scientific reports

Check for updates

OPEN

Optimizing process-based models to predict current and future soil organic carbon stocks at high-resolution

Derek Pierson[1]✉, Kathleen A. Lohse[1,2], William R. Wieder[3,4], Nicholas R. Patton[2,5], Jeremy Facer[1], Marie-Anne de Graaff[6], Katerina Georgiou[7], Mark S. Seyfried[8], Gerald Flerchinger[8] & Ryan Will[9]

From hillslope to small catchment scales (< 50 km^2), soil carbon management and mitigation policies rely on estimates and projections of soil organic carbon (SOC) stocks. Here we apply a process-based modeling approach that parameterizes the MIcrobial-MIneral Carbon Stabilization (MIMICS) model with SOC measurements and remotely sensed environmental data from the Reynolds Creek Experimental Watershed in SW Idaho, USA. Calibrating model parameters reduced error between simulated and observed SOC stocks by 25%, relative to the initial parameter estimates and better captured local gradients in climate and productivity. The calibrated parameter ensemble was used to produce spatially continuous, high-resolution (10 m^2) estimates of stocks and associated uncertainties of litter, microbial biomass, particulate, and protected SOC pools across the complex landscape. Subsequent projections of SOC response to idealized environmental disturbances illustrate the spatial complexity of potential SOC vulnerabilities across the watershed. Parametric uncertainty generated physicochemically protected soil C stocks that varied by a mean factor of 4.4 × across individual locations in the watershed and a − 14.9 to + 20.4% range in potential SOC stock response to idealized disturbances, illustrating the need for additional measurements of soil carbon fractions and their turnover time to improve confidence in the MIMICS simulations of SOC dynamics.

A large portion of the world's actively cycled carbon (C) is stored in surface soils[1,2]. With changes in global climate and land use accelerating at unprecedented rates[3–5], there is an emergent need to project soil organic carbon (SOC) dynamics under changing environmental conditions. A growing body of work focuses on potential global-scale SOC responses to climate change[6–8], but simulating soil carbon dynamics at more localized scales (hillslope and small watershed catchment, < 50 km^2) offers opportunities to parameterize process-based models at resolutions that have a greater relevance to policy and management[9,10]. This more local approach also better connects models and field observations, allowing stronger inference of the proximal controls over SOC persistence and turnover that ultimately builds confidence in model projections[11,12]. Such advances, however, require quantitative tools that can be used to link observations and models to calibrate model parameters, assess parametric uncertainty, and generate high resolution estimates of soil carbon stocks and potential vulnerabilities.

Current methodologies for estimating soil carbon fall into two main categories, digital soil mapping and process-based modeling. Digital soil mapping uses statistical correlations between climate, vegetation, topography, soil properties, and soil C stocks to generate high resolution estimates of SOC stocks[13–16]. While often accurate for predicting SOC stocks at fine scales[17], digital mapping methods generally have limited ability to project future SOC dynamics and responses since the statistical estimation of SOC does not adequately derive

[1]Department of Biological Sciences, Idaho State University, Pocatello, ID, USA. [2]Department of Geosciences, Idaho State University, Pocatello, ID, USA. [3]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA. [4]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA. [5]School of Earth and Environment, University of Canterbury, Christchurch, New Zealand. [6]Department of Biological Sciences, Boise State University, Boise, ID, USA. [7]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA. [8]Agricultural Research Service, Northwest Watershed Research Center, Boise, ID, USA. [9]Department of Geosciences, Boise State University, Boise, ID, USA. ✉email: derekpierson@isu.edu

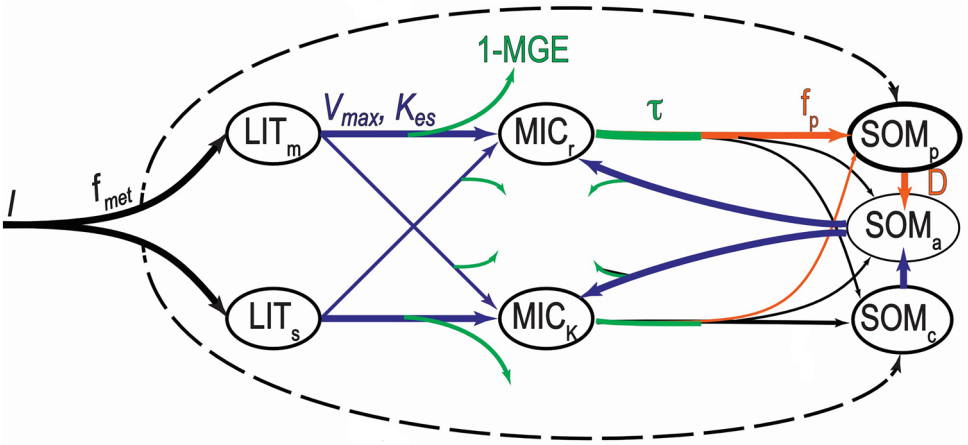

Figure 1. Diagram of the MIcrobial-MIneral Carbon Stabilization (MIMICS) model, which explicitly considers microbial functional diversity by simulating two functional groups (MIC_r, inefficient, fast-growers; MIC_K, conservative, slow-growers) and their potential effects on litter decomposition and soil organic matter persistence (available SOM_a; chemically protected SOM_c; physically protected SOM_p). Litter C inputs are initially partitioned (f_{met}) between litter pools (structural LIT_s; metabolic LIT_m), while a lesser fraction (f_i) transfers directly to the SOM pools. Parameters used for model calibration and validation related to microbial catabolic capacity (V_{max} and K_{es}, blue lines), microbial anabolism (MGE and τ, green lines), and physicochemical protection of SOM (f_p and D, orange lines). See also Table 1.

how biogeochemical interactions and environmental conditions combine to govern SOC stocks[18]. In contrast, process-based models have been extensively relied upon for projecting regional to global scale SOC stocks and their vulnerabilities to changing conditions[19–22]. Process-based models mathematically approximate the processes through which soil C enters the soil, decomposes, and then is either stored in soil or cycled back to the atmosphere. The use of process-based models for projecting SOC dynamics at high spatial resolution (e.g., < 250 m^2) is predicated by the availability of accurate, high-resolution forcing data for model calibration and the extrapolation of projections across complex landscapes. Rapid advances in monitoring technology, remote sensing, and machine learning present novel opportunities to generate such environmental data[23]. In addition to these technical challenges, major uncertainties also exist regarding the viability of using model parameters that were calibrated for regional- to global-scale simulations for projections at much smaller, hillslope- to watershed-scales.

Controls on SOC stocks are currently thought to change with the scale of observation and inference[24,25]. Global-scale simulations typically operate at a coarse spatial resolution (e.g. > 50 km^2). At these scales, process-based models of SOC have produced accurate estimates of SOC stocks based predominantly on the influences of soil temperature and moisture, net primary productivity and soil texture[26–29]. At smaller scales, additional environmental controls are often inferred to have more pronounced effects on SOC stocks across complex landscapes, including controls imposed by detrital input quality[30], microbial community composition[31] and physical factors such as landscape position[16] and erosion[32,33]. These additional influences on SOC dynamics often drive local heterogeneity in SOC stocks. Indeed, the importance of including such local biogeochemical controls in process model representations continues to be demonstrated and theorized[22,34], leading many modern process-based models to incorporate components relating to litter chemistry and microbial processing[26,27,35]. These advances in the process-based SOC model constructs are likely to improve the potential for fine scale process model applications.

Here we develop a method for model parameter optimization and high-resolution mapping of simulated SOC stocks across complex landscapes at spatial scales consistent with field observations and land management practices. This work combines the MIcrobial-MIneral Carbon Stabilization model (MIMICS; Fig. 1)[22,26] with high-resolution input data, soil carbon measurements, and a Monte Carlo approach for parameter optimization and associated uncertainty analysis. We specifically explore the impact of parametric uncertainty on model projections of SOC stocks and their potential sensitivities across a complex landscape with strong climate gradients and diverse soil properties at the Reynolds Creek Experimental Watershed and Critical Zone Observatory (RCEW-CZO)[36]. This work demonstrates how a variety of measurements collected across environmental gradients at the pedon and watershed scale improves our understanding and prediction of soil carbon storage and processes at management relevant scales. Ultimately, results from this study advance the general ability to model, map and project the sensitivity of SOC stocks at fine spatial scales, thus providing an important pathway for scientists, land managers and policymakers to gain more detailed spatial information about the existing size and future stability of SOC stocks across complex landscapes.

Results

Parameter estimation and uncertainty quantification. Initial simulations of the MIMICS model using high-resolution forcing data (Fig. SI1) and the parameters provided by Wieder et al.[26] (i.e., priors) resulted in simulated SOC stocks that showed biases relative to field observations across the Reynolds Creek Critical

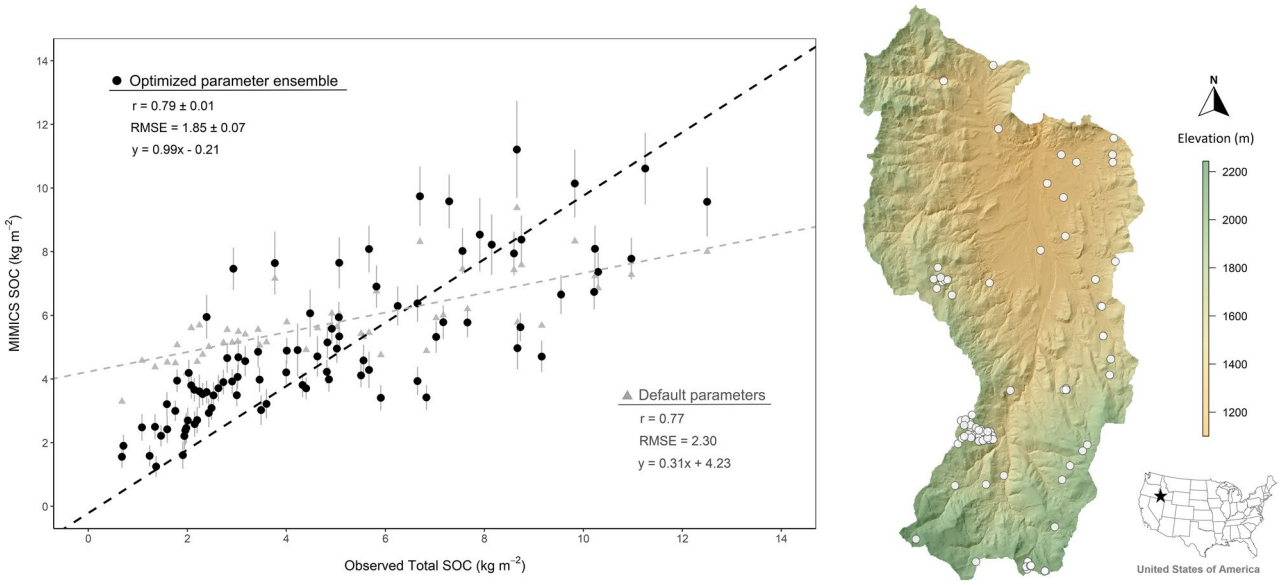

Figure 2. Comparison of the MIMICS model estimates with observed soil organic carbon stocks (kg C m^2, 0–30 cm soil depth) at locations across the Reynolds Creek Experimental Watershed and Critical Zone Observatory (n = 89). Black dots and associated error bars denote the mean ± standard deviation of the model estimates produced using each member of the optimized parameter ensemble. Grey triangles represent MIMICS estimates using the default MIMICS model parameterization. (Generated by free software R, https://www.R-project.org/).

