



Divergent controls of soil organic carbon between observations and process-based models

Katerina Georgiou · Avni Malhotra · William R. Wieder ·
Jacqueline H. Ennis · Melannie D. Hartman · Benjamin N. Sulman ·
Asmeret Asefaw Berhe · A. Stuart Grandy · Emily Kyker-Snowman ·
Kate Lajtha · Jessica A. M. Moore · Derek Pierson · Robert B. Jackson

Received: 31 January 2021 / Accepted: 4 June 2021 / Published online: 16 July 2021
© The Author(s) 2021, corrected publication 2021

Abstract The storage and cycling of soil organic carbon (SOC) are governed by multiple co-varying factors, including climate, plant productivity, edaphic properties, and disturbance history. Yet, it remains unclear which of these factors are the dominant predictors of observed SOC stocks, globally and within biomes, and how the role of these predictors varies between observations and process-based models. Here we use global observations and an ensemble of soil biogeochemical models to quantify the emergent importance of key state factors – namely, mean

annual temperature, net primary productivity, and soil mineralogy – in explaining biome- to global-scale variation in SOC stocks. We use a machine-learning approach to disentangle the role of covariates and elucidate individual relationships with SOC, without imposing expected relationships *a priori*. While we observe qualitatively similar relationships between SOC and covariates in observations and models, the magnitude and degree of non-linearity vary substantially among the models and observations. Models appear to overemphasize the importance of temperature and primary productivity (especially in forests and herbaceous biomes, respectively), while observations suggest a greater relative importance of soil minerals. This mismatch is also evident globally. However, we observe agreement between observations and model outputs in select individual biomes –

Responsible Editor: R. Kelman Wieder.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10533-021-00819-2>.

K. Georgiou · A. Malhotra · J. H. Ennis · R. B. Jackson
Department of Earth System Science, Stanford University,
94305 Stanford, CA, USA

K. Georgiou (✉)
Physical and Life Sciences Directorate, Lawrence
Livermore National Laboratory, Livermore,
CA 94551, USA
e-mail: georgiou1@llnl.gov

W. R. Wieder · M. D. Hartman
Climate and Global Dynamics Laboratory, National
Center for Atmospheric Research, 80307 Boulder,
CO, USA

W. R. Wieder
Institute of Arctic and Alpine Research, University of
Colorado, Boulder, CO 80309, USA

M. D. Hartman
National Resource Ecology Laboratory, Colorado State
University, CO 80523 Fort Collins, USA

B. N. Sulman
Environmental Sciences Division, Oak Ridge National
Laboratory, 37831 Oak Ridge, TN, USA

A. A. Berhe
Department of Life and Environmental Sciences,
University of California, 95343 Merced,
CA, USA

namely, temperate deciduous forests and grasslands, which both show stronger relationships of SOC stocks with temperature and productivity, respectively. This approach highlights biomes with the largest uncertainty and mismatch with observations for targeted model improvements. Understanding the role of dominant SOC controls, and the discrepancies between models and observations, globally and across biomes, is essential for improving and validating process representations in soil and ecosystem models for projections under novel future conditions.

Keywords Soil carbon · Microbial models · Global change · Earth system models · Global databases · Model benchmarking · Machine learning

Introduction

Soil organic matter is the largest terrestrial pool of actively-cycling carbon, with the potential for large carbon-feedbacks in response to climate and land-use change (Friedlingstein et al. 2014; Todd-Brown et al. 2014; Wieder et al. 2018). The storage and persistence of soil organic carbon (SOC) exhibit considerable spatial heterogeneity and are governed by multiple key factors, including climate, plant productivity, edaphic properties, and disturbance history (Doetterl et al. 2015; Jackson and Caldwell 1993; Jackson et al. 2017; Rasmussen et al. 2018; Sollins et al. 1996). However, it remains unclear which of these predictors dominate in distinct biomes and across spatial scales. Furthermore, as soil biogeochemical models continue to evolve,

benchmarking techniques that validate not only the soil carbon stocks, but also the environmental sensitivity of soil carbon pools are increasingly needed (Sulman et al. 2018; Wieder et al. 2019a).

Soil biogeochemical models vary widely in structure and, consequently, in predicted carbon stocks and response to future perturbations (Abramoff et al. 2018; Georgiou et al. 2017; Li et al. 2014; Sulman et al. 2018). For decades, soil and ecosystem model development largely reflected traditional ideas of inherent chemical recalcitrance as a dominant control, and, consequently, the role of plant productivity and climate in regulating SOC storage (Schmidt et al. 2011; Wang et al. 2010). In recent years, however, emerging ideas place a growing emphasis on the importance of microbial activity and mineralogical properties (Cotrufo et al. 2013; Dwivedi et al. 2019; Grandy and Neff 2008; Lehmann et al. 2020) that have since been incorporated explicitly in process-based soil carbon models (Allison et al. 2010; Robertson et al. 2019; Sulman et al. 2014; Wang et al. 2013; Wieder et al. 2015a, b). Representing these evolving theories in mathematical models generates novel emergent behavior and alternative hypotheses about the relative importance of key state factors and their interactions. To adequately compare such models, it is important to recognize that there are numerous sources of uncertainty that can cause apparent downstream divergences (Sulman et al. 2018). Thus, it is helpful to force models with identical inputs and environmental conditions when comparing SOC stocks and their response to perturbations (e.g., within a model testbed; Wieder et al. 2018; Ahlström et al. 2017), to reduce sources of uncertainty that may confound the interpretation of results.

While the global distribution of SOC stocks may look reasonably similar across models and measurements (Fig. 1), it is difficult to diagnose where the models predict correctly and whether they do so for the right reasons. Indeed, recent studies that compare SOC stocks from models against long-term field manipulations have concluded that there are still large uncertainties in both the models and observations, and that novel observational syntheses and benchmarking techniques are needed (Sulman et al. 2018; Wieder et al. 2019b). Diagnosing model temperature and moisture sensitivities, for example, can be an informative exercise for parameterizing and improving model climate sensitivities (Abramoff et al. 2019;

A. S. Grandy · E. Kyker-Snowman
Department of Natural Resources and the Environment,
University of New Hampshire, 03824 Durham,
NH, USA

K. Lajtha · D. Pierson
Department of Crop and Soil Sciences, Oregon State
University, 97331 Corvallis, OR, USA

J. A. M. Moore
Biosciences Division, Oak Ridge National Laboratory,
37831 Oak Ridge, TN, USA

R. B. Jackson
Woods Institute for the Environment and Precourt
Institute for Energy, Stanford University, 94305 Stanford,
CA, USA

