Received: 26 May 2021 | Revised: 22 February 2022 | Accepted: 24 February 2022

DOI: 10.1111/geb.13489

RESEARCH ARTICLE

Global Ecology and Biogeography A Journal of Macroecology WILEY

Continental United States may lose 1.8 petagrams of soil organic carbon under climate change by 2100

Sagar Gautam[1,2] | Umakant Mishra[1,2] | Corinne D. Scown[2,3,4,5] | Skye A. Wills[6] | Kabindra Adhikari[7] | Beth A. Drewniak[8]

[1]Bioscience Division, Sandia National Laboratory, Livermore, California, USA

[2]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California, USA

[3]Energy Analysis & Environmental Impact Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[5]Energy & Biosciences Institute, University of California, Berkeley, California, USA

[6]National Soil Survey Center, USDA-NRCS, Lincoln, Nebraska, USA

[7]Grassland, Soil and Water Research Laboratory, USDA-ARS, Temple, Texas, USA

[8]Environmental Science Division, Argonne National Laboratory, Lemont, Illinois, USA

Correspondence
Sagar Gautam, Bioscience Division, Sandia National Laboratory, Livermore, CA, USA.
Email: sgautam@lbl.gov

Funding information
Office of Biological and Environmental Research, Grant/Award Number: DE-AC02-05CH11231

Handling Editor: Bonnie G. Waring

Abstract

Aims: High-resolution information on soils' vulnerability to climate-induced soil organic carbon (SOC) loss can enable environmental scientists, land managers, and policy makers to develop targeted mitigation strategies. This study aims to estimate baseline and decadal changes in continental US surface SOC stocks under future emission scenarios.

Location: Continental United States.

Time period: 2014–2100.

Methods: We used recent SOC field observations (n = 6,213 sites), environmental factors (n = 32), and an ensemble machine learning (ML) approach to estimate baseline SOC stocks in surface soils across the continental United States at 100-m spatial resolution, and decadal changes under the projected climate scenarios of Coupled Model Intercomparison Project Phase Six (CMIP6) earth system models (ESMs).

Results: Baseline SOC projections from ML approaches captured more than 50% of variability in SOC observations, whereas ESMs represented only 6–16% of observed SOC variability. ML estimates showed a mean total loss of 1.8 Pg C from US surface soils under the high-emission scenario by 2100, whereas ESMs showed no significant change in SOC stocks with wide variation among ESMs. Both ML and ESM predictions agree on the direction of SOC change (net emissions or sequestration) across 46–51% of continental US land area. These differences are attributable to the high-resolution site-specific data used in the ML models compared to the relatively coarse grid represented in CMIP6 ESMs.

Main conclusions: Our high-resolution estimates of baseline SOC stocks, identification of key environmental controllers, and projection of SOC changes from US land cover types under future climate scenarios suggest the need for high-resolution simulations of SOC in ESMs to represent the heterogeneity of SOC. We found that the SOC change is sensitive to key soil related factors (e.g. soil drainage and soil order) that have not been historically considered as input parameters in ESMs, because currently more than 95% variability in the SOC of CMIP6 ESMs is controlled by net primary productivity, temperature, and precipitation. Using additional environmental factors to estimate the baseline SOC stocks and predict the future trajectory of SOC change can provide more accurate results.

KEYWORDS
climate, earth system model, environmental factor, future projection, machine learning, soil organic carbon

1 | INTRODUCTION

Soil organic carbon (SOC) constitutes the largest fraction of the terrestrial carbon cycle and can play an important role in buffering increases in atmospheric carbon dioxide (CO_2) concentrations (Paustian et al., 2016; Zimov et al., 2006). There is scientific consensus on the need to avoid and reverse SOC loss to enable sustainable land management, mitigate climate change, increase food security, and preserve biodiversity (Bradford et al., 2019; Lal, 2004; Paustian et al., 2016). Soil-based initiatives can mitigate the impacts of climate and land use changes, but implementation must be informed by an adequate understanding of the current and future trajectory of SOC change. However, the feedbacks between SOC stocks and the changing climate are still a major source of uncertainty in terrestrial carbon dynamics (Bellamy et al., 2005; Wiesmeier et al., 2019). Therefore, quantifying the magnitude and spatial heterogeneity of SOC vulnerability in future climate scenarios is a priority.