Zone Observatory (Fig. 2). The initial default parameters underpredicted higher SOC stocks and overpredicted lower SOC stocks (Fig. 2; r = 0.77, RMSE = 2.30 kgC m^{-2}). Optimizing the model parameters with a Markov Chain Monte Carlo (MCMC) based method revealed that many different parameter combinations were capable of producing similarly accurate estimates of SOC stocks across the Reynolds Creek Critical Zone Observatory (r = 0.77–0.82; root mean square error (RMSE) = 1.7–2.0 kgC m^{-2}; Examples of differing MCMC solutions are shown in Fig. SI2). Thus, we generated an ensemble of unique parameter combinations where each parameterization yielded a RMSE < 2 kgC m^{-2} between the model estimates and field observations of SOC stocks across the Reynolds Creek Critical Zone Observatory (Fig. 2; n = 30 unique parameter combinations). Overall, the ensemble of optimized parameter combinations for MIMICS improved soil carbon predictions by eliminating bias observed when using the priors (Fig. 2) and reducing the estimate RMSE by as much as 0.58 kgC m^{-2} (25%).

The parameter space covered by our ensemble produced near equivalent estimates for total SOC stocks across the RCEW-CZO (see histograms in Fig. 3). The calibrated model consistently found parameterizations that show higher temperature sensitivity of microbial catabolic capacity (V_{slope}, V_{int}, K_{slope}, K_{int}), relative to the initial model parameterizations used by Wieder et al.[26]. Similarly, parameters related to physicochemical protection of soil C (f_p and D) are lower than the initial model parameterizations, reflecting the longer turnover time of physicochemically protected organic matter (SOM_p) with our optimized parameter values. Finally, the parameter ensemble generated a normal distribution for parameter estimates of MGE and log-normal distributions for microbial turnover rates, however the 'best' parameter estimates for microbial anabolism were not dramatically different from initial parameter values.

Our parameter ensemble also revealed strong autocorrelations for particular parameter combinations in MIMICS. For example, the decomposition rate intercept terms in the model (V_{int} and K_{int}) were positively correlated (r = 0.94, see scatter plots on Fig. 3). We also observed correlations between the parameters controlling the fate of carbon through microbial anabolism with each other and terms associated with the persistence of soil C in the physicochemical protected pool (SOM_p, Fig. 1). Specifically, we observed that parameter estimates for microbial growth efficiency (*MGE*) were positively correlated with microbial turnover rates (τ, r = 0.74; Fig. 3) and negatively correlated with the fraction of SOM partitioned to the protected SOM pool (f_p, r = -0.59). Further, f_p was positively correlated with desorption (D) of SOM_p to the available SOM pool (r = 0.58). No correlations (r > 0.4) were observed between the decomposition rate parameters associated with microbial catabolic capacity and those related to microbial anabolism or persistence of soil C in physicochemical protected pools.

While all parameterizations in the parameter ensemble resulted in similarly robust estimates of total SOC stocks (RMSE < 2.0 kg C m^{-2}; Fig. 4A), the parametric uncertainty across the ensemble (i.e. the range of parameter combinations) led to greater variability in the model estimates of underlying litter, microbial biomass, and soil C pools than for estimates of total SOC stocks (Fig. 4). Across the parameter ensemble, the protected soil carbon stocks (SOM_p) simulated by MIMICS varied by a factor of 4.4 ± 2.0 at individual locations (mean ± 1σ; Fig. 4B). This variability in SOM_p pool sizes can in part be explained by uncertainty in the parameterization of the pool's turnover time, which ranged from 75 to over 400 years (Fig. 4D). Similarly, across individual locations, litter C pool estimates varied by a factor of 6.9 ± 3.3 and the proportion of total SOC made up of microbial biomass

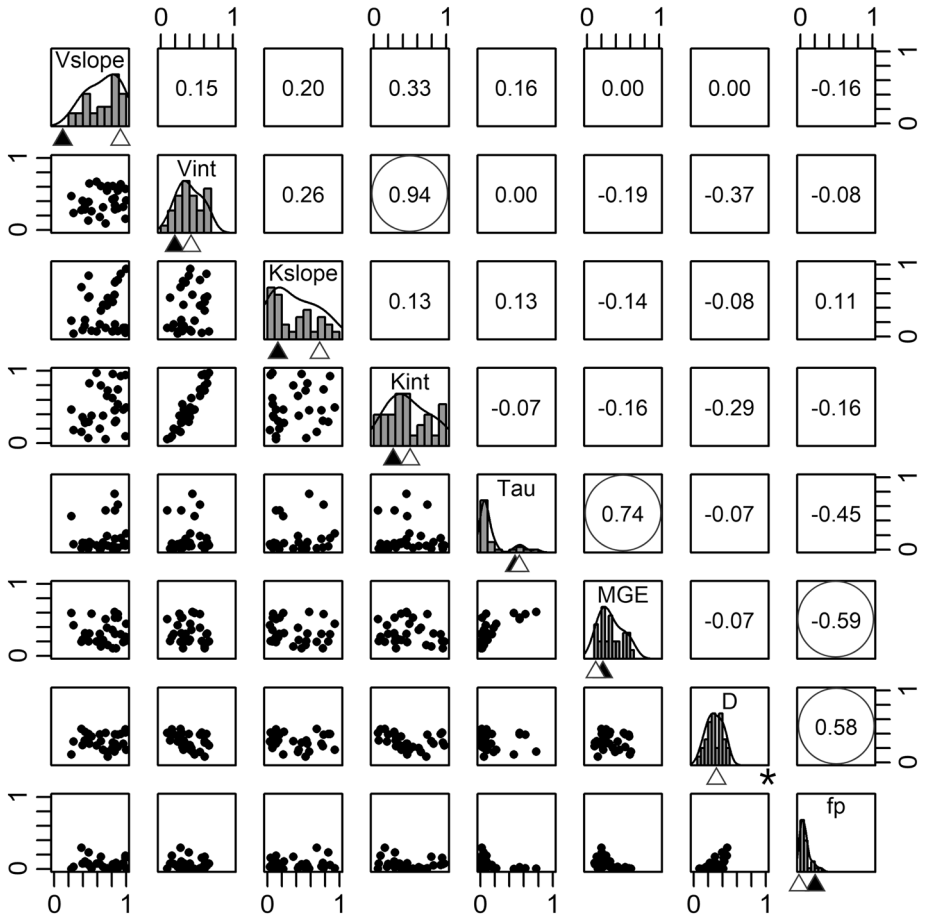

Figure 3. MIMICS model parameter space and relationships among parameters in the parameter ensemble that produced similarly accurate (RMSE 1.8–2.0) estimates of total soil organic C stocks across the Reynolds Creek Experimental Watershed and Critical Zone Observatory (n = 30 member parameter ensemble). Normalized parameter range coincides with the scaling factor proposal range for each parameter (see Table 1). Histograms on the diagonal represent the uncertainty in parameter estimates from the parameter ensemble. Triangles below histograms represent default (closed) and best-fit (open) parameter values. Off-diagonal panels display pairwise correlations and correlation coefficients between individual parameters, with statistically significant correlations circled ($p < 0.05$). *Default parameter line for D is beyond plot scale ($D_{default} = 3.1$). (Generated by free software R, https://www.R-project.org/).

varied by a factor of 7.7 ± 1.7 (Fig. 4C,E, respectively). Finally, the parametric uncertainty led to a large range in the relative abundance of copiotrophs (MIC_r) and oligotrophs (MIC_K) that were simulated across the RCEW-CZO, with a mean MIC_r:MIC_K ratio of 0.80 ± 0.48 across individual locations (Fig. 4F).

High-resolution mapping and sensitivities of SOC stocks. A spatially continuous map of estimated SOC stocks from 0 to 30 cm soil depth was generated for the RCEW-CZO by combining results from the parameter ensemble with gridded spatial data (*i.e.*, rasters; Fig. SI1) of the required model inputs (Fig. 5). We note that these rasters were also used to generate the forcing data used for model parameter optimization (above) based on field observations of SOC stocks. To generate a high-resolution map of 0–30 cm SOC stocks for the entire natural area of the RCEW-CZO, we calculated the steady-state carbon pools simulated by MIMICS at 2.4 million points evenly distributed across the watershed (10 m resolution). We repeated this with each of the 30 members of the parameter ensemble to calculate uncertainty estimates in the spatial distribution of belowground C stocks. Figure 5 shows the ensemble mean and associated parametric uncertainty (± 1σ) of soil C stocks across the watershed.

