it was developed primarily to improve the representation of the long-term dynamics of large perennial biofuel grasses such as sugarcane, switchgrass, and Miscanthus. The grass category differs from the model's traditional crop/grass plant type by representing aboveground biomass as both stems and leaves instead of as shoots only and representing belowground root biomass as both rhizomes and fine roots instead of as fine roots only. The stems, which contain a larger percentage of structural material than leaves, have a higher C:N ratio than leaves and the amount of leaves that can grow is dependent on the amount of stem biomass that can support them. Phenology, such as when senescence begins, can be prescribed or can be governed by growing degree days and moisture-related triggers. Shading can cause partial senescence of leaves which can either remain attached to the stems or fall to the ground. During the latter part of the growing season, perennial plants build up carbohydrate storage which will be available for growth in the subsequent growing season. Similarly, during senescence large perennial grasses re-translocate nitrogen from leaves to the rhizomes and this stored nitrogen will be available for growth in the subsequent growing season. Thus, the health of the plant in one growing season may affect yields in the subsequent growing season.

New PFT parameterizations in DayCent-CABBI, a version of DayCent developed by the Center for Advanced Bioenergy and Bioproducts Innovation, have improved the
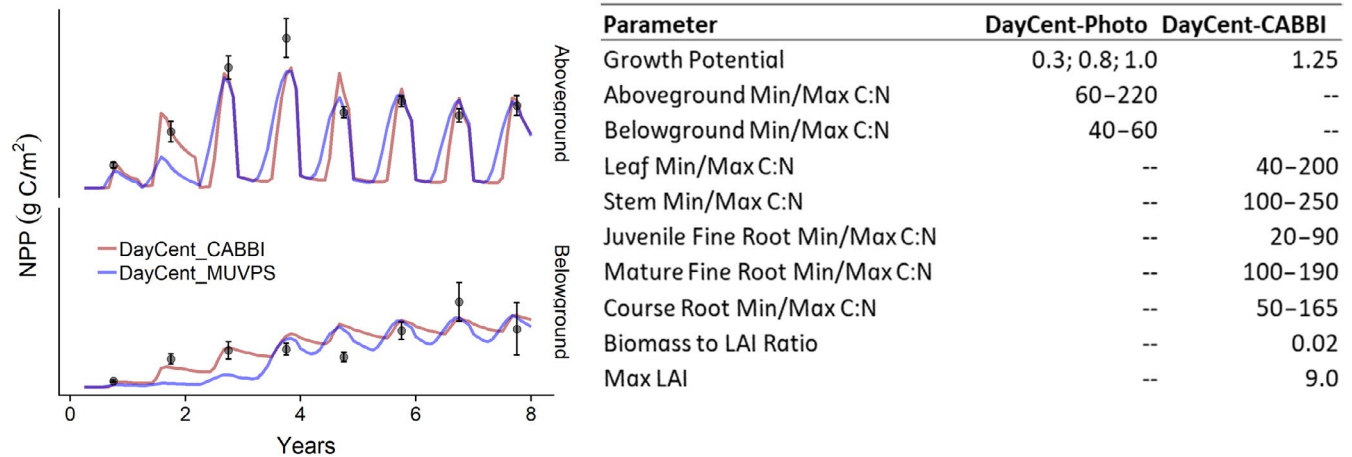
Parameter	DayCent-Photo	DayCent-CABBI
Growth Potential	0.3; 0.8; 1.0	1.25
Aboveground Min/Max C:N	60–220	--
Belowground Min/Max C:N	40–60	--
Leaf Min/Max C:N	--	40–200
Stem Min/Max C:N	--	100–250
Juvenile Fine Root Min/Max C:N	--	20–90
Mature Fine Root Min/Max C:N	--	100–190
Course Root Min/Max C:N	--	50–165
Biomass to LAI Ratio	--	0.02
Max LAI	--	9.0

FIGURE 3 DayCent-CABBI and DayCent-Photo (the previous version of the model) evaluated against measurements of peak aboveground and belowground peak biomass of switchgrass at the UIUC Energy Farm over 8 years following the initial planting. Points represent observed data. Error bars are standard error ($n = 5$). The table provides differences in major crop parameterizations between the old and the new version of the model. DayCent-Photo adjusting the growth potential coefficient during the first 2 years to limit growth during switchgrass establishment, whereas DayCent-CABBI achieves this though LAI parameters. The C:N and LAI values were derived from measurements taken at UIEF (Masters et al., 2016; Smith et al., 2013)

model representation of perennial plant establishment and growth (Figure 3; Moore et al., in press). Model-data agreement can be tuned with a comparable amount of success for both DayCent-CABBI and DayCent-Photo, the previous version of the model (Stenzel et al., 2019; Straube et al., 2018). However, to simulate lower yields during establishment years for perennials in DayCent-Photo, the model required three consecutive switchgrass PFT parameterizations to be used with varying growth potential (see table in Figure 3). In this comparison, DayCent-CABBI captured the effects of establishment in a single parameterization of switchgrass since simulated stem and rhizome growth required several years to achieve peak biomass through additional biomass to LAI and maximum LAI parameters. DayCent-CABBI also differentiates carbon and nitrogen allocation between leaves and stems and root tissues, whereas previous versions of the model only differentiate C:N ratios between aboveground and belowground pools.