Earth system models (ESMs) simulate Earth's climate as a function of interactions between the atmosphere, ocean, land, sea ice, and the biosphere, and project future changes in atmospheric greenhouse gas concentrations and associated climate changes at global scale (Flato, 2011). ESMs simulate SOC change and project carbon–climate feedbacks at coarse spatial resolutions (*c*. 100-km spatial scales; Eyring et al., 2016). More recently, machine learning (ML) models have been used to predict the future trajectory of SOC stocks using the climate forcing data from ESMs (Hugelius et al., 2020; T. Wang et al., 2020). Efforts to build data-informed ESMs are in progress due to their potential to capitalize upon the advancements in data assimilation and ability to learn from diverse data sources (Schneider, 2020).

Quantifying the effect of future climate on SOC stocks is a key research challenge; high resolution SOC change estimates can help identify hotspots for land use intervention needed to maintain SOC levels in soils to buffer the adverse impact of future climate changes. The magnitude of SOC change in soil depends on carbon inputs, SOC decomposition, and lateral fluxes from soil (i.e., erosion and runoff loss), and the resulting net balance of SOC changes are soil- and site-specific (Z. Luo et al., 2019). In ESMs, SOC change is estimated by using the balance between net primary productivity (NPP) and SOC decomposition. SOC decomposition in ESMs is represented by a first-order linear decomposition model, in which decomposition of SOC increases exponentially with increase in temperature using a Q10 (factor by which soil respiration increases when temperature is increased of 10°C) function (Lloyd & Taylor, 1994). The moisture dependence of SOC is represented as a function of water potential (limiting decomposition at low soil moisture) and oxygen stress, which can limit SOC decomposition when the soil is saturated (Todd-Brown et al., 2013).

Changes in the environmental controllers of SOC dynamics – temperature, precipitation, net primary productivity, and evapotranspiration – have been predicted under projected future emission scenarios (Kharin et al., 2013; Melillo et al., 1993). These changes can result in potential acceleration or deceleration of SOC emissions in the future, affecting atmospheric greenhouse gas concentrations. Current ESM baseline projections of SOC stocks and their environmental controllers are not consistent with observations (Mishra et al., 2017) and, as a result, there is a great deal of uncertainty in predicting future carbon–climate feedbacks (Friedlingstein et al., 2014). Therefore, spatio-temporal estimates of SOC stocks and their environmental controllers that are consistent with observations are needed to improve ESM representations of SOC dynamics (Y. Luo et al., 2016). Currently, the complex nature of ESMs, their coarse spatial resolution and their associated computational cost are some of the limitations for impact assessment of SOC under future changes. Due to these limitations, new ESM initiatives such as the Climate Modeling Alliance (CliMA) have been working to improve ESMs by leveraging advances in data science and machine learning (Schneider, 2020).

Field observations of SOC and data on environmental controllers are valuable for quantifying SOC change and validating model predictions. ML approaches can use these field observations as training data to develop both linear and nonlinear relationships among a large number of environmental factors and SOC observations for better predictions of SOC stocks and thus can help to benchmark environmental control representation in ESMs (Mishra et al., 2020). Earlier studies estimated continental US SOC stocks based on mean SOC stock from the State Soil Geographic (STATSGO) database (Guo et al., 2006) and a geographically weighted regression of observations from the Rapid Carbon Assessment database (Gonçalves et al., 2021; West et al., 2013). In contrast to previous studies, our approach uses (a) the most recently available SOC stock field observation data from the United States, (b) a large set of environmental controller data, and (c) an ensemble ML approach that has been documented to produce more accurate results (Mishra et al., 2020). ML methods provide an opportunity to improve our understanding of SOC stocks by using high resolution data and to better understand and address the limitations of existing models that rely on a substantial number of input parameters with user-defined values. In addition, this kind of ESM benchmarking exercise will help improve and benchmark existing ESMs (Collier et al., 2018; Y. Q. Luo et al., 2012; Mishra et al., 2017; Mishra & Riley, 2014).

In this study, we focused on understanding the magnitude and location of SOC changes in continental US topsoils (surface to 0.3 m depth) because these soils are the most vulnerable to human intervention and climate extremes (Minasny et al., 2017). Using an ensemble of three ML approaches and future climate forcing data from multiple CMIP6 models, we predicted high-resolution (100 m) baseline SOC stocks for the continental United States and projected decadal changes in SOC stocks for 2100. We compared our baseline and decadal SOC changes with CMIP6 ESM projections. The specific objectives of our study were to (a) predict the baseline continental US SOC stocks and compare their representation in CMIP6 models, (b) predict the decadal changes in SOC stocks under future emission scenarios, and (c) quantify the impact of projected changes in climatic factors on SOC stocks across different land cover types within the continental United States.