The high-resolution mapping with MIMICS shows that soil C stocks follow elevational trends at RCEW-CZO, while also reflecting more localized topographical differences in precipitation and mean annual temperature that drive shifts in vegetation communities, productivity, and soil development. Total SOC stocks are largest in the southern and western portion of the watershed, where higher elevations bring cooler temperatures and higher water availability, which supports more productive vegetation communities. Within elevational bands we also see strong influences of aspect on SOC stocks. On average across the watershed, soils on north-facing slopes hold 28% more SOC relative to south-facing slopes. For areas above 1600 m elevation, the aspect effects on SOC

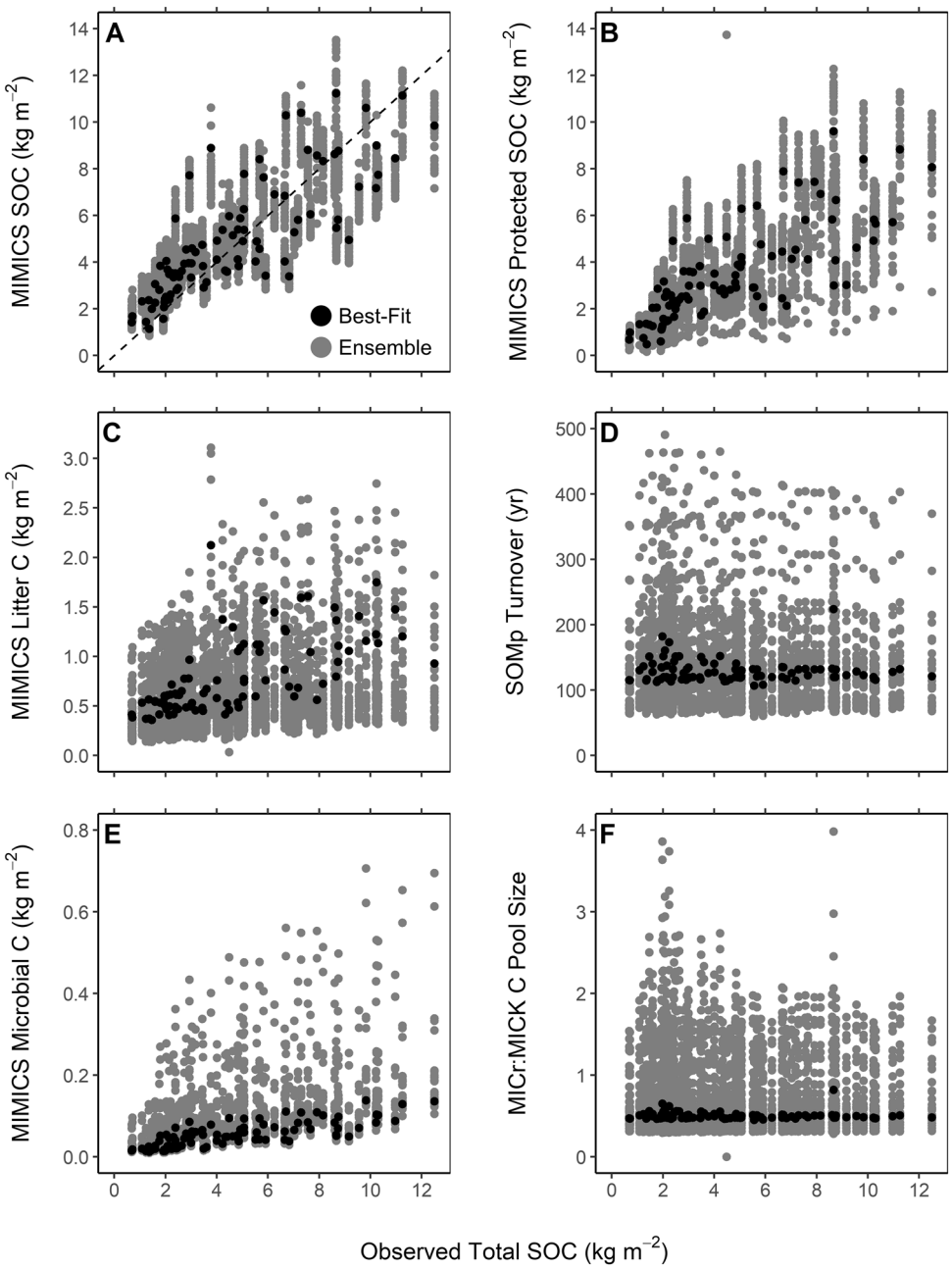

Figure 4. MIMICS estimate variability relative to observed SOC stocks for the best (lowest RMSE) MC determined parameterization and a larger ensemble of MIMICS model parameterizations (n = 30) with SOC stock estimate RMSE < 2. (Generated by free software R, https://www.R-project.org/).

storage increase, with 41% more SOC stored on north-facing slopes relative to south-facing slopes. The largest uncertainties in soil C stocks correspond to areas with higher productivity and larger SOC stock estimates. Finally, areas where the dominant litter source shifts from grass and shrub species to tree species with relatively poor litter quality (higher lignin:N) also show greater total SOC uncertainty.

Our simulations also generate spatially explicit estimates of individual soil C pools including mean litter C, microbial C and protected SOC stocks and their associated uncertainty from the parameter ensemble (Fig. 6). Litter C pools strongly reflect the aspect effects across Reynolds Creek, with 59% more C stored in litter pools on northern facing slopes relative to south-facing slopes, reflecting greater control by soil temperature on litter C pools. We estimate that litter C pools comprise 14 ± 7% of the total SOC stock from 0 to 30 cm across the natural areas of Reynolds Creek. Microbial C stocks largely reflect spatial patterns in vegetative productivity (Fig. 6A, Fig. SI1A). We estimate protected SOM accounts for 67 ± 16% of the total SOC stock at Reynolds Creek, which also reflects spatial distributions of NPP estimates and less dependence on soil clay content, temperature or litter quality.

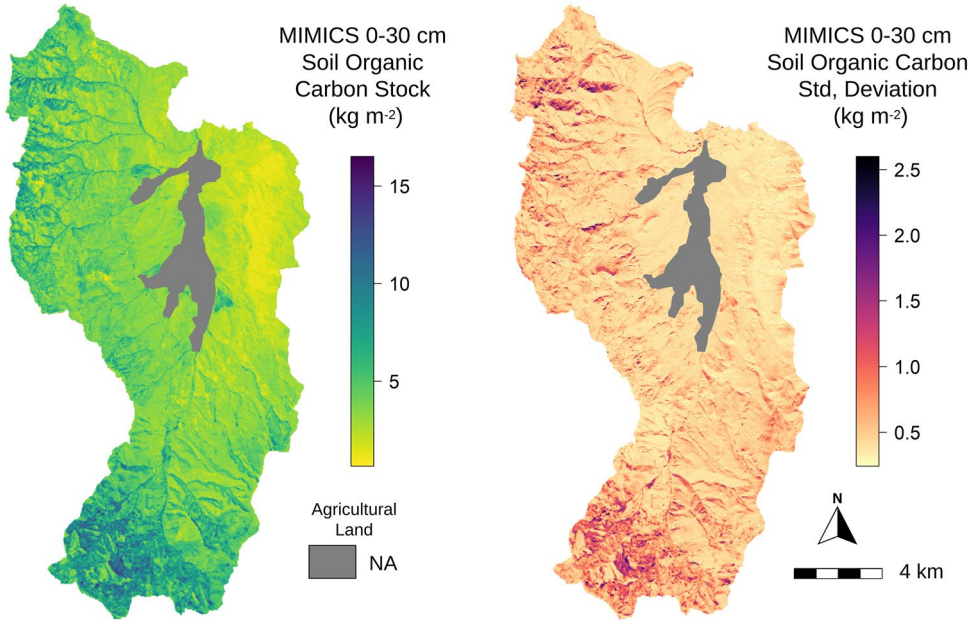

Figure 5. Ensemble (**A**) mean and (**B**) standard deviation of total soil organic carbon stocks simulated by the MIMICS model across the natural area of the Reynolds Creek Experimental Watershed and Critical Zone Observatory from 0–30 cm soil depth (n = 30 maps of soil C generated by the parameter ensemble produced from model calibration with site observations; Fig. 3a). (Generated by free software R, https://www.R-project.org/).

To project the sensitivity of existing soil C stocks to idealized environmental perturbations, we used the parameter ensemble to estimate soil C stock response to a uniform 10% increase in NPP and a 1 °C increase in mean annual soil temperature (Fig. 7). Gains in soil C from the increase in NPP were far greater in the southern, high elevation regions of the watershed where NPP is relatively high. The simulated 10% increase in NPP led to a total gain of 3.6 ± 0.6 GgC from 0–30 cm soil depth across Reynolds Creek, an increase of 3.5 ± 0.6% relative to initial stocks. Across ensemble member simulations, the mean estimated increase in SOC stocks across the watershed from the 10% increase in GPP ranged from 1.7—20.4% of existing stocks. In contrast, the 1 °C increase in soil temperature led to a more spatially consistent loss of soil C across Reynolds Creek, resulting in an estimated loss of 3.9 ± 2.8 GgC from 0–30 cm soil depth, or -3.7 ± 2.8% of initial stocks. Overall, projection uncertainty was much greater for the simulated change in soil temperature relative to the applied change in NPP (Fig. 7B). However, the range in net effect on SOC stocks was slightly smaller across the parameter ensemble. Across all ensemble simulations, the mean estimated change in SOC stocks from the 1 °C increase in soil temperature ranged from -1.1 to -14.9% of existing stocks.

Discussion

Overview. By coupling fine-resolution forcing data with model parameter optimization, we produced high resolution (10 m^2) estimates of soil carbon stocks across the natural land area of the RCEW-CZO (239 km^2 of complex terrain; Fig. 5). The use of high-resolution forcing data proved suitable for model performance when combined with parameter optimization, as the correlation between model projections and field observations of SOC stocks aligned with previous MIMICS applications (Fig. 2)[26] and studies involving digital soil mapping methods[17]. Optimization of the model parameters was critical for reducing estimate systematic bias from the priors used in the initial simulations and reducing the RMSE between simulated and observed SOC stocks (Fig. 2). These results directly demonstrate the viability of combining process-based modelling and remotely sensed data products for simulating SOC stocks across complex landscapes.