To improve long-term dynamics, we are investigating whether long-term changes in allocation between aboveground and belowground biomass are altered by prolonged drought and plant age. We are also investigating the best way to represent the age-related declines in Miscanthus and switchgrass production that have been observed in field trials in the USA since we are unsure if these yield declines are a result of reduced soil nutrient availability, changing soil conditions or pests and disease. Furthermore, heterogeneous conditions within a field may reduce average biomass harvest over time. The model already simulates nutrient uptake and therefore simulated yields will respond to nutrient supply, but other age-related productivity declines are not currently represented (Arundale et al., 2014). These changes to model

algorithms will also be useful for woody plants, especially age-related changes to allocation of carbon.

The role of carbon and nitrogen stored belowground is also important to understanding stress response in perennial grasses. Eichelmann, Wagner-Riddle, Warland, Deen, and Voroney (2016) found that water-use efficiency of mature switchgrass increased in the 2012 drought year relative to 2011 despite lower CO_2 fixation, suggesting the use of belowground stored carbon to "subsidize" drought year biomass production. The new grass category provides for carbon accumulation in rhizomes which will allow for plants to mobilize carbon stored in roots under drought stress for aboveground biomass production. However, algorithms involved with the new grass category will need to be validated against measurements such as these as they become available to enable accurate prediction of yield stability across wet and dry growing seasons.

7 | CURRENT MECHANISTIC UNDERSTANDING ABOUT THE TIMING AND MAGNITUDE OF TRACE GAS FLUXES (e.g., N₂O) IS INSUFFICIENT FOR MODEL DEVELOPMENT

Given the amplified global warming potential of N_2O relative to other GHGs, inaccurate estimates of N_2O fluxes under land use or climate change scenarios represent a large source of uncertainty on terrestrial ecosystems-climate feedbacks, particularly when considering the agricultural sector (Stein & Yung, 2003). Accurate predictions of

potential feedbacks are key for evaluating risks and designing mitigation strategies.

Because a direct measurement of all factors and interactions integrating the nitrogen (N) cycle is unfeasible and subject to large variability, models of varying complexity have been developed to reproduce the complex processes driving N dynamics and ultimately N_2O emissions. DayCent was developed to link to a daily land surface model to improve estimates of trace gas fluxes by incorporating daily soil hydrological and thermal dynamics in conjunction with modified parameterizations to accommodate a higher temporal resolution of ecosystem processes subject to significant daily variability (Parton, 1996; Parton et al., 1998). DayCent simulates N_2O and NO_x emissions from nitrification and denitrification and N_2 emissions from denitrification reproducing the regulation of each process separately. The model assumes that nitrification rates are controlled by the availability in soil of NH_4^+, and soil water content, temperature, and pH (Parton et al., 2001). Maximum nitrification rates occur at close to 50% water-filled pore space (WFPS) and are assumed to decreases as temperature and pH decrease. Denitrification is assumed to be a function of soil NO_3^- (e^- acceptor) concentration, labile C (e^- donor) availability, and O_2 (competing e acceptor). Heterotrophic respiration is used as a proxy for labile C availability while O_2 availability is estimated from WFPS, and soil physical properties related to texture that influence gas diffusivity (Del Grosso et al., 2000).

The DayCent model has been extensively used to generate N_2O flux estimates for regional GHG inventories, predict N_2O emissions and N_2O emission factors across land uses and management practices, and evaluate the impacts of climate change on agriculture (Adler, Del Grosso, & Parton, 2007; Davis et al., 2010, 2012; Del Grosso et al., 2006; Del Grosso, Mosier, Parton, & Ojima, 2005; Del Grosso, Wirth, Ogle, & Parton, 2008). However, despite yielding more accurate estimates than the IPCC recommended EF methodology (Del Grosso et al., 2005) and a generally better performance relative to alternative process-based models (Abdalla et al., 2010; Ye et al., 2019), DayCent estimates are not devoid of uncertainty, with reported deviations ranging from −57% to 38% (Abdalla et al., 2010; Del Grosso et al., 2005; Gaillard et al., 2018; Ye et al., 2019). The complexity of model predictions of ecosystem N_2O fluxes lies largely on the fact that N_2O fluxes result from multiple processes, whose regulation, drivers, and interactions are still not sufficiently understood or even identified.

The lack of mechanistic understanding of N_2O dynamics is largely due to the highly variable nature of these fluxes. In the last few decades, it has become apparent that the majority of N_2O emissions from soil and aquatic systems occurs in hot spots and during hot moments (McClain et al., 2003). Sampling is limited in space and time leading to an uneven representation in the literature (e.g., more accessible vs. less accessible areas; growing season vs. non-growing season),

making it difficult to disentangle the primary variables driving the response of N_2O to changes in the environment.