2 | MATERIALS AND METHODS

2.1 | Study area, SOC data, and environmental controllers

We used 6,213 field observations of topsoil (0.3 m) SOC from the Rapid Carbon Assessment (RaCA) project, which collected balanced samples across different land cover types, soil types, and ecological regions of the continental United States. At each site soil samples were collected from different genetic soil horizons to compute SOC concentration and bulk density using the Soil Survey Laboratory Methods Manual (Burt, 2004). Soil organic carbon (kg/m^2) was calculated based on SOC concentration (SOCC, g/kg), soil layer depth (*D*, cm), bulk density (BD, g/cm^3), and volumetric fraction of the coarse fragments (*CF*) using the following formulation:

$$\mathrm{SOC} = \left[(\mathrm{SOCC} \times \mathrm{BD} \times \mathrm{D}) \times \left(1 - \frac{\mathrm{CF}}{100} \right) \right] \quad (1)$$

Details on the sampling and analysis on the RaCA methodology can be found in Wills et al. (2014). The calculated SOC stock values were highly skewed, and therefore were transformed based on natural log function, and then back-transformed for final SOC stock calculations. We used spatially balanced split samples of SOC for training (70%) and model testing (30%; Supporting Information Figure S1). Multiple environmental factors, which represented major soil-forming factors (Jenny, 1983), were used to predict the SOC stocks between points. The environmental data used in this study can be broadly divided into four categories: (a) climate (temperature, precipitation, and potential evapotranspiration); (b) soil (soil order, soil drainage class, soil moisture); (c) topography (digital elevation model, slope related parameters); and (d) vegetation (land use land cover, potential vegetation cover, spectral bands, vegetation index, net primary productivity, and ecoregion). The details of all the environmental data, spatial resolution, time span and source are presented in Supporting Information Table S1. All the environmental variables were re-gridded to a 100-m spatial resolution.

2.2 | Machine learning models

Machine learning has been used as a useful tool for soil carbon prediction (Adhikari et al., 2019; Padarian et al., 2019; Wiesmeier et al., 2011). Here we employed ML models using spatially referenced data on climate, soil, topography, vegetation, and SOC measurements to predict SOC stocks at 100-m spatial resolution. Gradient boosting machine (GBM), random forest (RF), and eXtreme gradient boosting (XGB) methods were used to predict the surface SOC stocks across the continental United States. Predictions from the ML models and their ensemble mean were compared with the SOC stocks predicted by the ESM models and their ensemble mean for baseline and decadal changes. The space-for-time substitution approach was used for the future simulations. This approach uses the future environmental data to predict SOC stocks for future time periods. In order to identify key controls of SOC in current ESM models, we built ML models using ESM's environmental data and SOC datasets.

The GBM algorithm was originally proposed by Friedman (2001) to solve the regression problem by gradually reducing the error. The GBM model in this study is specified as SOC ~ climate + soil + topography + vegetation + ε, where $\varepsilon \sim N(0, \sigma^2)$, meaning errors are normally distributed with 0 mean and a variance of σ^2. GBM works on the concept of iterative learning by fitting a simple regression predictor of the data, calculating the error residuals, and learning a new model by fitting the next decision tree to the residual of the previous iteration. Over the iteration the accuracy and complexity of the GBM model increase. GBM uses a simple regression model, 'weak learners', and iteratively combines this simple model to obtain a 'strong learner' with improved accuracy by reducing the bias and the variance. RF is a tree-based modelling approach that uses aresponse variable (here SOC) to grow multiple decision trees (n_{tree}) with multiple randomized predictors (here environmental controllers) at each node (m_{try}; Breiman, 2001). As the name suggests, the RF creates the forest with multiple decision trees. In general, more trees in the forest result in robust prediction and higher accuracy. RF works on the rationale that the combination of learning models increases prediction accuracy. Multiple decision trees are built using algorithms such as information gain and the Gini index approach. To classify a new object based on attributes, each tree gives a classification and votes for that class. Finally, RF chooses the class with most votes among all the trees in the forest, taking the average of output by different trees in the forest. The XGB algorithm is based on classification and regression (Chen et al., 2015). It is the extended version of GBM, which iteratively combines the weak learners to obtain the strong learner. XGB uses a more structured model framework to control overfitting and to ensure computational efficiency. For a RF model setup, the node of each tree splits based on the best environmental covariates, which are randomly selected in subsets. The number of environmental covariates in each random subset (m_{try}) in the model and the number of trees in the forest (n_{tree}) are other user-defined parameters in addition to the selection of the environmental covariates. The m_{try} and n_{tree} values of 5 and 500, respectively, were based on lowest out-of-bag errors; these were determined by using the optimization from the R package CARAT. For the GBM and XGB models, the tuning of the parameters including the number of trees and tree depth was based on the best predictive model using the root mean square error (RMSE) function. A tree number of 150 and a tree depth of 3 were the optimized values for the best predictive model. The model training and testing were done using the ML packages *randomforest*, *gbm*, and *caret* (R Core Team, 2012). For future projections at 100-m resolution over the continental United States for multiple time steps, the ML predictions were conducted using multiple nodes at the Laboratory Computing Resource Center at Argonne National Laboratory in parallel environment using R (R Core Team, 2012). The high-resolution projection (100 m) was done for eight time steps, four climate models with three realizations, and two emission scenarios using three ML models. ML model fitting was done by using spatially balanced training and testing SOC observation datasets. Model parameters were optimized by using the 10-fold cross-validation technique. Seventy percent of the datasets were used to train the model parameters to yield the best model fit with the lowest RMSE and highest r^2. The trained model was used to test the predictivity of models using test datasets and projection was done spatially using the spatial grids of environmental variables.