Similar accurate, high-resolution estimates of SOC stocks across a complex environmental landscape have, to our knowledge, not been achieved previously using a process-based model such as MIMICS. In contrast to similar scale estimates of SOC stocks produced by statistical methods[16,17], our process-based modeling approach simulates the controls on the decomposition, accumulation, and loss of SOC, providing predictive capacity to project soil C responses to changing environmental conditions. Perhaps the most similar study to date was conducted by Lu et al.[37], who used spatial forcing data with a process-based model to project SOC stocks at 250 m^2 resolution (r = 0.64) for a 7700 km^2 watershed in the Loess Plateau, China. Our ability to project SOC stocks for the RCEW-CZO at a 10 m^2 resolution with high accuracy (r = 0.79, Fig. 2) is a combined product of recent advances in remote sensing, high performance computing, and process-based modeling of SOC dynamics. Combined together, these modern tools were essential for reducing error between model estimates and field observations. When standardized by the range in size of the SOC stocks projected, our observed RMSE between estimates and projections of SOC stocks is consistent with respective measures of error from digital soil mapping

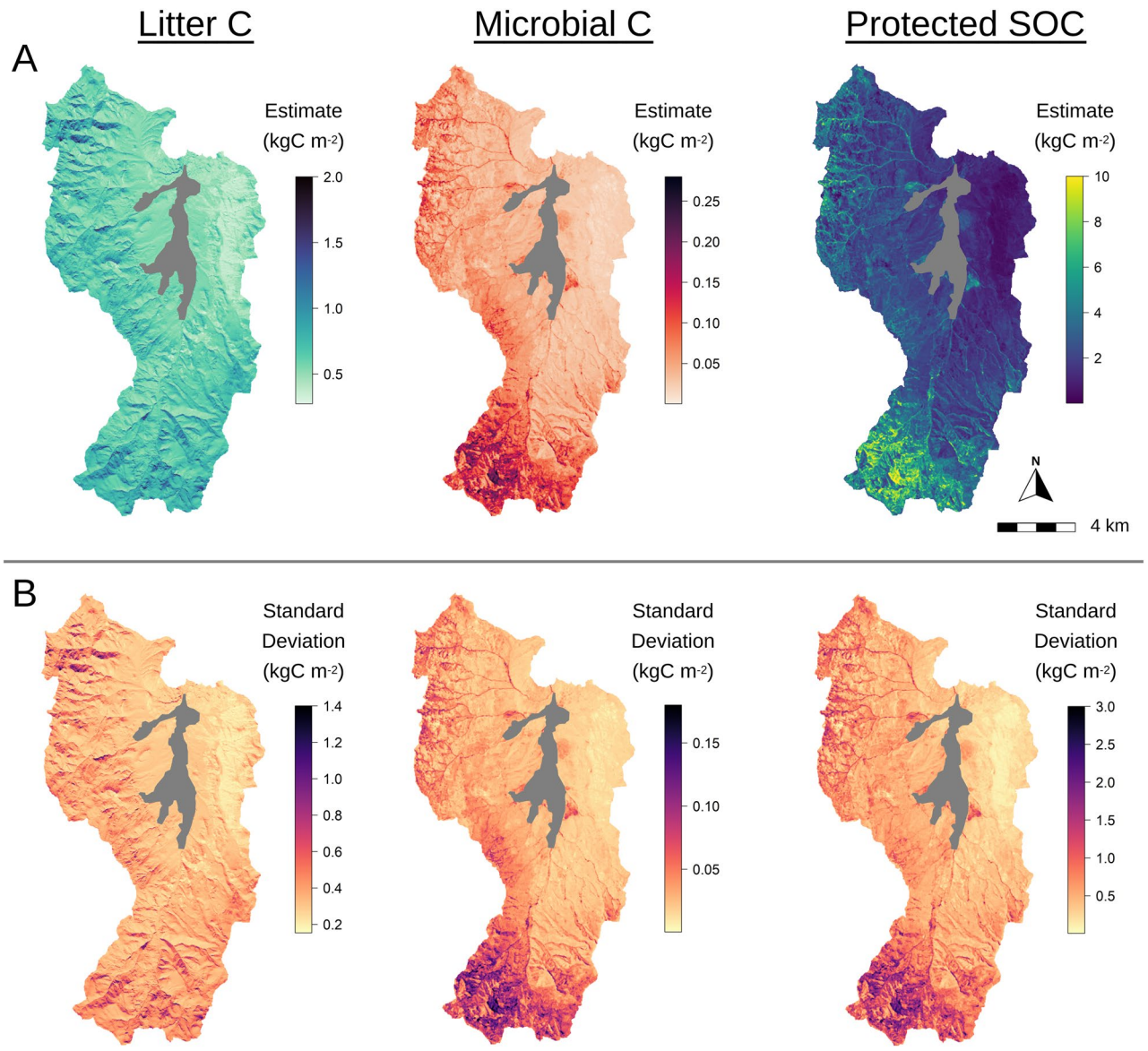

Figure 6. Ensemble (**A**) mean and (**B**) standard deviation of soil carbon pool stocks for 0–30 cm soil depth across the natural area of the Reynolds Creek Experimental Watershed and Critical Zone Observatory (n = 30 maps of soil C generated by the parameter ensemble produced from model calibration with site observations; Fig. 3a). (Generated by free software R, https://www.R-project.org/).

methods used at comparable scales[17,38–40]. Below, we discuss in greater detail the implications of combining parameter optimization and uncertainty analysis with high-resolution forcing data for future applications, as well as the unique information provided by high resolution modeling of SOC dynamics across complex landscapes.

Parameter estimation and uncertainty. Our estimates of SOC stocks across the RCEW-CZO suggest that the scales over which measurements are taken, and for which models are parameterized, influence the predicted sensitivity of SOC stocks to different environmental control factors. Differences between our optimized parameters and the default parameter estimates reflect the improvements made by calibrating the model to fit local observations (Fig. 2). Our findings also underscore important differences in parameter estimates generated by when calibrating MIMICS at local- versus continental-scales (Fig. 3, 4). For example, results from our model parameterization suggest a higher temperature sensitivity of microbial catabolism in our optimized parameterization of MIMICS across local gradients than was derived from the default parameterized generated from site level means across a continental-scale gradient (Table 1, Fig. 3)[26]. This parallels findings from Bradford et al.[24] who found strong scale dependence on the strength of causative relationships between environmental drivers and decomposition rates, suggesting that accounting for local heterogeneity may be important to consider for inferring proximal controls over SOC turnover that are used in model parameterizations[41].

Results from our parameter ensemble have higher temperature sensitivity for microbial kinetics and lower MGE than the initial parameterization (Fig. 3). These parametric changes related to microbial physiology

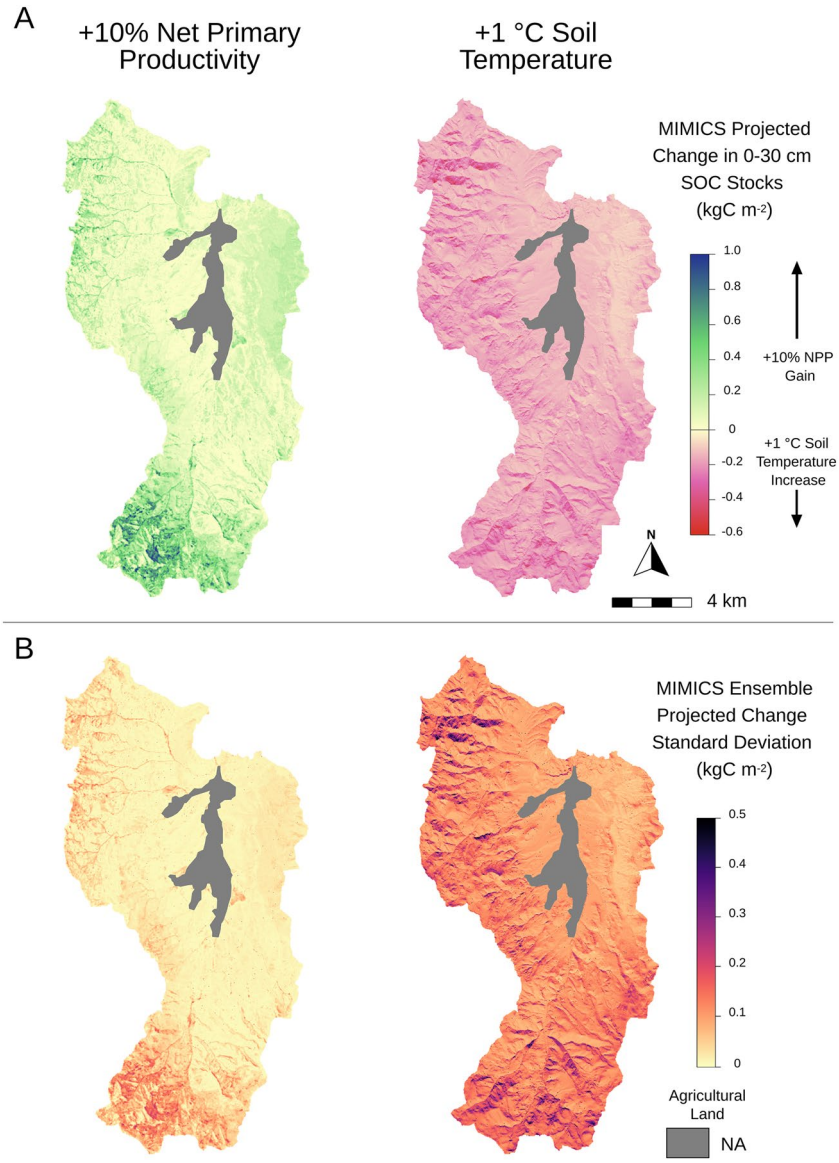

Figure 7. MIMICS projected (**A**) total change and (**B**) parametric uncertainty for change in 0–30 cm soil carbon stocks from 10% increase in net primary productivity (NPP) or 1 °C increase in mean annual soil temperature. (Generated by free software R, https://www.R-project.org/).

underscore how microbial communities can influence the rate of organic matter turnover through their catabolic capacity, as well as the fate of SOM via microbial anabolism[42]. Interestingly, our parameter ensemble did not indicate any statistical correlation between parameter combinations related to microbial catabolism or anabolism (Fig. 4), yet MIMICS represents a theoretical expectation that tradeoffs between microbial growth rates and growth efficiency should be linked in the parameterization of copiotrophic and oligotrophic communities[43,44]. Ultimately, the parameter estimates from our ensemble result in faster turnover and lower retention of fresh organic matter inputs– findings consistent with conclusions of Shi et al.[45]. The parameter ensemble also produced much longer turnover times for physicochemically protected soil C pools, correcting known biases in MIMICS[22,46]. Turnover times of SOM_p pools, however, still have large uncertainty, leading to large variation in the size of this soil C pool (Figs. 3, 4). Moving forward, we expect information about soil C fractions and their radiocarbon ages will provide an important additional constraint to reduce these uncertainties.