Predicting ecosystem N_2O fluxes requires integration of multiple processes operating simultaneously (i.e., nitrification and denitrification). These processes have different lag-times and sensitivities to a wide range of biotic factors (e.g., plant–soil interactions, natural inhibitors of the nitrogen metabolism, and dominant microbial communities) and environmental conditions which are still not fully understood or even identified. The role of soil pH in denitrification metabolism has only recently been recognized (Liu, Mørkved, Frostegård, & Bakken, 2010) and new evidence suggests that the contribution of N_2O emissions induced by freeze–thaw cycles to annual emissions has been underestimated by ~20%–30% (Wagner-Riddle et al., 2017). Neglecting to capture these processes in biogeochemical models likely contributes to strong biases in ecosystem N_2O annual budgets.

Moreover, static model site (or grid cell) parameters such as soil texture, drainage, or pH may not be constant as they can evolve over the course of the development of supported vegetation, buildup of SOM, or particularly following disturbance events. A recent review found that there is frequently high N_2O emissions in the first few years following conversion of annual crops to perennial bioenergy feedstocks (Whitaker et al., 2018). To accurately predict ecosystem N_2O emissions, models need be able to integrate these legacy effects, where the short-term effects of changes in land use and/or climate on the ecosystem physical and biological parameters are foundational to long-term responses.

The complex functional relationships between N_2O and its controlling factors make it difficult for process-based biogeochemical models to predict spatial and temporal N_2O dynamics. Furthermore, the popular process-based models for N_2O prediction such as DayCent and DNDC are often parameterized based on limited number of observations from controlled laboratory experiments that do not represent the complex variable interactions. The process-based models often fail to accurately predict fine-scale (daily) temporal N_2O predictions (Jarecki, Parkin, Chan, Hatfield, & Jones, 2008; Parton et al., 2001). We propose that data-driven machine learning models could be used to improve predictability as well as understand controlling variable sensitivity and identify their functional relationships with N_2O, which, in turn, could help to refine the process-based models through improved parameterization (Perlman, Hijmans, & Horwath, 2014; Philibert, Loyce, & Makowski, 2013; Reichstein et al., 2019). For example, deep learning models can improve prediction accuracy for poorly understood processes at the expense of interpretability (Brenowitz & Bretherton, 2018; de Bezenac, Pajot, & Gallinari, 2019). Whereas tree-based models such as Random Forest is a non-parametric machine learning technique that learns functional forms between the response and predictor variables
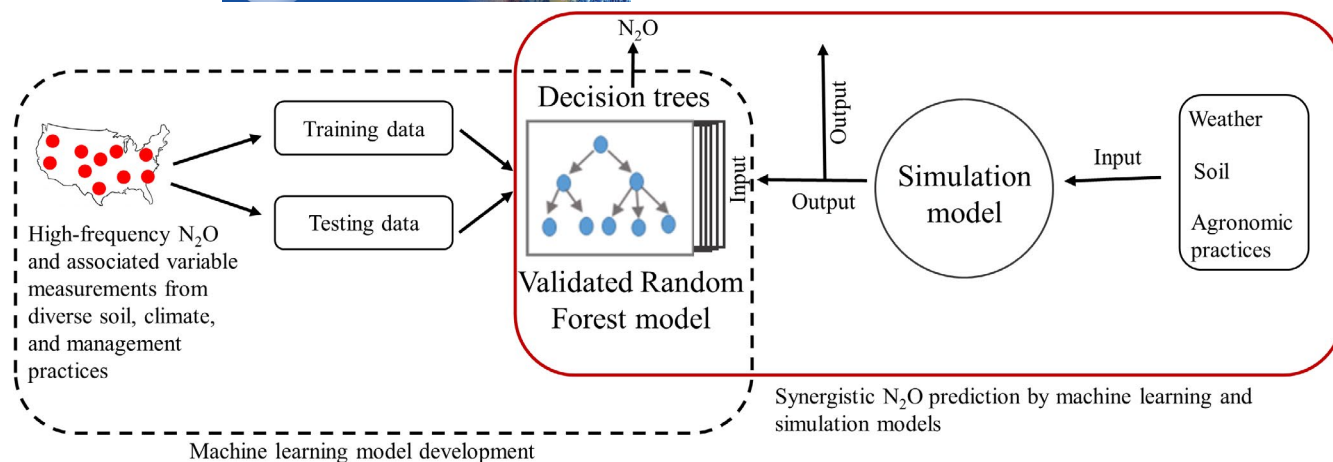
FIGURE 4 A schematic representation of integrating machine learning and biogeochemical simulation models to predict N_2O fluxes. The left broken box represents the development of machine learning model by learning from the data. To improve the machine learning model's ability to generalize predictions, the model should be trained and tested on data from diverse environmental and agronomic production systems. The right solid box represents output generation from process-based simulation model. The relevant outputs can serve as input variables for the machine model to predict N_2O fluxes at the same resolution as the simulation model (adapted from Saha, Basso, & Robertson, in review)

from the data (Breiman, 2001). Hence, can provide novel understanding on process controls while improving prediction accuracy. This process can operate in parallel or in fusion with biogeochemical or other process-based models (Walsh & Hudiburg, 2019). Unlike process-based models, the Random Forest algorithm learning is facilitated through an iterative process of recursive data partitioning and constructing hundreds of decision trees to partition the observations into distinct groups characterized by different properties of the predictor variables. Meta-modeling approach using Random Forest and process-based models has been used in predicting soil N biogeochemistry (Ramanantenasoa, Génermont, Gilliot, Bedos, & Makowski, 2019; Shahhosseini, Martinez-Feria, Hu, & Archontoulis, 2019); however, its use in predicting highly variable temporal soil N_2O fluxes is limited. Saha, Basso, and Robertson (in review) developed a Random Forest model based on automated flux chamber data from corn in the upper Midwest that predicted 51% variability in daily N_2O fluxes from an unknown site.