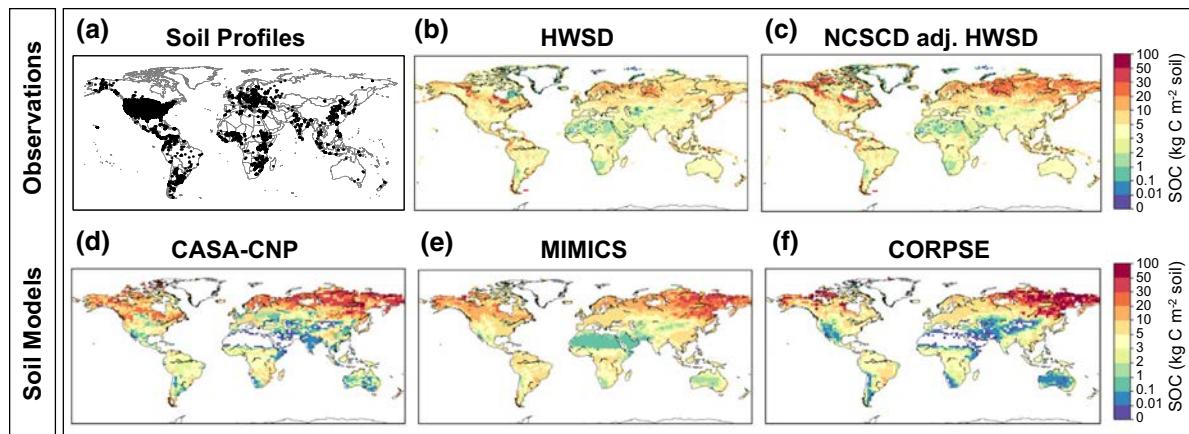


Fig. 1 Soil organic carbon observations and observationally-derived data products (a, b, c) and process-based soil model outputs (d, e, f) used in this study. Observations are from the a ISRIC WoSIS database soil profiles, b Harmonized World Soil Database (HWSD), and c Northern Circumpolar Soil

Carbon Database (NCSCD) adjusted HWSD. Model outputs are from the d CASA-CNP, e MIMICS, and f CORPSE soil biogeochemical models. Soil organic carbon (SOC) stocks are shown for all globally gridded data products and soil model outputs

Koven et al. 2017; Wieder et al. 2018). To examine a broader suite of controls that includes soil mineralogy and plant productivity, however, requires novel diagnostics. We also note that global soil data products used for model comparisons and benchmarking are often statistically-derived themselves (i.e., extrapolated from point observations) and may not carry forward the same underlying relationships with key environmental controls (Batjes 2016; Hengl et al. 2014). Therefore, it is critical to explore the environmental controls of SOC stocks in soil profiles, in addition to the controls of SOC stocks in commonly used global observationally-derived data products and biogeochemical models. Furthermore, dominant predictors may vary geographically across biomes in observations and biogeochemical models.

Here, we present a novel benchmarking approach that leverages machine-learning-based diagnostics. We explore key predictors of SOC stocks and how they differ between (i) soil profiles, (ii) observationally-derived data products, and (iii) biogeochemical models, globally and within biomes (Fig. 1). Specifically, we focus on climate, vegetation, and mineralogical controls that were present in all of the biogeochemical models compared in this study – namely, mean annual temperature (MAT), net primary production (NPP), and soil clay and silt content (i.e., soil texture; TEX). We anticipate that the predictability of SOC stocks with these select predictor variables will be lower for soil profiles and observationally-

derived data products, compared to outputs derived from biogeochemical models that have an imposed underlying structure. Models represent SOC dynamics as a function of only a select few input variables, and consequently model output may depict a lower variability in SOC stocks under a given set of environmental conditions compared to field measurements. Indeed, while other environmental controls are likely important for explaining SOC variability in field measurements (e.g., pH, precipitation, fire, nutrients), they are not yet incorporated into all of the soil biogeochemical models presented herein and are thus not compared in this benchmarking analysis. Evaluating SOC stocks and their dominant controls in global scale models is a critical step towards improving the understanding, model representation, and projection of factors controlling soil carbon persistence and potential responses under environmental change. Given temporal limitations in broad-scale SOC measurements, novel diagnostics using spatial gradients present opportunities to explore environmental sensitivities and long-term responses to perturbations.

Methods

Global observations

In this study, we used two types of observational and observationally-derived datasets: (i) a compilation of

soil profile measurements and (ii) observationally-derived gridded soil data products. Specifically, the ISRIC global soil database (World Soil Information System, WoSIS) was used for soil profile measurements and corresponding climate variables ($n > 10,000$ profiles) (Batjes 2009, 2016) (Fig. 1). For each soil profile, we summed SOC stocks to 1 m, and excluded profiles that were shallower than 80 cm, to best match the 1 m depth of the gridded soil data products and model outputs. For gridded soil data products, we used the Harmonized World Soil Database (HWSD) at $0.5^\circ \times 0.5^\circ$ resolution (Wieder et al. 2014b), but also ensured that our results were robust to the selected resolution (Table 1). Given that HWSD SOC stocks may be biased at high latitudes because of fewer observations, we also used a combined gridded data product where the Northern Circumpolar Soil Carbon Database (NCSCD) was used to replace HWSD values where overlap occurs (Hugelius et al. 2013). In our analyses, this product is called the NCSCD-adjusted HWSD (Fig. 1).

For gridded observational data we used soil texture (TEX; i.e., clay and silt content) from the HWSD at $0.5^\circ \times 0.5^\circ$ resolution. For the soil profiles, clay and silt content were reported directly in WoSIS and we used the corresponding profile averages in our analyses. Estimates of plant productivity were derived as 10-yr averages (2000–2010) from the MODIS net primary productivity (NPP) product at $0.5^\circ \times 0.5^\circ$ resolution and at soil profile locations (Koven et al. 2017; Zhao

et al. 2005). We used satellite-derived NPP estimates for all observational products in this analysis, but note that our findings were qualitatively insensitive to the choice of NPP, including using simulated NPP from the biogeochemical testbed forcing as used for the model output. Mean annual temperature (MAT) was estimated as a 10-yr average (2000–2010) from the CRU dataset at $0.5^\circ \times 0.5^\circ$ resolution and at soil profile locations (Harris et al. 2014). For a land classification map, we used the MODIS MCD12C1 landcover product for year 2010 at $0.5^\circ \times 0.5^\circ$ resolution (Friedl et al. 2010).