The parametric uncertainty analysis also identified what additional observations could help constrain model representations of microbial physiology and SOC persistence (Figs. 3, 4). Specifically, observations of litter decomposition rates, the relative size of SOC pools (e.g., from density fractionation), laboratory measurements of microbial carbon use efficiency (or microbial growth efficiency, MGE), and the turnover rate for mineral associated organic matter (from radiocarbon measurements on mineral associated organic matter) are directly relatable to the model and can be used as constraints during parameter optimization to better calibrate the model parameters. Auto-correlations between model parameters suggest that data pertaining to any one property may help to constrain other parameters in the model. For example, desorption rates (D) and the fraction of microbial

Parameter	Description	Default prior[a]	Best-fit scaling factor[b]	Scaling factor proposal range[c]	Units
V_{slope}	Decomposition max rate coefficient	0.063	3.85 ± 0.8	0.4–4	ln(mg Cs (mg MIC)$^{-1}$ h^{-1}) °C^{-1}
V_{int}	Decomposition max rate coefficient	5.47	1.16 ± 0.4	0.3–3	ln(mg Cs (mg MIC)$^{-1}$ h^{-1})
K_{slope}	Decomposition half saturation coefficient	0.017, 0.027, 0.017	0.84 ± 1.0	0.4–4	ln(mg C cm^{-3}) °C^{-1}
K_{int}	Decomposition half saturation coefficient	3.19	1.55 ± 0.7	0.3–3	ln(mg C cm^{-3})
τ	Microbial biomass turnover rate	$0.30 \times e^{1.3(fclay)}$, $0.20 \times e^{0.8(fclay)}$	0.46 ± 0.5	0.3–3	h^{-1}
MGE	Microbial growth efficiency	0.55, 0.25, 0.75, 0.35	0.54 ± 0.3	0.2–2	mg mg^{-1}
D	Desorption rate from SOMp to SOMa	$1.5 \times 10^{-5} \times e^{-1.5(fclay)}$	0.08 ± 0.03	0.001–0.3	h^{-1}
f_p	Fraction of τ partitioned to SOMp	$0.015 \times e^{1.3(fclay)}$, $0.01 \times e^{0.8(fclay)}$	0.34 ± 0.3	0.01–4	–

Table 1. MIMICS parameters included in the Monte Carlo (MC) optimization and the subsequent best-fit values found for the Reynolds Creek Experimental Watershed and Critical Zone Observatory (RCEW-CZO) calibration data. [a]Further information regarding MIMICS parameters and default initial values provided in[26] [b]Parameter proposals in the MC algorithm were conveyed as a factor of the initial prior value(s). Presented standard deviation is derived from the parameter ensemble (n = 30) with RMSE 1.8–2.0 [c]Factor proposals for each parameter were confined to agree with natural analogs and to reduce computational demands. See Methods for further details.

turnover that is physicochemically stabilized (f_p) show high correlation with each other (Fig. 3). This suggests that radiocarbon measurements that provide a constraint on mineral protected SOM turnover would reduce uncertainty ranges for both of these parameters. Similarly, measurements of microbial growth efficiency are becoming more common, and may also help constrain the parameterization of microbial turnover in the model. We note, however, that laboratory assays of MGE are short-term measurements of (near) instantaneous growth efficiency that are dependent on substrate quality and particulars of the chosen methodological approach[47,48]. As such, the absolute values of MGE from any incubation study may not be directly representative of these fixed parameters that are currently used in MIMICS, highlighting the importance of improving our understanding and representation of this key parametric uncertainty in soil biogeochemical models[47,49–51]. More relevant, may be developing an understanding of how microbial community composition reflects environmental gradients and may influence variation in MGE over space and time. Overall, this work illustrates improved capabilities and opportunities to exchange information between process-based models, field measurements, and laboratory observations. Such exchanges are critical for ensuring that model representations of SOC dynamics align with natural systems and produce accurate projections.

High resolution inputs, SOC stocks, and projections. Our study highlights the data challenges associated with trying to generate measurements that are needed for this kind of work. Even at a well-studied site like the RCEW-CZO that has a wealth of soils data, co-located measurements of additional driver data that are needed to parameterize MIMICS were largely absent (e.g. soil properties, soil temperature and moisture, plant productivity, and litter chemistry). Thus, we derived these from remote sensing and statistical modeling. Although these methods are still imprecise, they better account for local heterogeneity in biotic and abiotic factors known to influence SOM persistence and mark an important step forward in considering how we can parameterize and validate soil biogeochemical models.

Process-based models of SOC dynamics employ varied constructs and environmental properties to predict SOC[52]. The MIMICS model has been shown to have a relatively high dependence on plant productivity, as opposed to greater dependence on soil texture found in other, similar process-based models of SOC dynamics[7]. In practice, our high-resolution mapping application benefited from MIMICS dependence on NPP, since estimates of productivity are readily derived from high resolution multispectral imagery (< 30 m^2)[53,54]. Methods for estimating vegetative properties from multispectral imagery are broadly effective for most environments[55–57]. In contrast, relatively few methods exist for collecting information on soil properties (e.g. clay content) at similar resolutions[58]. The reduced influence of clay content relative to NPP on SOC dynamics in MIMICS has been proposed as a potential cause for observed discrepancies between SOC stocks simulated by MIMICS relative to field observations and current theory[7]. However, results from our parameter optimization procedure demonstrate that such discrepancies in MIMICS projections of protected SOC can be rectified through the model parameterization (Fig. 4D).

Surprisingly, we found that the inclusion of soil moisture in the MIMICS model, a factor with strong influence on SOC stocks at global scales[59], did not improve estimate accuracy across the RCEW-CZO watershed. This result may in part be explained by the fact that soil water availability along the elevational gradient in the RCEW-CZO is strongly correlated with productivity estimates from remote sensing at the site[60,61]. Thus, the spatially distributed productivity estimates that we developed likely captured the effects of moisture and was a significant factor explaining SOC stocks. Alternatively, accuracy of the spatial soil moisture data available may not have been sufficient to appropriately convey soil moisture conditions at the spatial resolution of this study, and ongoing work at the watershed is focused on generating high spatial and temporal resolution estimates for the site. We see such efforts as critical for estimating temporal variation in heterotrophic respiration fluxes and improving projections of soil C responses to climate change. In contrast, we found that spatially continuous

mean annual soil temperature (MAST) derived from soil temperature observations, elevation and estimated incoming solar radiation was a significant factor explaining SOC stocks. Seyfried et al.[62] and more recently Seyfried et al.[63] documented pronounced aspect effects on local soil temperature at the RCEW-CZO that rival larger elevational gradients. For example, differences in soil temperature associated with aspects in small catchments of the RCEW-CZO (< 2 km^2) are greater than differences in MAST observed along a 900 m elevation gradient (5.5 versus 4.4 °C), exemplifying the importance of using fine scales to model environmental processes in complex landscapes.

Our assessment of SOC stocks and sensitivities at high spatial resolution across the RCEW-CZO provides a robust example of how fine scale applications of process-based models of SOC dynamics can provide useful information to support land practice and policy decisions. High spatial resolution estimates of SOC stocks across the entire RCEW-CZO provide insight into landscape variation in SOC stocks that may be expected across the semi-arid environments of the Great Basin, USA (Fig. 5). We see this providing opportunity for land managers to better plan land use and focus valuable resources towards specific areas where the protection or management of SOC is warranted. The model also makes testable predictions about the size of particular soil C pools and their change over environmental gradients (Fig. 6), that could be validated with additional measurements taken at field relevant scales. Further, our idealized environmental change scenarios demonstrate the ability to assess SOC stability across a diverse landscape (Fig. 7).

The idealized scenarios used here illustrate the potential to use a calibrated model like MIMICS to make spatially-explicit projections of SOC responses to environmental change. Rangelands exist over ~ 40% of the Earth's ice-free land and account for ~ 30% of terrestrial SOC to a depth of 1 m[64]. The global potential for rangeland C sequestration has been estimated to range from 0.3 to as much as 1.6 Pg CO_2-eq per year[65]. Moreover, experiments like these can provide estimates of the magnitude of soil C change, and their associated uncertainty, that can be expected over large areas given incremental changes in local conditions (like temperature, productivity, and their interactions). High resolution process-based modeling of soil C dynamics provides a critical tool for evaluating such potential, as well as for identifying, protecting and harnessing sequestration potential in soil across earth's surface.

Conclusion

In this study, we show that accurate, high resolution (10 m^2) projections of 0–30 cm SOC stocks may be generated by combining remotely sensed environmental data and process-based modeling, with the aid of parameter optimization methods and high-performance computing. Leveraging the power of modern parameter optimization techniques and machine learning to improve Earth systems models remains a widespread area of emphasis across disciplines in environmental science[66,67]. The Monte Carlo simulation used for parameterizing the MIMICS model provided the cornerstone for this study, allowing us to calibrate model parameters, quantify the impacts of parametric uncertainty, and generate spatially explicit estimates of soil C stocks with high-resolution environmental data. For future studies and applications, we see great promise for similar approaches to assist with model parameterization and spatial extrapolations of SOC projections.

Required adjustments to the MIMICS parameters controlling SOC distributions and environmental sensitivity in simulations for the RCEW-CZO support a deepening discussion in ecology and Earth system science regarding the potential fallacies of inference when scales between observations and the environmental processes being studied are misaligned[24]. Observed differences in the best-fit parameters between fine and continental scale applications of the MIMICS model also support other recent studies in demonstrating the substantial variability found in soil processes at fine spatial scales[11,68,69]. The demonstrated ability to perform fine scale simulations of SOC dynamics thus provides a critical tool and resource for linking environmental observations with current theory and associated model representations. Such ability is critical for improved assessments and understanding of landscape drivers of SOC persistence, and for the projection of SOC response to changes in environmental conditions.

Moving forward, the proven approach and developed algorithms from this study provide a strong foundation for continued model improvement and further high-resolution applications. Our analysis of the parametric uncertainty in the model projections shows the direct opportunities for additional observational data pertaining to microbial properties and SOC persistence to further refine MIMICS parameter estimates. Further potential also exists to generate more complex, fine scale estimates and projections of SOC dynamics by adapting the methods from this study to similarly enhance the spatial application scale and parameterization of the CN-coupled[70] and soil depth resolved[71] versions of the MIMICS model.