2.3 | ESMs and emission scenarios

We evaluated the soil carbon projection in six CMIP6 ESMs (Supporting Information Table S2). We downloaded and aggregated the SOC, precipitation, NPP and temperature data from ESM projections. The ESM results were compared with the ML model results for the continental United States. The majority of models used in ESMs were designed to simulate the soil carbon for topsoil depth (Kelly et al., 1997; Todd-Brown et al., 2013, 2014). The ESMs consider multiple carbon pools and the decomposition of the carbon in each pool and transfer between pools are based on CENTURY ecosystem models, which are also designed to simulate soil carbon in topsoil (Parton et al., 1988). In this study, we downloaded the SOC stock (cSoil) from the CMIP6 outputs (Eyring et al., 2016; O'Neill et al., 2016) without a specific lower boundary for soil carbon representation. We found that the mean soil carbon stock for the continental United States from the CMIP6 models $\left(69.40^{+22.45}_{-38.11}\,\mathrm{Pg\,C}\right)$ was comparable with the mean baseline topsoil SOC stock $\left(64.2^{+0.4}_{-1.7}\,\mathrm{Pg\,C}\right)$ estimated in this study. We fitted the ML models to predict the baseline SOC (2014) stocks using the observed historic environmental control data (Supporting Information Table S1). We used the 30-year average value to represent the baseline climate variable; other variables including topographical, soil and land use variables are represented as points in time. Details of the time span for different environmental control datasets are presented in Supporting Information Table S1. Data on future SOC projections are based on environmental control datasets from two CMIP6 shared socioeconomic pathway (SSP)-radiative forcing scenarios (SSP2 4.5, representing middle-of-the-road SSP with 4.5 W/m^2 radiative forcing, and SSP5 8.5, fossil-fuelled developed SSP with 8.5 W/m^2 radiative forcing; Eyring et al., 2016). For future simulations, we averaged the annual total precipitation, temperature and NPP values using 10-year data of each decade to represent future precipitation, temperature and NPP trends. The space-for-time substitution approach was used to map SOC stock in the continental United States under future climate scenarios by substituting baseline data with future temperature, precipitation and NPP data from each of the CMIP6 ESMs and results are presented as the ensemble mean. Although space-for-time-substitution approach has its limitations, this approach has widely been used in the literature to predict SOC changes with the assumption that the observed relationships between SOC and environmental factors can be extrapolated to the future (Chadburn et al., 2017; Gonçalves et al., 2021; Heuvelink et al., 2021; Koven et al., 2017; Varney et al., 2020; B. Wang et al., 2021).