Models and simulations

We explored the primary controls of three global-scale soil carbon models: CASA-CNP (Carnegie-Ames-Stanford Approach model; Potter et al. 1993; Wang et al. 2010), MIMICS (MICROBIAL-MINERALIZATION Carbon Stabilization model; Wieder et al. 2015b), and CORPSE (Carbon, Organisms, Rhizosphere, and Protection in the Soil Environment model; Sulman et al. 2014) (Fig. 1; Table 1). These soil models form the foundation of the soil biogeochemical testbed (Wieder et al. 2018) and were chosen to represent different mechanistic representations actively used in global land models. We briefly describe the testbed and highlight features of each soil model, but refer readers to the papers above for more detailed

Table 1 Comparison of scales and predictability of soil organic carbon stocks from observations (soil profiles, HWSD, NCSCD adj. HWSD) and biogeochemical models (CASA-CNP, MIMICS, CORPSE)

	Type	Scale	Percent of variance explained by RF ^a	References
Soil profiles	Observation	1 m × 1 m	26%	WoSIS (Batjes 2009, 2016)
HWSD	Observation-derived	$0.5^\circ \times 0.5^\circ$ $1.9^\circ \times 2.5^\circ$	58%	Wieder et al. (2014b)
NCSCD adj. HWSD	Observation-derived	$0.5^\circ \times 0.5^\circ$	53%	Hugelius et al. (2013); Koven et al. (2017)
CASA-CNP	Model	$1.9^\circ \times 2.5^\circ$	63%	Wang et al. (2010)
MIMICS	Model	$1.9^\circ \times 2.5^\circ$	81%	Wieder et al. (2014a, 2015b)
CORPSE	Model	$1.9^\circ \times 2.5^\circ$	53%	Sulman et al. (2014)

See relevant publications and technical documentation for further details on the data and model details (Wieder et al. 2018)

^aRandom Forest (RF) models for each observation/model source of soil organic carbon (SOC) as a function of mean annual temperature (MAT), net primary production (NPP), and soil texture (TEX). Out-of-sample percent variance explained averaged over an ensemble of RF models with bootstrapped sampling

descriptions regarding specific model assumptions, structures, and parameterizations.

Our analysis used soil carbon stocks that were simulated from soil models in the biogeochemical testbed. All three soil carbon models in the testbed were forced with identical inputs and environmental conditions, thereby isolating the effect of underlying structural (and associated parametric) uncertainty. The testbed simulations used daily air temperature, gross primary production, soil temperature, and soil moisture that were generated by the Community Land Model (CLM version 4.5). The CLM4.5 simulations used satellite phenology and atmospheric forcing from the CRU-NCEP climate reanalysis from 1901 to 2010 (see Oleson et al. 2013). These simulations generated globally-gridded, daily data that were needed to run the CASA-CNP vegetation model (Randerson et al. 1996; Wang et al. 2010). Although CASA-CNP can simulate coupled carbon, nitrogen, and phosphorus biogeochemistry, here we used the carbon-only version of the model. The vegetation model in CASA-CNP calculates autotrophic respiration fluxes, allocation of carbon to different plant tissues, and the timing of senescence and litterfall carbon inputs to the soil models (CASA-CNP, MIMICS, and CORPSE). Thus, in these carbon-only simulations, the soil models experienced identical timing and magnitude of litter inputs and soil abiotic conditions (temperature, moisture, and texture). Additional details for the simulation are described in (Wieder et al. 2018, 2019a).

Litterfall inputs provide fresh carbon substrates into litter pools that decompose into soil carbon pools in each model. CASA-CNP follows a conventional decomposition scheme that uses first-order decay of each carbon pool. In MIMICS and CORPSE, litter and soil decomposition occur when organic matter passes through one (CORPSE) or one of two (MIMICS) microbial biomass pools before (re)forming soil carbon, with rates of decomposition determined by substrate availability and the size of microbial biomass pools. Thus, MIMICS and CORPSE explicitly represent soil microbial activity and consider microbial interactions with the surrounding physicochemical soil environment. In all three models, the decomposition of organic matter follows temperature-sensitive kinetics, although the specific parametrization of each model results in distinct emergent temperature

sensitivities of SOC turnover (see Wieder et al. 2018). All three models in the testbed assume that protected C pools are either inherently resistant to decomposition (e.g., long turnover times reflecting theories about the inherent chemical recalcitrance of passive C in CASA-CNP) or have restricted access to microbial decomposers (as in the protected pools simulated by MIMICS and CORPSE). Despite this distinction in theory, all three models in the testbed also use soil texture as a proxy that mediates the persistence of this passive or protected organic matter (Bailey et al. 2019; Rasmussen et al. 2018). Soil carbon stocks simulated by each model – CASA-CNP, MIMICS, and CORPSE – total 1380, 1420, and 1720 Pg C globally (0–100 cm depth; 10-yr average from 2000 to 2010), respectively, and, broadly, have similar spatial distributions (Fig. 1; Wieder et al. 2018).

Machine learning emulators and analyses

We used statistical modeling to identify key predictors influencing SOC variability and explored a suite of approaches, including multivariate linear regressions, gradient boosting machines, and random forests (Fig. S1). We trained the models using SOC content and corresponding predictors—here, MAT, NPP, and TEX—for each data source, globally and across land cover types. The random forest (RF) models are reported here, with results from the multivariate linear regressions shown in the supplement (Fig. S2; Table S1). For the RF results, the percent variance explained (on independent test data, with a 75 – 25 train-test split; Fig. S3–S5) and variable importance scores were averaged from an ensemble of 10 random forests (400 decision trees each) with bootstrapped sampling, which were sufficient for convergence and stable model results. All RF analyses were performed using the R package randomForest (version 4.6–12) (Breiman 2001; Liaw and Wiener 2002).