Methods

Study area. The Reynolds Creek Experimental Watershed and Critical Zone Observatory (RCEW-CZO) in southwestern Idaho, USA covers 239 km^2, with 13 km^2 of active cropland near the valley bottom. The RCEW-CZO is situated in a mountainous region and contains a large elevational gradient (~ 1100–2100 m) that gives rise to diverse climates, soils and vegetation throughout the watershed[36,63]. Vegetative productivity is severely limited by low water availability in lower elevations (< 1400 m) of the watershed, which receive an average of 230–300 mm of precipitation each year. Mean annual temperature across the lower elevations of the watershed is 9.1 °C. In the lowland regions, the vegetation is predominately sagebrush (*Artemisia spp.*) with a sparse underlying mix of grasses and forbs. At the mid-elevations of the watershed (1400–1700 m), pockets of aspen (*Populus tremuloides*) and willow (*Salix spp.*) begin to emerge in riparian areas, as well as on some of the northern facing slopes, where winter snowpacks supply sufficient water. At higher elevations, Juniper (*Juniperus spp.*) and Douglas fir (*Pseudotsuga menziesii*) stands are common, though not pervasive, supported by far greater annual precipitation and lower mean annual air temperatures, averaging 350–994 mm per year and 5.4–8.5 °C. Soils across the

RCEW-CZO watershed are mixed and predominantly originate from either felsic material (i.e., granitic Idaho Batholith[72]) or mafic basaltic flows[73]. Soils are generally deep across the RCEW-CZO (> 1 m not including colluvium)[74]. Soil development is typically greatest at higher elevation in convergent topography. For the top 30 cm of soil, the mafic derived soils are generally loams (average sand, silt, clay: 41%, 39%, and 20%), while the felsic are generally sandy loams (average sand, silt, clay: 59%, 25%, and 16%[16,75]). Across the RCEW-CZO, approximately 85% of the soils are classified as Mollisols, with another 13% classified as Aridisols[76].

Field observations of SOC stocks. Soil C stocks were previously determined at unique locations across the RCEW between 2014–2016[75,77,78]. For model calibration we used previously collected data pertaining to total soil C stocks from 0–30 cm at 89 locations across the RCEW-CZO. The median 0–30 cm soil C stock was used where soil data overlapped within a 10 m^2 area.

The MIMICS model and parameter optimization methods. The MIcrobial-Mineral Carbon Stabilization (MIMICS) model has been widely used for estimation of soil C across diverse ecosystems and has been found to perform well across ecosystem gradients and at global scales[22,26,43,46,70,79]. Here we use a version of the model that calculates equilibrium SOC stocks based on the mean annual values for net primary productivity, soil temperature, litter lignin:N ratio and the soil clay content. Soil moisture may also be used as a control on soil C dynamics in the MIMICS model[22]. However, model testing using soil moisture inputs derived from aerial imagery and a snow model did not improve model accuracy of SOC stocks estimates at RCEW, and thus soil moisture was not included as a model input. We assumed all field observations of SOC included in the study are at steady state and ignored the interannual and seasonal dynamics of climate and vegetation. In kind, all model simulations were carried out to steady state. These assumptions present uncertainty and limitations regarding the insights and projections from the model, but it remains common practice to allow for best use of available data[43,46,79].

The flow of C through the MIMICS model is controlled by a number of rate parameters (Table 1, Fig. 1)[26]. These parameters control the simulated turnover and persistence of litter, microbial biomass, and soil C pools (Fig. 1; Methods). Model parameters broadly fit into three categories related to microbial catabolic capacity (decomposition kinetics described by V_{max} and K_{es}), microbial anabolism (microbial growth efficiency and turnover; MGE and τ, respectively), and physicochemical protection of SOC (Fig. 1).

Initially, we used a Hamiltonian Markov Chain Monte Carlo (MCMC) algorithm[80] to determine the best fit values for eight of the MIMICS model parameters with the strongest control over litter, microbial biomass, and soil C pools. Uniform prior distributions for the MCMC were given a mean and standard deviation equal to the parameter values provided by Wieder et al.[26]. In practice, the best-fit parameterizations determined over multiple MCMC runs were not consistent, suggesting that substantial equifinality exists in the model construct. To account for and gain further insight into how equifinality in model parameterization impacts model estimates, we proceeded to perform a Monte Carlo (MC) simulation involving 500,000 MIMICS model runs with unique, random parameter combinations. To improve performance of the MC simulation, we proceeded to tune the minimum and maximum limits on each of the uniform parameter proposal distributions (i.e. the scaling factor proposal range, see Table 1) to ensure that the range encompassed all values with the potential to produce viable model outcomes (see criteria below). Specifically, using repeated trials of the MC simulation, the scaling factor proposal range was extended when viable model outcomes were observed within 10% of the proposal value limit Conversely, scaling factor ranges were narrowed when parameter proposals above or below a certain limit resulted in implausible model outcomes (i.e., Desorption rates (D) > 0.3 produce protected SOM pool (SOMp) turnover rates below the constraint of 50 years (See further model outcome constraint details below)). Ultimately, all parameterizations from the MC simulation yielding a RMSE between field and estimated bulk SOC values of < 2.0 kgC m^{-2} were included in the parameter ensemble used throughout the study. Further, a parameter set was only deemed viable if it passed a set of binary filters to confirm that the parameterization produced plausible values for microbial C (1–3% of the total SOC), the relative abundance of microbial r- vs. K-strategist (0.1–10), litter C stocks (10–40% of total SOC), and the turnover time of C in the protected SOM pool (50–400 years). For brevity, we refer to these accurate (RMSE < 2.0) and plausible parameterizations throughout the manuscript as the parameter ensemble (n = 30). The lowest RMSE observed over all combinations was 1.72 kgC m^{-2}.

Previous studies have optimized MIMICS parameters using MCMC[79] and shuffled complex evolution (SCE) algorithms[46]. We chose to employ a MC simulation to optimize the model parameters due to the simplicity of the algorithm relative the MCMC and SCE methods. The MC simulation requires no statistical assumptions and has no inherent potential to bias the optimized parameters chosen. The MC optimization approach is computationally inefficient compared with MCMC and SCE, which required us to use high performance computing clusters to complete the necessary number of model simulations (n = 500,000). We performed a threefold cross-validation to ensure the MC optimization method results were not biased by the nature of the RCEW forcing data or the number of simulations performed. Each cross-validation of the MC optimization procedure used a unique sub-sample (n = 60) of the field observations of SOC stocks to derive the optimized parameter distributions. Cross-validation sub-samples were stratified by co-located estimates of GEP to ensure consistency in the scope of the model projections. The resulting optimized parameter distributions from cross-validation simulations showed no concerning disparities between the subsample distributions and the distributions derived from the full dataset (Figure SI3).

We used the following approach to quantify the variability in MIMICS C pool estimates produced by the unique parameterizations in the parameter ensemble. For each location in the dataset, we calculated the range of estimated values and then divided that by the minimum estimated value. We then calculated the average and standard deviation of those location specific estimate variability factors across all locations.

Spatial data analysis. Model estimation of soil C stocks across the entirety of the RCEW-CZO required continuous spatial data layers for elevation and incoming solar radiation, soil temperature, net primary productivity, soil clay content and the lignin:N ratio for the dominant vegetative source of surface litter. Digital elevation and soil clay content layers were used as provided online courtesy of the U.S. Geological Survey (https://earthexplorer.usgs.gov/) and USDA Web of Soil Survey[76] respectively. Clay content was procured from the Web Soil Survey as an average value for the soil layer from 0–30 cm. The underlying source of the WSS clay data is the SSURGO database[81]. To create a spatial map of lignin:N content of the dominant vegetative species across the RCEW, the vegetation map provided in Seyfried et al.[82] was paired with corresponding litter lignin:N values reported in the literature. From the digital elevation map, incoming solar radiation was calculated using the ESRI Solar Radiation Toolset, which calculates net incoming solar radiation by simulating the solar path over the digital elevation map surface while accounting for atmospheric effects, site latitude and elevation, slope, aspect, daily shifts of the sun angle, and effects of shadows cast by surrounding topography[83].

Soil temperature. A continuous spatial layer for mean annual soil temperature (MAST) was created based on the relationship found between soil temperature, elevation and incoming solar radiation. Across the full extent of the elevation gradient of Reynolds Creek, and for soils with similar aspects, there is a strong relationship between soil temperature and elevation. However, soil temperatures differ substantially for soils with opposing north and south facing aspects, as shown by Godsey et al.[84] and Seyfried et al. [62,63]. To compensate for these aspect effects on MAST, estimated annual incoming solar radiation was used as an adjustment factor as shown in the following equation:

$$MAST = -0.0087(ELEV) + 22.43 - (solarMID - solarIN/tMOD)$$

where ELEV represents elevation (m); solarMID represents the average incoming solar radiation across the RCEW-CZO (553,833 W m^{-2}); solarIN represents the incoming solar radiation at the estimate location (W m^{-2}) and tMOD is employed as a scaling factor (46,385.8 W m^{-2}). All coefficients, and the tMOD scaling factor, were determined using a linear optimization algorithm set to minimize error between observed and estimated MAST values for a series of locations across the RCEW-CZO (n = 35, data from Seyfried et al.[62,63] and Godsey et al.[84]). The spatially dispersed and depth stratified MAST observation data showed no significant differences in MAST across soil depth increments of 5, 10, 20 and 30 cm. Based on this evidence, all MAST data available for soil depths < 30 cm were used to calibrate the MAST estimation equation. Correlation between observed and predicted MAST values equaled $R^2 = 0.73$ (Fig. SI4). The resulting continuous spatial map of MAST produced is assumed to be representative of MAST from 0–30 cm soil depth.