3 | RESULTS AND DISCUSSION

3.1 | Environmental controls of soil organic carbon

With the observed SOC datasets for the continental United States (Supporting Information Figure S1) and a large dataset of environmental factors, we used ML models – random forest (RF), gradient boosting machine (GBM), and eXtreme gradient boosting (XGB) – to rank the relative importance of environmental factors (climate, soil, topography, and vegetation parameters) regulating the spatial pattern and magnitude of surface SOC stocks. The statistical properties of the surface SOC stocks used for training and testing are presented in Supporting Information Figure S1. The average SOC stock of continental US surface soil was 9.5 kg/m^2, ranging from 0.06 to 127 kg/m^2. The observed SOC stocks showed unimodal (kurtosis = 24.7) and positively skewed (coefficient of skewness = 4.3) distributions. All three ML models showed similar sensitivity ranking for different environmental factors in predicting continental US surface SOC stocks. Based on the increase in the residual sum of squares, mean annual precipitation had the largest influence on predicting the SOC distribution over the continental United States (Figure 1). Annual precipitation explained 21–34% of the total variation of SOC stocks (Figure 1) across different ML approaches. Potential evapotranspiration and soil drainage conditions were the second group of environmental variables controlling spatial distribution of SOC stocks with relative contributions of 10–17% (Figure 1). Mean annual temperature, net primary productivity, land cover types, and soil moisture were found to explain about 3–9% of total variation of SOC stocks. We also

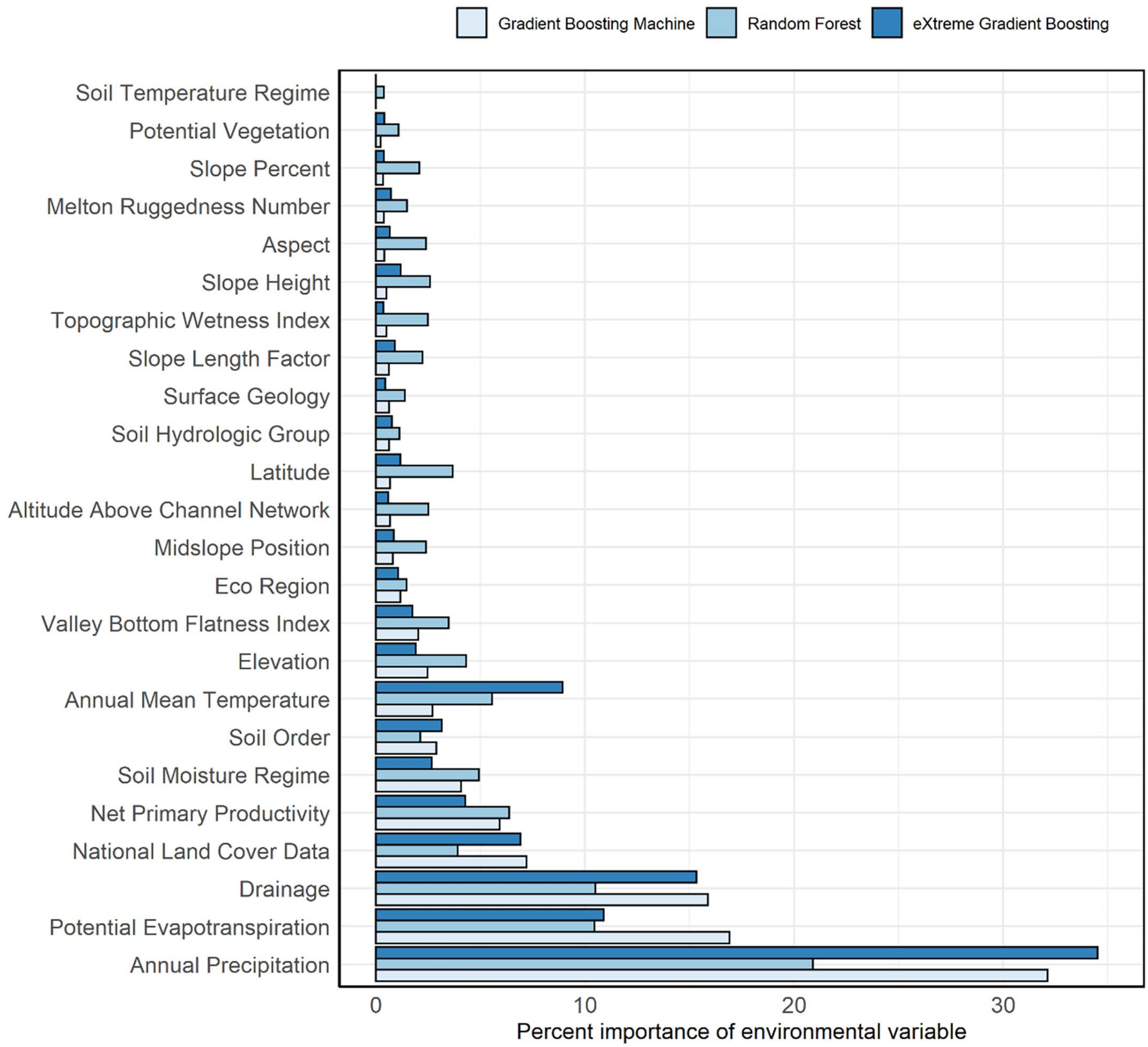

FIGURE 1 Variable importance of environmental predictors in the machine learning (ML) models. The variable importance is scaled to represent percentage importance of each variable