Variable importance scores depict the degradation in model performance, i.e., the increase in mean squared error (MSE), following the exclusion of a given predictor from the RF model and were normalized to sum to 1. Namely, in the case of two hypothetical predictors x_j and x_k , the variable importance (VI) of predictor x_k can be written as

$$\begin{aligned}
 VI_k &= \frac{\Delta MSE_k}{\Delta MSE_k + \Delta MSE_j} \\
 &= \frac{MSE(x_j) - MSE(x_j, x_k)}{[MSE(x_j) - MSE(x_j, x_k)] + [MSE(x_k) - MSE(x_j, x_k)]}
 \end{aligned}
 \quad (1)$$

and that of predictor x_j can be written as

$$\begin{aligned}
 VI_j &= \frac{\Delta MSE_j}{\Delta MSE_k + \Delta MSE_j} \\
 &= \frac{MSE(x_k) - MSE(x_j, x_k)}{[MSE(x_j) - MSE(x_j, x_k)] + [MSE(x_k) - MSE(x_j, x_k)]}
 \end{aligned}
 \quad (2)$$

where ΔMSE_k is the increase in mean squared error when x_k is removed from the RF model compared to the model with all variables included, and analogously for ΔMSE_j with the removal of x_j . This increase in mean squared error when a given variable is removed means that, in the case of x_k for example, $\Delta MSE_k > 0$ since $MSE(x_j) > MSE(x_j, x_k)$, where $MSE(x_j)$ is the mean squared error when x_k is removed from the RF model and $MSE(x_j, x_k)$ when all variables are included. The variable importance scores then sum to 1 across predictors; that is, $VI_k + VI_j = 1$ in the example above. The importance ranking is linear, so it is preserved in the normalization. Importance scores reflect how important each predictor is in explaining the spatial variability of SOC for each data source, where a higher score signifies a greater importance (Figs. S6–S8).

Partial dependence relationships were used to explore the effect of each climatic and edaphic predictor variable on SOC stocks from each data source, while the other predictor variables were held constant at their mean value. These relationships emerged from the RF emulators, without imposing expected relationships *a priori*. For comparison, we also include results from a multivariate linear regression (Fig. S2; Table S1), which show similar qualitative relationships without the potential for emergent non-linearities. Standardized regression beta coefficients are given for the multivariate linear regressions from each data source, where the underlying data have been standardized so that the variances of dependent and independent variables are equal to 1. This allows the comparison of regression coefficients on the same scale. However, the empirical and modeled relationships are known to be non-linear for given predictors

(e.g., with temperature), and thus we focus our study on these results and urge the adoption of methods that allow non-linearities to emerge.

Biome-specific analyses were conducted on a subset of the global datasets for each data source. Using the MODIS MCD12C1 landcover product (Friedl et al. 2010) for classification, we first grouped forest (deciduous/evergreen broad/needleleaf and mixed forests; i.e., IGBP land classes 1 to 5) and herbaceous (savannas and grasslands; i.e., IGBP land classes 9 and 10) biomes for broad landcover comparisons, with the HWSO and NCSCD-adjusted gridded data products and biogeochemical model output. We then explored underlying biomes within these broad categories (Fig. S6–S7). We excluded the soil profiles from these biome-specific analyses due to low predictability (variance explained < 20 %) and higher uncertainty in landcover classification, focusing instead on the gridded data products and biogeochemical model outputs (Fig. S6–S7). We trained random forest models on each data subset to learn the role of the underlying climatic and edaphic predictors on SOC stocks, and compared the percent variance explained and variable importance for each biome subset across the data sources.

Results and Discussion

Predictability and variance explained

The predictability of SOC varies substantially between observations and process-based models and across scales. Indeed, using random forest models with key state factors—namely, MAT, NPP, and TEX—to predict SOC from different sources showed a range of variance explained, from 26 % for global soil profile measurements up to 81 % for the MIMICS model (Table 1). The soil profile measurements were taken at the plot-level and exhibited greater heterogeneity due to fine-scale controls (e.g., proximity to vegetation, topographic wetness, soil aggregation; Wiesmeier et al. 2019) that are often not included in broad-scale analyses or models. As a result, the predictability of such SOC measurements is often lower (Doetterl et al. 2015; Hengl et al. 2014, 2017)—here, only 26 % of the variance in soil profiles is explained by MAT, NPP, and TEX using a random forest model, indicating that important controls at this plot-scale are still

missing and that local conditions that drive SOC accumulation may not be well-matched with the relatively coarse spatial scales of MAT and NPP. Future studies should thus consider using finer-scale observationally-derived products of covariates, when feasible and available. Furthermore, other fine-scale controls (e.g., pH, exchangeable calcium, extractable iron and aluminum, among other properties) are important in explaining the variability in SOC across soil profiles (Heckman et al. 2020; Nave et al. 2021; Rasmussen et al. 2018), but these controls are not yet incorporated in most biogeochemical models and, therefore, could not be compared herein.

In contrast, observationally-derived data products (e.g., HWSD and NCSCD-adj. HWSD) and process-based biogeochemical models have imposed underlying structures that rely on fewer input variables and are more easily learned by a random forest algorithm. Indeed, these gridded observational and model products exhibit less fine-scale heterogeneity in SOC stocks than that seen in soil profile measurements. Using a random forest model for each of the observational and biogeochemical model products globally, the same three covariates – MAT, NPP, and TEX – explained 58 and 53 % of the variance in the HWSD and NCSCD-adj. HWSD versus 63 %, 81 %, and 53 % in the CASA-CNP, MIMICS, and CORPSE model outputs, respectively (Table 1). The predictability of SOC also varied between individual biomes and broad biome classifications (Fig. S4–S5). For example, temperate deciduous broadleaf and evergreen needleleaf forests showed greater predictability than tropical evergreen broadleaf forests, in both observational data products and process-based models (Fig. S5).

While we focused our analyses on the more parsimonious RF model that included MAT, NPP, and TEX, we briefly note that adding mean annual precipitation (MAP) into the random forest models only marginally increased the percent variance explained by 1–5 % across the different data sources globally (Table S2). The smallest difference was observed for MIMICS, agreeing with the fact that the version of the model used in the biogeochemical testbed did not contain a soil moisture function and thus did not use MAP as an input (Wieder et al. 2018). Therefore, subsequent analyses focused on the environmental variables (MAT, NPP, TEX) that were direct inputs to all of the biogeochemical models.

Variable importance

In addition to overall SOC predictability across data and models, we also explored the variable importance and emergent relationships for MAT, NPP, and TEX for each of the SOC sources. We found an apparent mismatch between observations and biogeochemical models, where TEX was the most important predictor for the observations (both soil profiles and global datasets), in contrast to MAT and NPP for the model outputs (Fig. 2). This result also holds for clay as a predictor, as often used in biogeochemical models (Fig. S8; though only 48 % of the variance in HWSD was explained). Furthermore, we note that the importance of MAT was higher in the NCSCD-adj. HWSD compared to the HWSD alone, suggesting potential differences in underlying temperature sensitivity and freeze-thaw dynamics that are important to consider when benchmarking biogeochemical models globally.