High temporal and spatial resolution input data are critical to facilitate this learning process; hence, measured data from diverse soil, climate, and cropping system management practices are critical for model training (Figure 4). As more long-term observed flux data from automated flux chamber sites become available, the opportunity to use machine learning methods to improve process-based biogeochemical models, like DayCent, is increasing. The predictor variables may include soil properties (texture organic matter, pH), weather variables (precipitation, temperature), management practices (tillage, cover crop, fertilization, and land use change), and dynamic soil biogeochemistry (mineral nitrogen availability, soil moisture, and temperature). As a way forward, we suggest first identifying the available measured N_2O data sources

from diverse soil, climate, and production systems. Second, creation of a database of N_2O fluxes and associated predictor variables that are empirically identified as important. Process-based models can be used to fill in the gaps in input data with a certain level of confidence to facilitate an integration of data-driven and physics-based N_2O modeling approach (Figure 4). Third, machine learning models should be trained and then tested with separate validation data. Fourth, performance comparisons between the process-based models and Random Forest method in predicting N_2O fluxes can be used as a metric of improved model confidence. This method could extend to other ecosystem fluxes that are also difficult to model under quickly changing environmental conditions such as CO_2 from heterotrophic respiration and CH_4 production and consumption.

8 | CONCLUSION

Modeling is a powerful tool for generating hypotheses and predictions about how major changes in land use, driven by the expansion of bioenergy crops and other changes in managed landscapes, will affect the land–atmosphere exchange of GHGs. Century-based models have been keystone models representing processes affecting SOC and GHG exchange. Assimilation of new understanding of plant physiology, new paradigms explaining changes in SOC, the ability to quantify feedbacks between the microbial community and decomposition, and the use of machine learning to identify and optimize key parameters when high-resolution data is lacking will improve the predictive power of DayCent when confronted with rapidly changing environmental conditions.

AUTHORS' CONTRIBUTIONS

D.B., T.W.H., and E.B. planned and designed the research. D.B., D.S., M.D.H., and T.W.H. designed figures and performed analysis. D.B., T.W.H., E.B., E.H.D., M.D.H., J.K., E.B.-B., B.D., and D.S. wrote the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Danielle Berardi https://orcid.org/0000-0003-1414-3080
Evan H. DeLucia https://orcid.org/0000-0003-3400-6286
Debasish Saha https://orcid.org/0000-0001-9425-675X
Tara W. Hudiburg https://orcid.org/0000-0003-4422-1510

REFERENCES

Abdalla, M., Jones, M., Yeluripati, J., Smith, P., Burke, J., & Williams, M. (2010). Testing DayCent and DNDC model simulations of N$_2$O fluxes and assessing the impacts of climate change on the gas flux and biomass production from a humid pasture. *Atmospheric Environment*, *44*(25), 2961–2970. https://doi.org/10.1016/j.atmosenv.2010.05.018

Abramoff, R., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., … Mayes, M. A. (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

Adler, P. R., Del Grosso, S. J., & Parton, W. J. (2007). Life-cycle assessment of net greenhouse-gas flux for bioenergy cropping systems. *Ecological Applications*, *17*, 675–691. https://doi.org/10.1890/05-2018

Alexopoulou, E., Zanetti, F., Scordia, D., Zegada-Lizarazu, W., Christou, M., Testa, G., … Monti, A. (2015). Long-term yields of switchgrass, giant reed, and Miscanthus in the Mediterranean basin. *Bioenergy Research*, *8*(4), 1492–1499. https://doi.org/10.1007/s12155-015-9687-x

Allison, S. D., Wallenstein, M. D., & Bradford, M. A. (2010). Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience*, *3*(5), 336–340. https://doi.org/10.1038/ngeo846

Alvarez, G., Shahzad, T., Andanson, L., Bahn, M., Wallenstein, M. D., & Fontaine, S. (2018). Catalytic power of enzymes decreases with temperature: New insights for understanding soil C cycling and microbial ecology under warming. *Global Change Biology*, *24*(9), 4238–4250. https://doi.org/10.1111/gcb.14281

Arora, V. K. (2003). Simulating energy and carbon fluxes over winter wheat using coupled land surface and terrestrial ecosystem models. *Agricultural and Forest Meteorology*, *118*(1–2), 21–47. https://doi.org/10.1016/S0168-1923(03)00073-X