investigated the spatial correlation of surface SOC stocks with environmental factors and found that mean annual precipitation, net primary productivity, potential evapotranspiration, mean annual temperature, and soil drainage class were significantly correlated with continental US surface SOC stocks (spatial average r = .50, .53, −.59, −.40, and .25, respectively). As expected, the regional pattern of SOC stocks was positively associated with mean annual precipitation, net primary productivity, and soil drainage conditions, and negatively associated with temperature and potential evapotranspiration. When ML models were fitted to ESM inputs and ESM-predicted SOC stocks, more than 90% variability in ESM-predicted SOC was controlled by precipitation, NPP and temperature. In contrast, ML models based on observations showed higher importance of drainage class and soil order, which have not been historically incorporated in ESMs. Precipitation drives NPP and thus the carbon input to the soil system (Wiesmeier et al., 2019), and soil drainage conditions determine whether there will be anaerobic conditions in the soil; both are indicative of higher SOC storage at a location (Kumar et al., 2014).

Temperature affects the microbial decomposition of soil organic matter and soil water evaporation; both decrease SOC stocks (Jobbágy & Jackson, 2000; Koven et al., 2017; Lloyd & Taylor, 1994; Smith et al., 2005). A long-term (26-year) soil-warming experiment showed increased SOC decomposition with warming because of changes in the microbial population, mineralization, and soil respiration (Melillo et al., 2017). Earlier meta-analysis of experimental warming studies between 1990 and 2010 indicated around 1 ± 0.7% decrease in soil carbon under elevated air temperature of 1.8 ± 0.2 °C (Lu et al., 2013). A meta-analysis of the observed warming response to SOC and soil CO_2 flux revealed high variation, with experiments with higher increases in temperature showing greater soil respiration responses; however, the SOC loss was not significantly impacted by temperature (Sulman et al., 2018). A meta-analysis of warming experiments across the globe for different time-scales (3–6 years) and temperature changes (1–4 °C) indicated a statistically significant decrease in soil carbon (4.5%) due to warming (Yan et al., 2020). Note that the majority of these experimental studies were conducted by comparing a heated plot with a control plot under natural rainfall conditions.

Researchers have also developed empirical relationships based on meta-analysis of warming experiments to predict the soil carbon change which shows net SOC loss due to future warming (Crowther et al., 2016); however, this approach negates nonlinear relationships. A community land model (CLM) based prediction of warming impacts indicated that the underlying biochemical mechanism in current soil carbon model in CLM are insufficient to represent high latitude perturbation experiments (Bouskill et al., 2014). Evidence also suggests that soil respiration (and therefore soil carbon) is directly proportional to precipitation amount (Wu et al., 2011). Given the multiple drivers of SOC, predicting SOC with a single factor or linear function can cause large biases. ML models capable of capturing nonlinear relationships may improve the research community's mechanistic understanding of soil carbon change and provide a lightweight method for representing complex soil carbon dynamics in ESMs with reduced bias and uncertainty. Given the wide range of temperature and precipitation patterns found over the continental United States, our approach should capture part of the future climate trends. However, as the future climate trend is uncertain, there will be uncertainty in predicted SOC change for extreme values of climatic factors not accounted for by the baseline ML predictions. One of the main reasons for using multi-model ensemble method was to represent these sources of uncertainty and present the estimates with ranges.

3.2 | New estimates of continental US SOC stocks using ML models and comparison with CMIP6 ESMs

In this study, we quantified a spatially explicit baseline and projected decadal changes of topsoil SOC stocks in the continental United States by integrating recently available SOC observations with a large number of environmental datasets and ESM-projected precipitation, temperature and NPP data. The baseline surface SOC stock in the continental United States as estimated by ML approaches was $64.2^{+0.4}_{-1.7}$ Pg C (Figures 2a and 3b). The highest prediction accuracy was obtained from the RF model with a coefficient of determination of .57 and RMSE of 1.91 kg/m^2 (Figure 3a). GBM and XGB models resulted in a similar coefficient of determination (r^2 = .54 and .54, respectively) and RMSE (1.95 and 1.98 kg/m^2, respectively; Figure 3a).