While biogeochemical models placed a higher importance on both temperature and plant productivity globally (Fig. 2), greater nuance exists at the biome-level. In our RF emulators of the biogeochemical model outputs, MAT stood out as the most important variable in forests, whereas NPP was most important in herbaceous biomes (Fig. 3). In contrast, TEX was again the most important variable for HWSD in both forest and herbaceous biomes (Fig. 3; Fig. S6), though the NCSCD-adj. HWSD showed an increased importance of MAT for herbaceous biomes at high-latitudes (Fig. 3; Fig. S6). The RF models performance in broad forest and herbaceous land classes were comparable to that of the global results (Fig. S4). Interestingly, among individual biomes where the RF also achieved similar performance to the global results (Fig. S5), we find agreement between the observations and model outputs in select biomes (Fig. S7). Namely, MAT had high explanatory power in temperate deciduous broadleaf and mixed forests and NPP had high explanatory power in temperate grasslands. This suggests that biogeochemical models can match observations in select biomes (e.g., temperate forests and grasslands), but targeted model improvements are needed in biomes where observations and models show divergent controls on SOC stocks (Fig. S6–S7). These targeted improvements could include a closer examination of model parameterizations in specific biomes (for example, increasing mineralogical controls or increasing/decreasing temperature

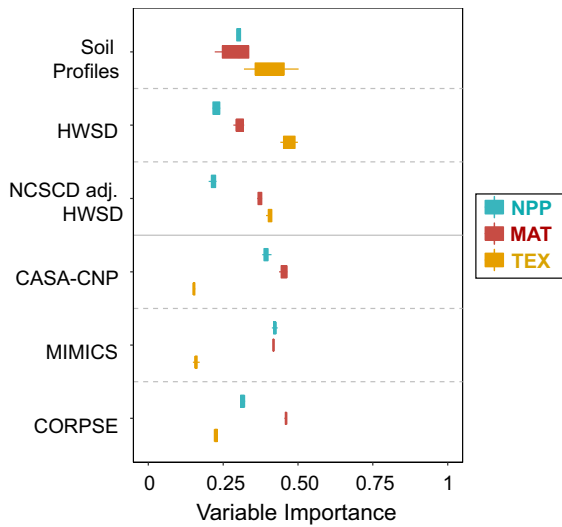


Fig. 2 Variable importance for soil organic carbon (SOC) stocks globally, using Random Forests (RFs) that learned the SOC from observations and biogeochemical models with net primary productivity (NPP), mean annual temperature (MAT), and soil texture (TEX). Results shown are averaged over an ensemble of RFs with bootstrapped sampling, and box plots depict the 95 % confidence intervals (error bars) and first and third quartiles (boxes). Variable importance equal to 1 indicates the highest importance

dependencies) and the distribution of SOC among different model pools. Modifications may also require incorporating missing controls into biome-specific model formulations or conducting additional

experiments and sampling campaigns in data-poor regions to further inform parameterizations.

Predictive relationships

Mineralogical controls

Using random forest models trained on the observations and model output, we explored the dependence of measured and modeled SOC on individual covariates. These random forest models act as emulators of the observations and biogeochemical models, allowing us to explore individual relationships (i.e. partial dependence plots) while controlling for all other covariates. This method also allows for emergent non-linearities without *a priori* imposed relationships.

We observed stark differences between the observations and biogeochemical models (Fig. 4). Both the soil profiles and observational data products (HWSD and NCSCD-adj. HWSD) showed a greater dependence (steeper positive slope) on TEX (clay and silt content) compared to the models (near zero or negative slope; Fig. 4a; Fig. S2; Table S1). By contrast, the soil biogeochemical models were more sensitive to MAT and NPP, compared to the observational products. Though the soil biogeochemical models do contain some representation of mineral-organic associations (where MIMICS and CORPSE do so in a more explicit, mechanistic way), the

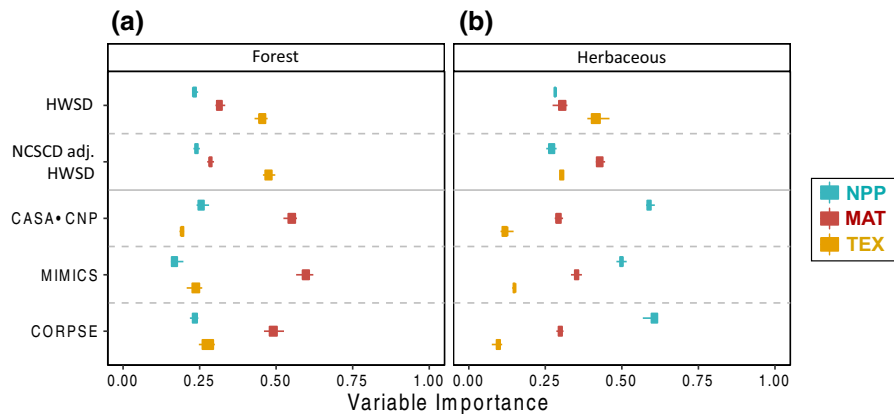


Fig. 3 Variable importance for soil organic carbon (SOC) stocks in **a** forest and **b** herbaceous biomes, using Random Forests (RFs) that learned the SOC from observations and biogeochemical models with net primary productivity (NPP), mean annual temperature (MAT), and soil texture (TEX). Forests included deciduous/evergreen broadleaf/needleleaf and

mixed forests, and herbaceous biomes included grasslands and savannas. Results shown are averaged over an ensemble of RFs with bootstrapped sampling, and box plots depict the 95 % confidence intervals (error bars) and first and third quartiles (boxes). Variable importance equal to 1 indicates the highest importance

importance of TEX appeared insignificant in explaining SOC content globally. Rather, climate and vegetation seemed to play a predominant role in driving the distribution of SOC in soil biogeochemical models. This discrepancy motivates a closer look at how mineralogical controls are implemented in each biogeochemical model and the resulting distributions of SOC stocks between modeled pools.