Arundale, R. A., Dohleman, F. G., Heaton, E. A., Mcgrath, J. M., Voigt, T. B., & Long, S. P. (2014). Yields of *Miscanthus × giganteus* and *Panicum virgatum* decline with stand age in the Midwestern USA. *GCB Bioenergy*, *6*(1), 1–13. https://doi.org/10.1111/gcbb.12077

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289–6298. https://doi.org/10.1029/2018GL078510

Brzostek, E. R., Fisher, J. B., & Phillips, R. P. (2014). Modeling the carbon cost of plant nitrogen acquisition: Mycorrhizal trade-offs and multipath resistance uptake improve predictions of retranslocation. *Journal of Geophysical Research: Biogeosciences*, *119*(8), 1684–1697. https://doi.org/10.1002/2014JG002660

Castellano, M. J., Mueller, K. E., Olk, D. C., Sawyer, J. E., & Six, J. (2015). Integrating plant litter quality, soil organic matter stabilization, and the carbon saturation concept. *Global Change Biology*, *21*(9), 3200–3209. https://doi.org/10.1111/gcb.12982

Clifton-Brown, J., & Lewandowski, I. (2000). Overwintering problems of newly established Miscanthus plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. *The New Phytologist*, *148*(2), 287–294. https://doi.org/10.1046/j.1469-8137.2000.00764.x

Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K., & Paul, E. (2013). The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: Do labile plant inputs form stable soil organic matter? *Global Change Biology*, *19*(4), 988–995. https://doi.org/10.1111/gcb.12113

Davidson, E. A., Janssens, I. A., & Luo, Y. (2006). On the variability of respiration in terrestrial ecosystems: Moving beyond Q10. *Global Change Biology*, *12*(2), 154–164. https://doi.org/10.1111/j.1365-2486.2005.01065.x

Davis, S. C., Parton, W. J., Del Grosso, S. J., Keough, C., Marx, E., Adler, P. R., & DeLucia, E. H. (2012). Impact of second-generation biofuel agriculture on greenhouse-gas emissions in the corn-growing regions of the US. *Frontiers in Ecology and the Environment*, *10*(2), 69–74. https://doi.org/10.1890/110003

Davis, S. C., Parton, W. J., Dohleman, F. G., Smith, C. M., Del Grosso, S., Kent, A. D., & DeLucia, E. H. (2010). Comparative biogeochemical cycles of bioenergy crops reveal nitrogen-fixation and low greenhouse gas emissions in a *Miscanthus* x *giganteus* agro-ecosystem. *Ecosystems*, *13*(1), 144–156.

de Bezenac, E., Pajot, A., & Gallinari, P. (2019). Deep learning for physical processes: Incorporating prior scientific knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, *2019*(12), 124009. https://doi.org/10.1088/1742-5468/ab3195

Del Grosso, S. J., Mosier, A. R., Parton, W. J., & Ojima, D. S. (2005). DAYCENT model analysis of past and contemporary soil N$_2$O and net greenhouse gas flux for major crops in the USA. *Soil and Tillage Research*, *83*(1), 9–24. https://doi.org/10.1016/j.still.2005.02.007

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Ojima, D. S., Kulmala, A. E., & Phongpan, S. (2000). General model for N$_2$O and N$_2$ gas

emissions from soils due to dentrification. *Global Biogeochemical Cycles*, *14*(4), 1045–1060. https://doi.org/10.1029/1999GB001225

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Walsh, M. K., Ojima, D. S., & Thornton, P. E. (2006). DAYCENT national-scale simulations of nitrous oxide emissions from cropped soils in the United States. *Journal Environmental Quality*, *35*, 1451–1460. https://doi.org/10.2134/jeq2005.0160

Del Grosso, S. J., Wirth, T., Ogle, S. M., & Parton, W. J. (2008). Estimating agricultural nitrous oxide emissions. *EOS, Transactions, American Geophysical Union*, *89*(51), 529–529. https://doi.org/10.1029/2008EO510001

Eichelmann, E., Wagner-Riddle, C., Warland, J., Deen, B., & Voroney, P. (2016). Evapotranspiration, water use efficiency, and energy partitioning of a mature switchgrass stand. *Agricultural and Forest Meteorology*, *217*, 108–119.

Evans, S., Dieckmann, U., Franklin, O., & Kaiser, C. (2016). Synergistic effects of diffusion and microbial physiology reproduce the Birch effect in a micro-scale model. *Soil Biology and Biochemistry*, *93*, 28–37. https://doi.org/10.1016/j.soilbio.2015.10.020

Fisher, J., Sitch, S., Malhi, Y., Fisher, R., Huntingford, C., & Tan, S. Y. (2010). Carbon cost of plant nitrogen acquisition: A mechanistic, globally applicable model of plant nitrogen uptake, retranslocation, and fixation. *Global Biogeochemical Cycles*, *24*(1). https://doi.org/10.1029/2009GB003621

Gaillard, R. K., Jones, C. D., Ingraham, P., Collier, S., Izaurralde, R. C., Jokela, W., … Ruark, M. D. (2018). Underestimation of N₂O emissions in a comparison of the DayCent, DNDC, and EPIC models. *Ecological Application*, *28*, 694–708. https://doi.org/10.1002/eap.1674

Goll, D. S., Brovkin, V., Liski, J., Raddatz, T., Thum, T., & Todd-Brown, K. E. O. (2015). Strong dependence of CO₂ emissions from anthropogenic land cover change on initial land cover and soil carbon parametrization. *Global Biogeochemical Cycles*, *29*(9), 1511–1523.