The baseline SOC stocks represented in ESMs show different magnitudes and spatial distributions across different models; none of the ESMs we evaluated represented geographic or climatic gradients in the estimated baseline SOC stocks (Figure 2a,b). Individual ESM SOC projections showed diverse magnitudes of total SOC stocks and their spatial distributions across the continental United States (Figures 2b and 3b) and did not show significant correlations with environmental factors. The mean SOC stock estimated using the ensemble of CMIP6 ESMs was $69.40^{+22.45}_{-38.11}$ Pg C (Figure 3b).

Baseline SOC projections from ML approaches captured more than 50% of variability in SOC observations compared to CMIP6-ESMs, which represented around 6–16% of SOC variability (Figure 3a). The spatial distribution of baseline ML-predicted SOC stocks showed significant geographic and climatic control across US ecoregions, which are not represented in the ESM due to their coarse spatial resolution (100 km) and use of one set of assumptions to predict SOC. In contrast, ML model tends to learn from data and build relation based on multiple environmental controller data to predict SOC. ML-predicted SOC stocks were higher at the higher latitudes of the eastern United States and lower at the lower latitudes of the western United States. The estimates of spatial distribution of SOC stocks in this study are consistent with the SOC estimates generated by earlier studies (Guo et al., 2006; Kern, 1994). Previous continental US SOC stocks for top 1-m depth ranged from 57.2 to 78 Pg C (Bliss et al., 2014; Kern, 1994; Sundquist et al., 2009; Tao et al., 2020). Our SOC stock estimate is within the reported range and shows fine-scale spatial details in SOC distribution across different latitudes and geographic regions of the continental United States.

The large uncertainty range in baseline ESM SOC estimates demonstrates the need for land model benchmarking with additional environmental controls, which may also help to reduce uncertainty in the future projections. These large uncertainties in the CMIP6 ESMs are attributed to coarse grid resolution of model environemtal control datasets. In addition to the large uncertainty range, the magnitude and distribution of ESMs' SOC stock simulations were not consistent with baseline ML predictions, which further reduces confidence in these estimates (Z. Luo et al., 2017, 2019; Todd-Brown et al., 2013). However, our results show that the mean estimates of SOC stocks obtained from the ensemble of CMIP6-ESMs better represents the spatial pattern of SOC stocks obtained from the ensemble ML approach. The results suggest a need for improvement in the spatial representation of SOC stocks by including key observed environmental controllers in ESMs.

3.3 | Projected decadal changes in US surface SOC stocks across different land cover types

The mean ensemble estimates of ML models showed a decrease in future SOC stocks from continental US surface soils under the high (SSP5 8.5) emission scenario and no change under the moderate (SSP2 4.5) emission scenario (Figure 4a). At lower latitudes, the ML models predict a decrease in SOC stocks driven by faster carbon cycling (Trumbore, 1997); losses are higher in SSP5 8.5 with higher warming. In contrast, ESM predictions showed SOC