Temperature sensitivity

Both observations and models showed the highest SOC in grid cells with colder MAT (Fig. 4b). Indeed, SOC had the highest temperature sensitivity (steeper slope) in colder regions ($\text{MAT} < 0\text{ }^\circ\text{C}$) and a lower temperature sensitivity (shallow slope) in warmer regions ($\text{MAT} > 10\text{ }^\circ\text{C}$) (Fig. 4b), corroborating recent studies (Koven et al. 2017; Wieder et al. 2019a). This higher temperature sensitivity of SOC at low MAT primarily occurs in high-latitude forests, and therefore, a clear difference in temperature sensitivity was seen across herbaceous (here grasslands and savannas) and forest biomes (Fig. 3; Fig. S6). Without the NCSCD-adjusted values at high latitudes, the HWSO data product on its own showed a muted temperature sensitivity, far lower than that of the soil profiles and biogeochemical models (Fig. 4b), suggesting that the HWSO alone may not fully capture the underlying temperature sensitivity that is needed to

benchmark biogeochemical models. Replacing the high latitudes with the NCSCD data product, known to be more accurate in those geographies (Hugelius et al. 2013; Koven et al. 2017), improved the emergent temperature sensitivity such that it approached that of the soil profiles.

While the soil biogeochemical models use similar temperature functions, their parameterizations and which fluxes or carbon pools they are applied to differ between the models (Wieder et al. 2018). As a result, the biogeochemical models exhibited a range of temperature sensitivities (Fig. 4b). The degree of temperature sensitivity of the soil profiles most closely matched the MIMICS and CASA-CNP models (Fig. 4b). CORPSE had the highest MAT dependence (Fig. 4b), which may be due in part to its ceased decomposition in frozen soils (Wieder et al. 2019a), and is reflected by its higher SOC content at high latitudes (Fig. 1).

Influence of plant productivity

Plant inputs are the main source of carbon to the soil and, thus, plant productivity across biomes plays an important role in carbon accumulation. We found that, in both the observations and soil biogeochemical models, SOC content exhibited a saturating relationship with increasing net primary productivity (NPP; Fig. 4c). However, the soil profiles and the

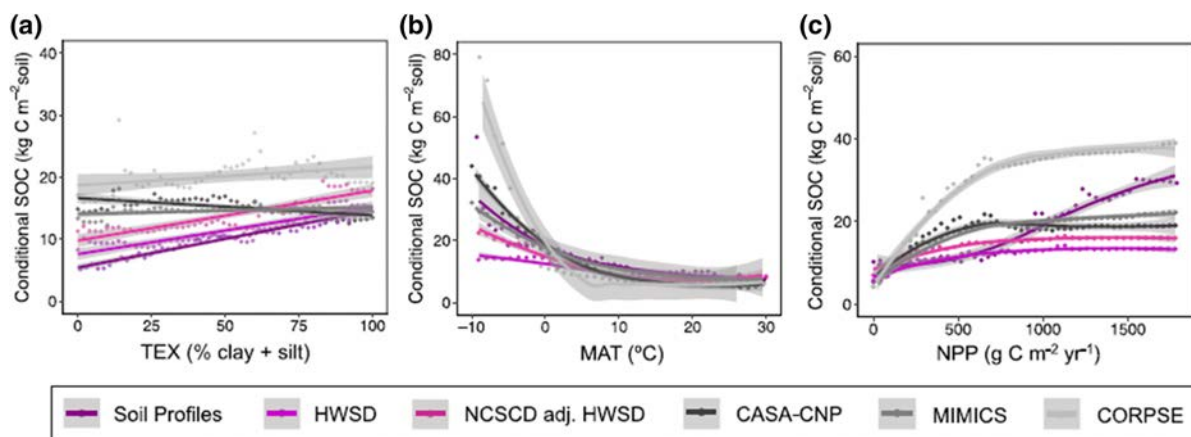


Fig. 4 Partial dependence relationships of soil organic carbon (SOC) as a function of individual controlling factors, namely **a** soil texture (TEX), **b** mean annual temperature (MAT), and **c** net primary productivity (NPP). Machine-learning (Random Forest; RF) derived global relationships of individual covariates

conditional on all other covariates being held constant at their mean. RFs were trained on each observational dataset and model output separately. Points show RF predictions; best fit lines show 99 % confidence intervals on the relationship

biogeochemical models showed a significantly higher sensitivity (steeper increase) to NPP than the HWSO data product. This suggests that underlying environmental sensitivities seen in the soil profiles may not be effectively represented in the HWSO data product, and thus, care should be taken when using such data products to benchmark soil biogeochemical models. Indeed, the soil profiles showed a NPP sensitivity that fell in the middle of those seen in the models; specifically, at high NPP, the soil profiles showed a lower sensitivity than CORPSE and higher sensitivity than MIMICS and CASA-CNP, while at low NPP, the soil profiles were less sensitive than all three of the models (Fig. 4c). This qualitative difference in the shape of the soil profile curve is interesting, and further investigation with other datasets (especially those with site-level NPP measurements) is warranted for verification – especially given the low variance explained for the soil profiles – and to understand the root cause and potential implications for model representations. It is important to note that gridded NPP (and MAT) estimates may differ from the actual conditions experienced by the soil profiles being measured, as NPP can be highly variable at the site-level within a grid cell. Elucidating the sensitivity of long-term SOC storage to changes in plant inputs, and to underlying site-level heterogeneity, is critical for understanding potential feedbacks to land-use and land-cover change globally.

Conclusions and implications

Our results suggest a potential mismatch between observations and biogeochemical models in the emergent role of environmental controls in explaining SOC variability. Namely, MAT and NPP have the most explanatory power for SOC stocks from the biogeochemical models – CASA-CNP, MIMICS, and CORPSE – and TEX has the least explanatory power. In contrast, TEX has the most explanatory power for SOC stocks among the observations, both at the soil profile scale and in the global observationally-derived data products (Fig. 2; Fig. S8). This mismatch motivates a closer look at how mineralogical controls modulate SOC persistence and microbial function in biogeochemical models, and their resulting distribution of mineral-associated organic carbon, which

could improve projections of not only SOC stocks, but also SOC ages and turnover times.

Furthermore, the covariate relationships observed in soil profiles and observationally-derived gridded data products were not always aligned (e.g., Fig. 4b-c), indicating that gridded data products may not effectively capture the underlying climatic and edaphic controls of the measured soil profiles due in part to upscaling uncertainty. The NCSCD-adj. HWSO, however, was more closely aligned with the soil profiles than was the HWSO alone, supporting recent studies that preferentially use it for model comparisons (Koven et al. 2017). Other gridded observational products (Hengl et al. 2017) can also be evaluated in this way to understand their effective representation of emergent environmental controls. This is an important step in validating observational data products for future use as model benchmarks.