Guenet, B., Moyano, F. E., Peylin, P., Ciais, P., & Janssens, I. A. (2016). Towards a representation of priming on soil carbon decomposition in the global land biosphere model ORCHIDEE (version 1.9. 5.2). *Geoscientific Model Development*, *9*(2), 841–855.

Hamilton, S., Hussain, M., Bhardwaj, A., Basso, B., & Robertson, G. (2015). Comparative water use by maize, perennial crops, restored prairie, and poplar trees in the US Midwest. *Environmental Research Letters*, *10*(6), 064015. https://doi.org/10.1088/1748-9326/10/6/064015

Harper, A. B., Wiltshire, A. J., Cox, P. M., Friedlingstein, P., Jones, C. D., Mercado, L. M., … Duran-Rojas, C. (2018). Vegetation distribution and terrestrial carbon cycle in a carbon cycle configuration of JULES4.6 with new plant functional types. *Geoscientific Model Development*, *11*(7), 2857–2873. https://doi.org/10.5194/gmd-11-2857-2018

Heaton, E. A., Dohleman, F. G., & Long, S. P. (2009). Seasonal nitrogen dynamics of *Miscanthus × giganteus* and *Panicum virgatum. GCB Bioenergy*, *1*(4), 297–307.

Hudiburg, T. W., Davis, S. C., Parton, W., & Delucia, E. H. (2015). Bioenergy crop greenhouse gas mitigation potential under a range of management practices. *GCB Bioenergy*, *7*(2), 366–374. https://doi.org/10.1111/gcbb.12152

Ings, J., Mur, L. A., Robson, P. R., & Bosch, M. (2013). Physiological and growth responses to water deficit in the bioenergy crop *Miscanthus* x *giganteus. Frontiers in Plant Science*, *4*, 468. https://doi.org/10.3389/fpls.2013.00468

Jarecki, M. K., Parkin, T. B., Chan, A. S., Hatfield, J. L., & Jones, R. (2008). Comparison of DAYCENT-simulated and measured nitrous oxide emissions from a corn field. *Journal of Environmental Quality*, *37*(5), 1685–1690. https://doi.org/10.2134/jeq2007.0614

Joo, E., Zeri, M., Hussain, M. Z., DeLucia, E. H., & Bernacchi, C. J. (2017). Enhanced evapotranspiration was observed during extreme drought from Miscanthus, opposite of other crops. *GCB Bioenergy*, *9*(8), 1306–1319. https://doi.org/10.1111/gcbb.12448

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., … Swenson, S. C. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, *10*(11), 7109–7131. https://doi.org/10.5194/bg-10-7109-2013

Laurent, A., Pelzer, E., Loyce, C., & Makowski, D. (2015). Ranking yields of energy crops: A meta-analysis using direct and indirect comparisons. *Renewable and Sustainable Energy Reviews*, *46*, 41–50. https://doi.org/10.1016/j.rser.2015.02.023

Lavallee, J. M., Soong, J. L., & Cotrufo, M. F. (2020). Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. *Global Change Biology*, *26*(1), 261–273. https://doi.org/10.1111/gcb.14859

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Vertenstein, M., … Kennedy, D. (2019). CLM5 documentation. Tech. rep., National Center for Atmospheric Research, Boulder, CO.

Li, C., Frolking, S., & Frolking, T. A. (1992). A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. *Journal of Geophysical Research: Atmospheres*, *97*(D9), 9759–9776. https://doi.org/10.1029/92JD00509

Li, C., Frolking, S., & Harriss, R. (1994). Modeling carbon biogeochemistry in agricultural soils. *Global Biogeochemical Cycles*, *8*(3), 237–254. https://doi.org/10.1029/94GB00767

Li, J., Wang, G., Mayes, M. A., Allison, S. D., Frey, S. D., Shi, Z., … Melillo, J. M. (2019). Reduced carbon use efficiency and increased microbial turnover with soil warming. *Global Change Biology*, *25*(3), 900–910. https://doi.org/10.1111/gcb.14517

Liu, B., Mørkved, P. T., Frostegård, Å., & Bakken, L. R. (2010). Denitrification gene pools, transcription and kinetics of NO, N₂O and N₂ production as affected by soil pH. *FEMS Microbiology Ecology*, *72*(3), 407–417.