Gross primary productivity. Hyperspectral satellite-based imagery was used to generate a continuous spatial map of gross ecosystem productivity (GEP) across the RCEW-CZO. Following work previously completed by Fellows et al.[60], Landsat 5 imagery courtesy of the U.S. Geological Survey was collected across 10 years, from 2002 to 2011. From this data, specific images were chosen from each year to determine the maximum modified soil-adjusted vegetation index (MSAVI2) values at a 30 m^2 resolution across the RCEW-CZO. The MSAVI2 index has been demonstrated as a viable method for identifying differences in vegetative productivity[85–87]. MSAVI2 was calculated using the following equation:

$$MSAVI2 = \left(2 * NIR + 1 - \sqrt{((2 * NIR + 1)^2 - 8 * (NIR - RED))}\right)/2$$

where NIR is the near infrared band reflectance and RED is the red band reflectance value. Generation of a continuous map of GEP was then based on the observed relationship between maximum mean annual MSAVI2 and observed measures of GEP across the RCEW-CZO (GEP = 1972.8(MSAVI2) + 101; r = 0.69). Observations of GEP were also collected between 2002–2011 and were generated by eddy covariance at three elevation stratified flux towers. The relationship between observed and predicted GEP is shown in (Fig. SI4). To run the MIMICS model, which requires a measure of vegetative productivity in the form of annual net primary productivity, we estimated ANPP to equal half of GEP.

Spatial extrapolation and environmental change scenarios. As needed, all spatial data layers were resampled into raster format with 10 m^2 resolution using nearest neighbor resampling method, and all layers were re-projected to a common projection and coordinate reference system (World Geodetic System (WGS) 1984, Universal Transverse Mercator (UTM) Zone 11). For model calibration and validation datasets, pixel values from the spatial data layers for GEP, MAST, clay content and litter lignin:N content were extracted at the coordinates of sampling points.

We performed MIMICS simulations for each grid cell in the RCEW-CZO spatial data (2.4 million grid cells). Each grid cell simulation involved producing estimates from each member of the parameter ensemble (n = 30). The average estimate at each grid cell was then used to construct the continuous maps (Figs. 5, 6, 7). The corresponding parametric uncertainty maps were similarly generated by calculating the standard deviation of the estimates for each grid cell. Continuous spatial estimates of SOC responses to idealized scenarios were similarly performed. Idealized scenarios were applied through a uniform shift in the corresponding forcing data layer (+ 10% increase in GPP or + 1 °C MAST) prior to the model simulations. The High-Performance Computing (HPC) resources provided by Idaho State University and the U.S. Department of Energy Idaho National Laboratory were invaluable for the high-resolution spatial projections, allowing us to complete over 300 million MIMICS simulations.

Statistical analysis and code. All modeling, analysis and spatial mapping for this study were performed in R[88]. Root mean square error and correlation between model estimates and field data were specifically performed using the Metrics (https://github.com/mfrasco/Metrics) and base R packages respectively[88]. The MIMICS model and the developed Markov Chain Monte Carlo and Monte Carlo simulation algorithms were composed in R programming scripts and are available in open-access repositories on GitHub (https://github.com/piersond/MIMICS_HiRes, https://github.com/wwieder/MIMICS/tree/sandbox). The MIMICS_HiRes repository provides all modeling and analysis code associated with this study, as well as useful startup information and simplified examples to assist future use. Scripts in the repository were assembled specifically for use with HPC to reduce the time required for model parameterization and simulations.

Data availability

The spatial and tabular datasets used for model forcing, calibration and validation within this study are publicly available from Boise State University ScholarWorks. (https://scholarworks.boisestate.edu/reynoldscreek/26/; https://doi.org/10.18122/reynoldscreek.26.boisestate).

Code availability

Codes used for MIMICS parameterizations, estimates, projections and spatial extrapolations associated with this manuscript are publicly available on GitHub. (https://github.com/piersond/MIMICS_HiRes; https://doi.org/10.5281/zenodo.5794655).

Received: 11 February 2022; Accepted: 2 June 2022
Published online: 25 June 2022

References

1. Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **12**, 1-40 (2017).
2. Janzen, H. H. Carbon cycling in earth systems - A soil science perspective. *Agric. Ecosyst. Environ.* **104**, 399–417 (2004).
3. Biskaborn, B. K. *et al.* Permafrost is warming at a global scale. *Nat. Commun.* **10**, 1–11 (2019).
4. Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science (80-.)* **366**, 1-10 (2019).
5. Nerem, R. S. *et al.* Climate-change–driven accelerated sea-level rise detected in the altimeter era. *Proc. Natl. Acad. Sci. USA* **115**, 2022–2025 (2018).
6. Todd-Brown, K. E. O. *et al.* Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* **11**, 2341–2356 (2014).
7. Georgiou, K. *et al.* Divergent controls of soil organic carbon between observations and process-based models. *Biogeochemistry* **2**, 5–17 (2021).
8. Beillouin, D. *et al.* A global overview of studies about land management, land-use change, and climate change effects on soil organic carbon. *Glob. Change Biol.* **28**, 1690–1702 (2022).
9. Ogle, S. M. *et al.* Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. *Glob. Change Biol.* **16**, 810–822 (2010).
10. Smith, P. *et al.* Towards an integrated global framework to assess the impacts of land use and management change on soil carbon: Current capability and future vision. *Glob. Change Biol.* **18**, 2089–2101 (2012).
11. Bradford, M. A. *et al.* Managing uncertainty in soil carbon feedbacks to climate change. *Nat. Clim. Change* **6**, 751–758 (2016).
12. Bradford, M. A. *et al.* Soil carbon science for policy and practice. *Nat. Sustain.* **2**, 1070–1072 (2019).
13. Mulder, V. L., de Bruin, S., Schaepman, M. E. & Mayr, T. R. The use of remote sensing in soil and terrain mapping—A review. *Geoderma* **162**, 1–19 (2011).
14. Sanderman, J. *et al.* Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy. *Biogeochemistry* **156**, 97–114 (2021).
15. Kunkel, M. L., Flores, A. N., Smith, T. J., McNamara, J. P. & Benner, S. G. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. *Geoderma* **165**, 1–11 (2011).
16. Patton, N. R., Lohse, K. A., Seyfried, M. S., Godsey, S. E. & Parsons, S. B. Topographic controls of soil organic carbon on soil-mantled landscapes. *Sci. Rep.* **9**, 1–15 (2019).
17. Minasny, B., McBratney, A. B., Malone, B. P. & Wheeler, I. *Digital Mapping of Soil Carbon. Advances in Agronomy* Vol. 118 (Elsevier, 2013).
18. Cuddington, K. *et al.* Process-based models are required to manage ecological systems in a changing world. *Ecosphere* **4**, 1–12 (2013).
19. Schimel, D. S. *et al.* Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Global Biogeochem. Cycles* **8**, 279–293 (1994).
20. Koven, C. D., Lawrence, D. M. & Riley, W. J. Permafrost carbon-climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics. *Proc. Natl. Acad. Sci. USA* **112**, 3752–3757 (2015).
21. Lacoste, M., Viaud, V., Michot, D. & Walter, C. Model-based evaluation of impact of soil redistribution on soil organic carbon stocks in a temperate hedgerow landscape. *Earth Surf. Process. Landforms* **41**, 1536–1549 (2016).
22. Wieder, W. R. *et al.* Beyond static benchmarking: using experimental manipulations to evaluate land model assumptions. *Global Biogeochem. Cycles* **33**, 1289–1309 (2019).
23. Toth, C. & Jóźków, G. Remote sensing platforms and sensors: A survey. *ISPRS J. Photogramm. Remote Sens.* **115**, 22–36 (2016).
24. Bradford, M. A. *et al.* A test of the hierarchical model of litter decomposition. *Nat. Ecol. Evol.* **1**, 1836–1845 (2017).
25. Wiesmeier, M. *et al.* Soil organic carbon storage as a key function of soils—A review of drivers and indicators at various scales. *Geoderma* **333**, 149–162 (2019).
26. Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G. & Bonan, G. B. Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geosci. Model Dev.* **8**, 1789–1808 (2015).
27. Sulman, B. N., Phillips, R. P., Oishi, A. C., Shevliakova, E. & Pacala, S. W. Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO 2. *Nat. Clim. Chang.* **4**, 1099–1102 (2014).
28. Parton, W. J., Cole, C. V., Stewart, J. W. B., Ojima, D. S. & Schimel, D. S. Simulating regional patterns of soil C, N, and P dynamics in the US central grasslands region. Ecology of arable land—Perspectives and challenges, 99–108 (1989).
29. Jenkinson, D. S. & Rayner, J. H. The turnover of soil organic matter in some of the Rothamsted classical experiments. *Soil Sci.* **123**(5), 298–305 (1977).