The overall predictability of SOC stocks from soil profiles and observational data products was generally lower than that of the biogeochemical models, globally and within select biomes (Table 1; Figs. S4-S5). While other predictors are certainly important for explaining the variability of SOC in soil profiles and observational data products (Delgado-Baquerizo et al. 2017; Doetterl et al. 2015; Rasmussen et al. 2018), these predictors are often not direct inputs into biogeochemical model simulations and therefore cannot be compared using the random forest emulator approach presented here. However, further exploration of the underlying, biome-specific and global controls driving SOC variability in observations (e.g., metal oxides, nutrients, fauna, topography; also MAP in Table S2) (Crowther et al. 2019; Doetterl et al. 2015; Rasmussen et al. 2018) is essential for motivating their incorporation in future process-based formulations of soil biogeochemical models.

Acknowledgements K.G. was supported by a USDA NIFA Postdoctoral Fellowship and as a Lawrence Fellow at Lawrence Livermore National Laboratory (LLNL). Work at LLNL was conducted under the auspices of DOE Contract DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project No. 21-ERD-045. This study stems from a synthesis group *Advancing Soil Organic Matter Research: Synthesizing Multi-scale Observations* supported through the Long-Term Ecological Research Network Office (LNO; NSF award numbers 1545288 and 1929393) and the National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California, Santa Barbara, and lead by W.R.W. and K.L. W.R.W. was supported by the Niwot Ridge LTER

program (NSF DEB – 1637686) and USDA-NIFA (2020-67019-31395). J.A.M.M. was supported by Postdoctoral Development funds at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

Author contributions All co-authors contributed to the study conceptualization. K.G., A.M., and R.B.J. designed the implementation. K.G. and W.R.W. contributed data and W.R.W., M.D.H., and B.N.S. contributed model output. K.G. and J.H.E. performed statistical and model analyses. K.G. wrote the first draft and all authors contributed to the manuscript.

Data Availability All data used in this study are freely available. The gridded global maps can be found from their corresponding references, listed in the Materials and Methods. The global profile measurements are available from ISRIC WoSIS. A compiled dataset of observational and modeled soil organic carbon and key covariates, as used in this study, and all code are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare no conflict of interest.

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

References

- Abramoff R, Xu X, Hartman M, O'Brien S, Feng W, Davidson E et al (2018) The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry* 137(1–2):51–71. <https://doi.org/10.1007/s10533-017-0409-7>
- Abramoff RZ, Torn MS, Georgiou K, Tang J, Riley WJ (2019) Soil Organic Matter Temperature Sensitivity Cannot be Directly Inferred From Spatial Gradients. *Global Biogeochem Cycles* 33(6):761–776. <https://doi.org/10.1029/2018GB006001>
- Ahlström A, Schurgers G, Smith B (2017) The large influence of climate model bias on terrestrial carbon cycle simulations. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/12/1/014004>
- Allison SD, Wallenstein MD, Bradford MA (2010) Soil-carbon response to warming dependent on microbial physiology. *Nat Geosci* 3(5):336–340. <https://doi.org/10.1038/ngeo846>
- Bailey VL, Pries CH, Lajtha K (2019) What do we know about soil carbon destabilization? *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/ab2c11>
- Batjes NH (2009) Harmonized soil profile data for applications at global and continental scales: Updates to the WISE database. *Soil Use Manag* 25(2):124–127. <https://doi.org/10.1111/j.1475-2743.2009.00202.x>
- Batjes NH (2016) Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* 269:61–68. <https://doi.org/10.1016/j.geoderma.2016.01.034>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. https://doi.org/10.1007/978-3-662-56776-0_10
- Cotrufo MF, Wallenstein MD, Boot CM, Deneff 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? *Glob Change Biol* 19(4):988–995. <https://doi.org/10.1111/gcb.12113>
- Crowther TW, van den Hoogen J, Wan J, Mayes MA, Keiser AD, Mo L et al (2019) The global soil community and its influence on biogeochemistry. *Science*. <https://doi.org/10.1126/science.aav0550>
- Delgado-Baquerizo M, Eldridge DJ, Maestre FT, Karunarathne SB, Trivedi P, Reich PB, Singh BK (2017) Climate legacies drive global soil carbon stocks in terrestrial ecosystems. *Sci Adv* 3(e1602008 12):1–7. <https://doi.org/10.1126/sciadv.1701482>
- Doetterl S, Stevens A, Six J, Merckx R, Van Oost K, Pinto C, M., et al (2015) Soil carbon storage controlled by interactions between geochemistry and climate. *Nat Geosci* 8(10):780–783. <https://doi.org/10.1038/ngeo2516>
- Dwivedi D, Tang J, Bouskill N, Georgiou K, Chacon SS, Riley WJ (2019) Abiotic and biotic controls on soil organo – mineral interactions: developing model structures to analyze why soil organic matter persists. *Rev Mineral Geochem* 85:329–348
- Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ* 114(1):168–182. <https://doi.org/10.1016/j.rse.2009.08.016>
- Friedlingstein P, Meinshausen M, Arora VK, Jones CD, Anav A, Liddicoat SK, Knutti R (2014) Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *J Clim* 27:511–526. <https://doi.org/10.1175/JCLI-D-12-00579.1>
- Georgiou K, Abramoff RZ, Harte J, Riley WJ, Torn MS (2017) Microbial community-level regulation explains soil carbon