Masters, M. D., Black, C. K., Kantola, I. B., Woli, K. P., Voigt, T., David, M. B., & DeLucia, E. H. (2016). Soil nutrient removal by four potential bioenergy crops: *Zea mays*, *Panicum virgatum*, *Miscanthus × giganteus*, and prairie. *Agriculture, Ecosystems & Environment*, *216*, 51–60. https://doi.org/10.1016/j.agee.2015.09.016

McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., … Pinay, G. (2003). Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems*, *6*(4), 301–312. https://doi.org/10.1007/s10021-003-0161-9

Moore, C., Berardi, D., Blanc-Betes, E., DeLucia, E. H., Dracup, E. C., Egenriether, S., … Bernacchi, C. J. (in press). The carbon and nitrogen cycle impacts of reverting perennial bioenergy switchgrass to an annual maize crop rotation. *GCB Bioenergy*.

Pankhurst, C., Magarey, R., Stirling, G., Blair, B., Bell, M., Garside, A., & Venture, S. Y. D. J. (2003). Management practices to improve soil health and reduce the effects of detrimental soil biota associated with yield decline of sugarcane in Queensland, Australia. *Soil and Tillage Research*, *72*(2), 125–137. https://doi.org/10.1016/S0167-1987(03)00083-7

Parton, W. (1996). The CENTURY model. *Evaluation of Soil Organic Matter Models*, *38*, 283–291.

Parton, W. J., Hartman, M., Ojima, D., & Schimel, D. (1998). DAYCENT and its land surface submodel: Description and testing. *Global and Planetary Change*, *19*(1–4), 35–48. https://doi.org/10.1016/S0921-8181(98)00040-X

Parton, W., Holland, E., Del Grosso, S., Hartman, M., Martin, R., Mosier, A., … Schimel, D. (2001). Generalized model for NOx and N$_2$O emissions from soils. *Journal of Geophysical Research: Atmospheres*, *106*(D15), 17403–17419.

Parton, W., Schimel, D. S., Cole, C., & Ojima, D. (1987). Analysis of factors controlling soil organic matter levels in Great Plains Grasslands. *Soil Science Society of America Journal*, *51*(5), 1173–1179. https://doi.org/10.2136/sssaj1987.03615995005100050015x

Parton, W. J., Stewart, J. W., & Cole, C. V. (1988). Dynamics of C, N, P and S in grassland soils: A model. *Biogeochemistry*, *5*(1), 109–131. https://doi.org/10.1007/BF02180320

Perlman, J., Hijmans, R. J., & Horwath, W. R. (2014). A metamodelling approach to estimate global N$_2$O emissions from agricultural soils. *Global Ecology and Biogeography*, *23*(8), 912–924.

Philibert, A., Loyce, C., & Makowski, D. (2013). Prediction of N$_2$O emission from local information with Random Forest. *Environmental Pollution*, *177*, 156–163. https://doi.org/10.1016/j.envpol.2013.02.019

Pignon, C. P., Jaiswal, D., McGrath, J. M., & Long, S. P. (2017). Loss of photosynthetic efficiency in the shade. An Achilles heel for the dense modern stands of our most productive C4 crops? *Journal of Experimental Botany*, *68*(2), 335–345.

Ramanantenasoa, M. M. J., Génermont, S., Gilliot, J.-M., Bedos, C., & Makowski, D. (2019). Meta-modeling methods for estimating ammonia volatilization from nitrogen fertilizer and manure applications. *Journal of Environmental Management*, *236*, 195–205. https://doi.org/10.1016/j.jenvman.2019.01.066

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

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

Saha, D., Basso, B., & Robertson, G. P. (in review). Machine learning reveals simplified path for predicting N$_2$O fluxes from agriculture [Manuscript submitted for publication].

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

Shahhosseini, M., Martinez-Feria, R. A., Hu, G., & Archontoulis, S. V. (2019). Maize yield and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters*, *14*(12), 124026. https://doi.org/10.1088/1748-9326/ab5268

Sinsabaugh, R. L., Moorhead, D. L., Xu, X., & Litvak, M. E. (2017). Plant, microbial and ecosystem carbon use efficiencies interact to stabilize microbial growth as a fraction of gross primary production. *New Phytologist*, *214*(4), 1518–1526. https://doi.org/10.1111/nph.14485

Smith, C. M., David, M. B., Mitchell, C. A., Masters, M. D., Anderson-Teixeira, K. J., Bernacchi, C. J., & DeLucia, E. H. (2013). Reduced nitrogen losses after conversion of row crop agriculture to perennial biofuel crops. *Journal of Environmental Quality*, *42*(1), 219–228. https://doi.org/10.2134/jeq2012.0210

Smith, D., Inman-Bamber, N., & Thorburn, P. (2005). Growth and function of the sugarcane root system. *Field Crops Research*, *92*(2–3), 169–183. https://doi.org/10.1016/j.fcr.2005.01.017

Sokol, N. W., & Bradford, M. A. (2019). Microbial formation of stable soil carbon is more efficient from belowground than aboveground input. *Nature Geoscience*, *12*(1), 46–53. https://doi.org/10.1038/s41561-018-0258-6

Stein, L. Y., & Yung, Y. L. (2003). Production, isotopic composition, and atmospheric fate of biologically produced nitrous oxide. *Annual Review of Earth and Planetary Sciences*, *31*(1), 329–356.