30. Lajtha, K. *et al.* The detrital input and removal treatment (DIRT) network: Insights into soil carbon stabilization. *Sci. Total Environ.* **640–641**, 1112–1120 (2018).
31. Aas, A. B. *et al.* Fine-scale diversity patterns in belowground microbial communities are consistent across kingdoms. *FEMS Microbiol. Ecol.* **95**, 1–11 (2019).
32. Elliot, W. J. *Soil Erosion in Forest Ecosystems and Carbon Dynamics. The Potential of US Forest Soils to Sequester Carbon and Mitigate the Greenhouse Effect* 175–190 (CRC Press, 2002).
33. Chappell, A., Baldock, J. & Sanderman, J. The global significance of omitting soil erosion from soil organic carbon cycling schemes. *Nat. Clim. Chang.* **6**, 187–191 (2016).
34. Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K. & Paul, E. 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. Chang. Biol.* **19**, 988–995 (2013).
35. Abramoff, R. Z. *et al.* Improved global-scale predictions of soil carbon stocks with Millennial Version 2. *Soil Biol. Biochem.* **164**, 108466 (2022).
36. Seyfried, M. *et al.* Reynolds creek experimental watershed and critical zone observatory. *Vadose Zo. J.* **17**, 1–20 (2018).
37. Lu, N. *et al.* Changes in soil carbon stock predicted by a process-based soil carbon model (Yasso07) in the Yanhe watershed of the Loess Plateau. *Landsc. Ecol.* **30**, 399–413 (2015).
38. Mora-Vallejo, A., Claessens, L., Stoorvogel, J. & Heuvelink, G. B. M. Small scale digital soil mapping in Southeastern Kenya. *CATENA* **76**, 44–53 (2008).
39. Marchetti, A., Piccini, R., Francaviglia, R., Santucci, S. & Chiuchiarelli, I. Estimating soil organic matter content by regression kriging. Progress in Soil Science 2. In *Digital Soil Mapping. Bridging Research, Environmental Application, and Operation* (eds Boettinger, J. L. *et al.*) 241–254 (Springer, 2010).
40. Vasques, G. M., Grunwald, S., Comerford, N. B. & Sickman, J. O. Regional modelling of soil carbon at multiple depths within a subtropical watershed. *Geoderma* **156**, 326–336 (2010).
41. Bradford, M. A. *et al.* Quantifying microbial control of soil organic matter dynamics at macrosystem scales. *Biogeochemistry* **156**, 19–40 (2021).
42. Schimel, J. P. & Schaeffer, S. M. Microbial control over carbon cycling in soil. *Front. Microbiol.* **3**, 1–11 (2012).
43. Wieder, W. R., Boehner, J. & Bonan, G. . Global Biogeochemical Cycles in Earth system models with observations. AGU Publ. 211–222 (2014).
44. Kallenbach, C. M., Frey, S. D. & Grandy, A. S. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nat. Commun.* **7**, 1–10 (2016).
45. Shi, Z. *et al.* The age distribution of global soil carbon inferred from radiocarbon measurements. *Nat. Geosci.* **13**, 555–559 (2020).
46. Zhang, H. *et al.* Microbial dynamics and soil physicochemical properties explain large-scale variations in soil organic carbon. *Glob. Chang. Biol.* **26**, 2668–2685 (2020).
47. Frey, S. D., Lee, J., Melillo, J. M. & Six, J. The temperature response of soil microbial efficiency and its feedback to climate. *Nat. Clim. Change* **3**, 395–398 (2013).
48. Geyer, K. M., Dijkstra, P., Sinsabaugh, R. & Frey, S. D. Clarifying the interpretation of carbon use efficiency in soil through methods comparison. *Soil Biol. Biochem.* **128**, 79–88 (2019).
49. Allison, S. D., Wallenstein, M. D. & Bradford, M. A. Soil-carbon response to warming dependent on microbial physiology. *Nat. Geosci.* **3**, 336–340 (2010).
50. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Chang.* **3**, 909–912 (2013).
51. Liang, C., Schimel, J. P. & Jastrow, J. D. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* **2**, 1-6 (2017).
52. Sulman, B. N. *et al.* Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry* **141**, 109–123 (2018).
53. Dechant, B., Ryu, Y. & Kang, M. Making full use of hyperspectral data for gross primary productivity estimation with multivariate regression: Mechanistic insights from observations and process-based simulations. *Remote Sens. Environ.* **234**, 111435 (2019).
54. Wieneke, S. *et al.* Airborne based spectroscopy of red and far-red sun-induced chlorophyll fluorescence: Implications for improved estimates of gross primary productivity. *Remote Sens. Environ.* **184**, 654–667 (2016).
55. Scher, C. L. *et al.* Application of remote sensing technology to estimate productivity and assess phylogenetic heritability. *Appl. Plant Sci.* **8**, 1–14 (2020).
56. Song, C., Dannenberg, M. P. & Hwang, T. Optical remote sensing of terrestrial ecosystem primary productivity. *Prog. Phys. Geogr.* **37**, 834–854 (2013).
57. Yu, T. *et al.* Estimation of global vegetation productivity from Global LAnd Surface Satellite data. *Remote Sens.* **10**, 327–346 (2018).
58. Ciampalini, A. *et al.* Improved estimation of soil clay content by the fusion of remote hyperspectral and proximal geophysical sensing. *J. Appl. Geophys.* **116**, 135–145 (2015).
59. Carvalhais, N. *et al.* Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature* **514**, 213–217 (2014).
60. Fellows, A. W., Flerchinger, G. N., Seyfried, M. S., Lohse, K. A. & Patton, N. R. Controls on gross production in an aspen–sagebrush vegetation mosaic. *Ecohydrology* **12**, 1–14 (2019).
61. Flerchinger, G. N., Fellows, A. W., Seyfried, M. S., Clark, P. E. & Lohse, K. A. Water and Carbon Fluxes Along an Elevational Gradient in a Sagebrush Ecosystem. *Ecosystems* **23**, 246–263 (2020).
62. Seyfried, M., Link, T., Marks, D. & Murdock, M. Soil temperature variability in complex terrain measured using fiber-optic distributed temperature sensing. *Vadose Zo. J.* **15**, vzj2015.09.0128 (2016).
63. Seyfried, M. *et al.* Slope and aspect controls on soil climate: Field documentation and implications for large-scale simulation of critical zone processes. *Vadose Zo. J.* **10**, 1–17 (2021).
64. Harden, J. W. *et al.* Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter. *Glob. Change Biol.* **24**, e705–e718 (2018).
65. Paustian, K. *et al.* Climate-smart soils. *Nature* **532**, 49–57 (2016).
66. Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
67. Zennaro, F. *et al.* Exploring machine learning potential for climate change risk assessment. *Earth-Science Rev.* **220**, 103752 (2021).
68. Keiser, A. D., Knoepp, J. D. & Bradford, M. A. Disturbance decouples biogeochemical cycles across forests of the southeastern US. *Ecosystems* **19**, 50–61 (2016).
69. Waring, B. G., Adams, R., Branco, S. & Powers, J. S. Scale-dependent variation in nitrogen cycling and soil fungal communities along gradients of forest composition and age in regenerating tropical dry forests. *New Phytol.* **209**, 845–854 (2016).
70. Kyker-Snowman, E., Wieder, W., Frey, S. & Grandy, A. S. Stoichiometrically coupled carbon and nitrogen cycling in the MIcrobial-MIneral Carbon Stabilization model (MIMICS-CN). *Geosci. Model Dev. Discuss.* **10**, 1–32. https://doi.org/10.5194/gmd-2019-320 (2019).
71. Wang, Y. P. *et al.* Microbial activity and root carbon inputs are more important than soil carbon diffusion in simulating soil carbon profiles. *J. Geophys. Res. Biogeosci.* **126**, 1–19 (2021).
72. Pansze, A. J. *Geology and Ore Deposits of the Silver City-De Lamar-Flint Region, Owyhee County, Idaho* (Idaho Bureau of Mines and Geology, 1975).

73. Mclntyre, D. H. Cenozoic geology of the Reynolds Creek experimental watershed Owyhee County, Idaho. *Idaho Bur. Mines Geol. Pam.* **151**, 115 (1972).
74. Patton, N. R., Lohse, K. A., Godsey, S. E., Crosby, B. T. & Seyfried, M. S. Predicting soil thickness on soil mantled hillslopes. *Nat. Commun.* **9**, 1-10 (2018).
75. Patton, N. R., Lohse, K. A., Seyfried, M. S. & Murdock, M. D. Dataset for Soil Properties Determined at the Reynolds Creek Experimental Watershed (RCEW), Idaho [Idaho State University Collections 2010–2016]. (2018).
76. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: http://websoilsurvey.sc.egov.usda.gov. Accessed January 10, 2021.
77. Will, R. M. et al. Reynolds Creek--A Collection of Near-Surface Soil Organic Carbon (SOC) Maps, GIS/Map Data (2017). (2017).
78. Delvinne, H. H. Temperature Impacts on Soil Organic Carbon Decomposition Across an Environmental Gradient in a Semi-Arid Ecosystem. (2016).
79. Shi, Z., Crowell, S., Luo, Y. & Moore, B. Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nat. Commun.* **9**, 1–11 (2018).
80. Betancourt, M. The Convergence of Markov Chain Monte Carlo Methods: From the Metropolis Method to Hamiltonian Monte Carlo. *Ann. Phys.* **531**, 1–6 (2019).
81. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Available online at https://sdmdataaccess.sc.egov.usda.gov. Accessed January 10, 2021.
82. Seyfried, M. S., Harris, R. C. & Marks, D. A geographic database for watershed research, Reynolds Creek Experimental Watershed, Idaho, USA. U.S. Department of Agriculture, Agricultural Research Service, 800 Park Blvd, Suite 105 Boise, ID 83712–7716. Database (2000).
83. ESRI (2011). ArcGIS Desktop: Release 10. Environmental Systems Research Institute.
84. Godsey, S. E. *et al.* Eleven years of mountain weather, snow, soil moisture and streamflow data from the rain-snow transition zone—The Johnston Draw catchment, Reynolds Creek Experimental Watershed and Critical Zone Observatory, USA. *Earth Syst. Sci. Data* **10**, 1207–1216 (2018).
85. Wiesmair, M., Feilhauer, H., Magiera, A., Otte, A. & Waldhardt, R. Estimating vegetation cover from high-resolution satellite data to assess grassland degradation in the Georgian Caucasus. *Mt. Res. Dev.* **36**, 56–65 (2016).
86. Miller, G. J., Morris, J. T. & Wang, C. Estimating aboveground biomass and its spatial distribution in coastal wetlands utilizing planet multispectral imagery. *Remote Sens.* **11**, 15 (2019).
87. Xue, J. & Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sensors* **2017**, 1–18 (2017).
88. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Acknowledgements

This study was conducted in collaboration and cooperation with the USDA Agricultural Research Service, Northwest Watershed Research Center, Boise, Idaho, and landowners within the Reynolds Creek Experimental Watershed and Critical Zone Observatory (RCEW-CZO). The Reynolds Creek Experimental Watershed is supported by the USDA Long-Term Agroecosystem Research (LTAR) network. Support for this research was provided by the U.S. National Science Foundation (NSF) via RCEW-CZO Cooperative agreement NSF EAR-1331872 (Lohse, Seyfried) and NSF EAR-2012878. DP and WRW were supported by USDA-NIFA-AFRI 2020-67019-31395. WRW was supported by NSF ARCSS-2031238 and NSF DEB-1926413.

Author contributions

K.A.L., M.S.S. conceived and funded project; D.P., W.R.W., K.A.L. designed the methods and modeling approach; N.R.P., R.W., K.A.L., M.G. collected and analyzed soil observation data; GF collected eddy covariance data; J.F. analyzed MSAVI2; DP analyzed soil temperature data, part of which was collected by M.S.S.; DP performed modeling analyses and developed the MCMC, MC and spatial extrapolation algorithms; M.G. and K.G. contributed to study design and edited the manuscript. DP wrote the manuscript with K.A.L., W.RW. All reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-14224-8.

Correspondence and requests for materials should be addressed to D.P.

Reprints and permissions information is available at www.nature.com/reprints.

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