- responses to long-term litter manipulations. *Nat Commun* 8(1):1–10. <https://doi.org/10.1038/s41467-017-01116-z>
- Grandy AS, Neff JC (2008) Molecular C dynamics downstream: The biochemical decomposition sequence and its impact on soil organic matter structure and function. *Sci Total Environ* 404(2–3):297–307. <https://doi.org/10.1016/j.scitotenv.2007.11.013>
- Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int J Climatol* 34(3):623–642. <https://doi.org/10.1002/joc.3711>
- Heckman KA, Nave LE, Bowman M, Gallo A, Hatten JA, Matosziuk LM et al (2020) Divergent controls on carbon concentration and persistence between forests and grasslands of the conterminous US. *Biogeochemistry*, 0123456789. <https://doi.org/10.1007/s10533-020-00725-z>
- Hengl T, De Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E et al (2014) SoilGrids1km - Global soil information based on automated mapping. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0105992>
- Hengl T, De Jesus M, Heuvelink J, Gonzalez GBM, Kilibarda MR, Blagotić M, A., et al (2017) *SoilGrids250m: Global gridded soil information based on machine learning*. *PLoS One*. <https://doi.org/10.1371/journal.pone.0169748>
- Hugelius G, Tarnocai C, Broll G, Canadell JG, Kuhry P, Swanson DK (2013) The northern circumpolar soil carbon database: Spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth System Science Data* 5(1):3–13. <https://doi.org/10.5194/essd-5-3-2013>
- Jackson RB, Caldwell MM (1993) Geostatistical patterns of soil heterogeneity around individual perennial plants. *J Ecol* 81(81):683–692. <https://doi.org/10.2307/2261666>
- Jackson RB, Lajtha K, Crow SE, Hugelius G, Kramer MG, Piñeiro G (2017) The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls. *Annu Rev Ecol Evol Syst* 48(1):419–445. <https://doi.org/10.1146/annurev-ecolsys-112414-054234>
- Koven CD, Hugelius G, Lawrence DM, Wieder WR (2017) Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nature Climate Change* 7(11):817–822. <https://doi.org/10.1038/nclimate3421>
- Lehmann J, Hansel CM, Kaiser C, Kleber M, Maher K, Manzoni S et al (2020) Persistence of soil organic carbon caused by functional complexity. *Nat Geosci* 13(8):529–534. <https://doi.org/10.1038/s41561-020-0612-3>
- Li J, Wang G, Allison SD, Mayes MA, Luo Y (2014) Soil carbon sensitivity to temperature and carbon use efficiency compared across microbial-ecosystem models of varying complexity. *Biogeochemistry* 119(1–3):67–84. <https://doi.org/10.1007/s10533-013-9948-8>
- Liaw A, Wiener M (2002) Classification and Regression by randomForest. *R News* 2(3):18–22
- Nave LE, Bowman M, Gallo A, Hatten JA, Heckman KA, Matosziuk L et al (2021) Patterns and predictors of soil organic carbon storage across a continental-scale network. *Biogeochemistry*, 9. <https://doi.org/10.1007/s10533-020-00745-9>
- Oleson KW, Lawrence DM, Bonan GB, Drewniak B, Huang M, Charles D et al (2013) Technical description of version 4.5 of the Community Land Model (CLM) (No. NCAR/TN-503 + STR). <https://doi.org/10.5065/D6RR1W7M>
- Potter CS, Randerson JT, Field CB, Matson PA, Vitousek PM, Mooney HA, Klooster SA (1993) Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochem Cycles* 7(4):811–841. <https://doi.org/10.1029/93GB02725>
- Randerson JT, Thompson MV, Malmstrom CM, Field CB, Fung YI (1996) Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO₂. *Global Biogeochem Cycles* 10(4):585–602
- Rasmussen C, Heckman K, Wieder WR, Keiluweit M, Lawrence CR, Asefaw A et al (2018) Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* 137:297–306. <https://doi.org/10.1007/s10533-018-0424-3>
- Robertson AD, Paustian K, Ogle S, Wallenstein MD, Lugato E, Cotrufo F, M (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>
- Schmidt MWI, Torn MS, Abiven S, Dittmar T, Guggenberger G, Janssens IA et al (2011) Persistence of soil organic matter as an ecosystem property. *Nature* 478:49–56. <https://doi.org/10.1038/nature10386>
- Sollins P, Homann P, Caldwell BA (1996) Stabilization and destabilization of soil organic matter: mechanisms and controls. *Geoderma* 74:65–105
- Sulman BN, Phillips RP, Oishi aC, Shevliakova E, Pacala SW (2014) Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO₂. *Nature Climate Change* 4(December):1099–1102. <https://doi.org/10.1038/nclimate2436>
- Sulman BN, Moore JAM, Abramoff R, Averill C, Kivlin S, Georgiou K et al (2018) Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry* 141(2):109–123. <https://doi.org/10.1007/s10533-018-0509-z>
- Todd-Brown KEO, Randerson JT, Hopkins F, Arora V, Hajima T, Jones C et al (2014) Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* 11(8):2341–2356. <https://doi.org/10.5194/bg-11-2341-2014>
- Wang YP, Law RM, 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>
- Wang G, Post WM, Mayes Ma (2013) Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. *Ecol Appl* 23(1):255–272. <https://doi.org/10.1890/12-0681.1>
- Wieder WR, Grandy AS, Kallenbach CM, Bonan GB (2014a) Integrating microbial physiology and physio-chemical principles in soils with the Microbial-MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences* 11(14):3899–3917. <https://doi.org/10.5194/bg-11-3899-2014>
- Wieder WR, Boehnert J, Bonan GB, Langseth M (2014b) RegridDED Harmonized World Soil Database v1.2. *Data Set. Available on-Line [Http://Daac.Ornl.Gov] from Oak Ridge National Laboratory Distributed Active Archive*

- Center, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1247>
- Wieder WR, Allison SD, Davidson EA, Georgiou K, Hararuk O (2015a) Explicitly representing soil microbial processes in Earth system models. *Global Biogeochem Cycles* 29:1782–1800
- Wieder WR, Grandy AS, Kallenbach CM, Taylor PG, Bonan GB (2015b) Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geosci Model Dev* 8(6):1789–1808. <https://doi.org/10.5194/gmd-8-1789-2015>
- Wieder WR, Hartman MD, Sulman BN, Wang YP, Koven CD, Bonan GB (2018) Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Glob Change Biol* 24(4):1563–1579. <https://doi.org/10.1111/gcb.13979>
- Wieder WR, Sulman BN, Hartman MD, Koven CD, Bradford MA (2019a) Arctic Soil Governs Whether Climate Change Drives Global Losses or Gains in Soil Carbon. *Geophys Res Lett* 46(24):14486–14495. <https://doi.org/10.1029/2019GL085543>
- Wieder WR, Lawrence DM, Fisher RA, Bonan GB, Cheng SJ, Goodale CL et al (2019b) Beyond Static Benchmarking: Using Experimental Manipulations to Evaluate Land Model Assumptions. *Global Biogeochem Cycles* 33(10):1289–1309. <https://doi.org/10.1029/2018GB006141>
- Wiesmeier M, Urbanski L, Hobbey E, Lang B, von Lützow M, Marin-Spiotta E et al (2019) Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. *Geoderma*, 333: 149–162. <https://doi.org/10.1016/j.geoderma.2018.07.026>
- Zhao M, Heinsch FA, Nemani RR, Running SW (2005) Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens Environ* 95(2):164–176. <https://doi.org/10.1016/j.rse.2004.12.011>

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