Stenzel, J. E., Bartowitz, K. J., Hartman, M. D., Lutz, J. A., Kolden, C. A., Smith, A. M. S., … Hudiburg, T. W. (2019). Fixing a snag in carbon emissions estimates from wildfires. *Global Change Biology*, *25*(11), 3985–3994. https://doi.org/10.1111/gcb.14716

Straube, J. R., Chen, M., Parton, W. J., Asso, S., Liu, Y.-A., Ojima, D. S., & Gao, W. (2018). Development of the DayCent-Photo model and integration of variable photosynthetic capacity. *Frontiers of Earth Science*, *12*(4), 765–778. https://doi.org/10.1007/s11707-018-0736-6

Sulman, B. N., Brzostek, E. R., Medici, C., Shevliakova, E., Menge, D. N., & Phillips, R. P. (2017). Feedbacks between plant N demand and rhizosphere priming depend on type of mycorrhizal association. *Ecology Letters*, *20*(8), 1043–1053. https://doi.org/10.1111/ele.12802

Sulman, B. N., Moore, J. A. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., … Classen, A. T. (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

Sulman, B. N., Phillips, R. P., Oishi, A. C., Shevliakova, E., & Pacala, S. W. (2014). Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO$_2$. *Nature Climate Change*, *4*(12), 1099. https://doi.org/10.1038/nclimate2436

Sulman, B. N., Shevliakova, E., Brzostek, E. R., Kivlin, S. N., Malyshev, S., Menge, D. N., & Zhang, X. (2019). Diverse mycorrhizal associations enhance terrestrial C storage in a global model. *Global Biogeochemical Cycles*, *33*(4), 501–523. https://doi.org/10.1029/2018GB005973

Takriti, M., Wild, B., Schnecker, J., Mooshammer, M., Knoltsch, A., Lashchinskiy, N., … Richter, A. (2018). Soil organic matter quality exerts a stronger control than stoichiometry on microbial substrate use efficiency along a latitudinal transect. *Soil Biology and Biochemistry*, *121*, 212–220. https://doi.org/10.1016/j.soilbio.2018.02.022

Tejera, M., Boersma, N., Vanloocke, A., Archontoulis, S., Dixon, P., Miguez, F., & Heaton, E. (2019). Multi-year and multi-site establishment of the perennial biomass crop *Miscanthus × giganteus* using a staggered start design to elucidate N response. *BioEnergy Research*, *12*(3), 471–483. https://doi.org/10.1007/s12155-019-09985-6

Wagner-Riddle, C., Congreves, K. A., Abalos, D., Berg, A. A., Brown, S. E., Ambadan, J. T., … Tenuta, M. (2017). Globally important nitrous oxide emissions from croplands induced by freeze–thaw cycles. *Nature Geoscience*, *10*(4), 279–283. https://doi.org/10.1038/ngeo2907

Walker, A. P., Hanson, P. J., De Kauwe, M. G., Medlyn, B. E., Zaehle, S., Asao, S., … Iversen, C. M. (2014). Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO$_2$ enrichment experiments: Model performance

at ambient CO_2 concentration. *Journal of Geophysical Research: Biogeosciences*, *119*(5), 937–964.

Walsh, E. S., & Hudiburg, T. (2019). An integration framework for linking avifauna niche and forest landscape models. *PLoS One*, *14*(6). https://doi.org/10.1371/journal.pone.0217299

Wang, Y., Law, R., & Pak, B. (2010). A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. *Biogeosciences*, *7*(7), 2261–2282. https://doi.org/10.5194/bg-7-2261-2010

Whitaker, J., Field, J. L., Bernacchi, C. J., Cerri, C. E. P., Ceulemans, R., Davies, C. A., … McNamara, N. P. (2018). Consensus, uncertainties and challenges for perennial bioenergy crops and land use. *GCB Bioenergy*, *10*(3), 150–164. https://doi.org/10.1111/gcbb. 12488

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

Wieder, W., Grandy, A., Kallenbach, C., & Bonan, G. (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

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., … Liu, X. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*(4), 1573–1600. https://doi.org/10.5194/gmd-12-1573-2019

Ye, J. S., Bradford, M. A., Dacal, M., Maestre, F. T., & García-Palacios, P. (2019). Increasing microbial carbon use efficiency with warming predicts soil heterotrophic respiration globally. *Global Change Biology*, *25*(10), 3354–3364. https://doi.org/10.1111/gcb.14738

Zhu, Q., Riley, W. J., Tang, J., Collier, N., Hoffman, F. M., Yang, X., & Bisht, G. (2019). Representing nitrogen, phosphorus, and carbon interactions in the E3SM land model: Development and global benchmarking. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2238–2258. https://doi.org/10.1029/2018MS001571

Zhu, X., Liang, C., Masters, M. D., Kantola, I. B., & DeLucia, E. H. (2018). The impacts of four potential bioenergy crops on soil carbon dynamics as shown by biomarker analyses and DRIFT spectroscopy. *GCB Bioenergy*, *10*(7), 489–500. https://doi.org/10.1111/gcbb.12520