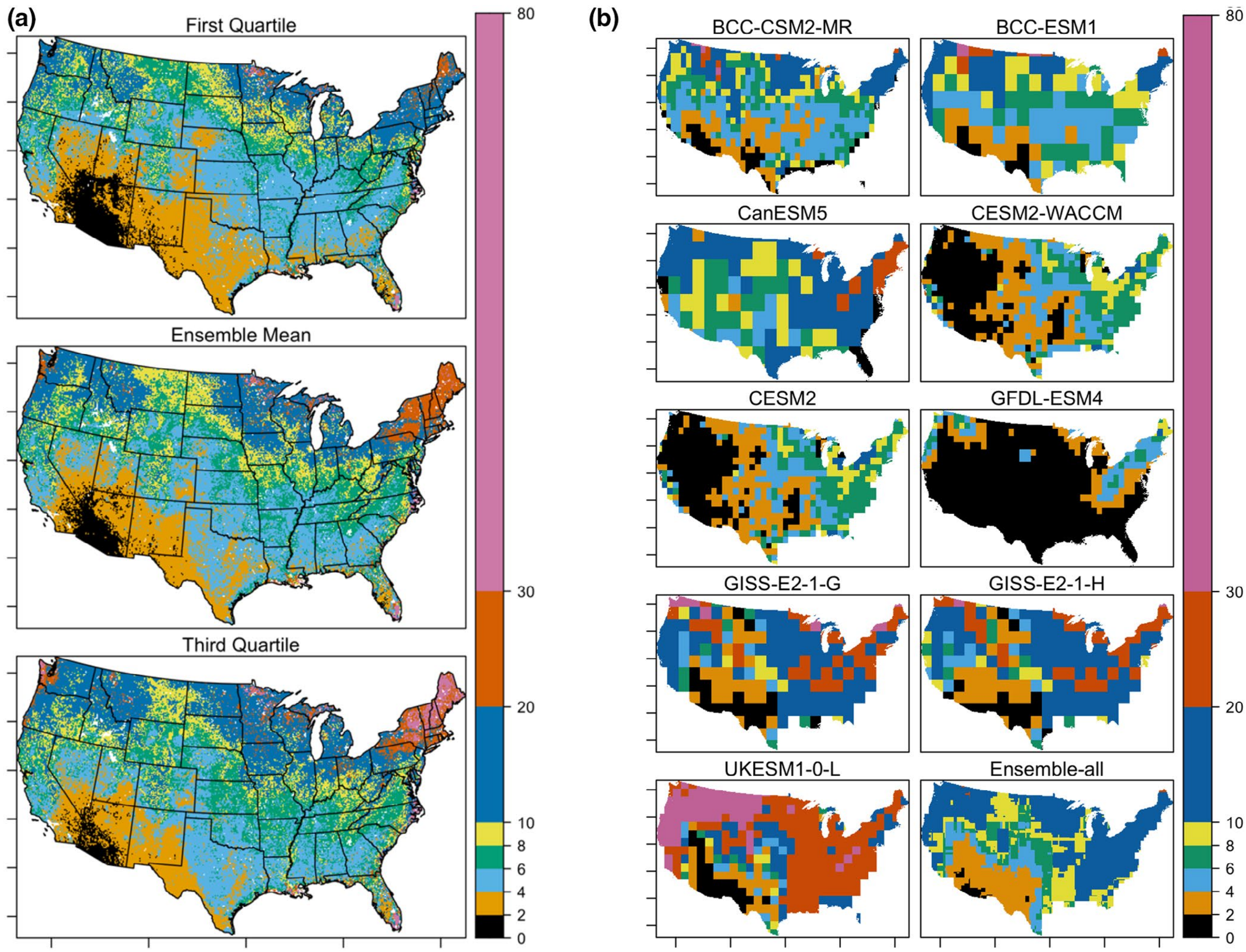

FIGURE 2 Comparison of machine learning (ML) and earth system models (ESMs). Ensemble first quartile, mean, and third quartile of predicted baseline soil organic carbon (SOC) stocks (kg/m^2 C) using (a) three ML models and environmental datasets and (b) baseline projection from multiple ESMs and their ensemble average

increases at lower latitudes (Supporting Information Figure S2). The estimates of decadal SOC stock change from multiple ESMs showed wide variability among models, and the ensemble of ESM projections showed no significant change in SOC stocks over time under both emission scenarios (Figure 4b). At higher latitudes, ML models predicted higher loss of SOC in area with higher SOC stock, which are also represented in ESM projections except for north-eastern states (Supporting Information Figure S2). The differences in SOC change predictions across space indicate different environmental controllers regulating the SOC change in ML predictions and ESMs (Ahlström et al., 2017). The ML models showed a cumulative 1.8 Pg C decrease in SOC stocks under the SSP5 8.5 emission scenario by 2100 compared to the baseline projection of 64.2 Pg C (Figure 4a). The temporal changes in SOC stocks under both emission scenarios were similar through 2050. Beyond 2050, the high-emission scenario showed a more rapid SOC loss from surface soils than the low-emission scenario, which showed no change in SOC (Figure 4a). The trajectory of SOC change predicted by the ML models differed based on future climate forcing data, and the extent of change differed depending upon the climate model and emission scenario (Gottschalk et al., 2012), resulting in a wide range in possible SOC loss. There was no consensus among the ESMs for the trajectory and the magnitude of SOC change across investigated emission scenarios. These results indicate need for common modelling assumptions across ESM and need for inclusion of key environmental controllers (Conant et al., 2011; Schmidt et al., 2011; Wieder et al., 2013). Since the SOC stock in soil is three times more than that in the atmosphere and terrestrial vegetation combined, current representation of SOC dynamics may have significant bias in representation of the carbon cycle and climate-carbon feedback in ESMs (Friedlingstein et al., 2014).

We compared the ensemble ML SOC change predictions with the ensemble ESM predictions to identify spatial locations where both models agree and disagree in the direction and magnitude of SOC change (Figure 5). The ML and ESM projections disagree on the direction (accumulation versus net emissions) of SOC change across 49% of the total land area in the United States under the SSP2 4.5 scenario, and this fraction increased slightly to 54%

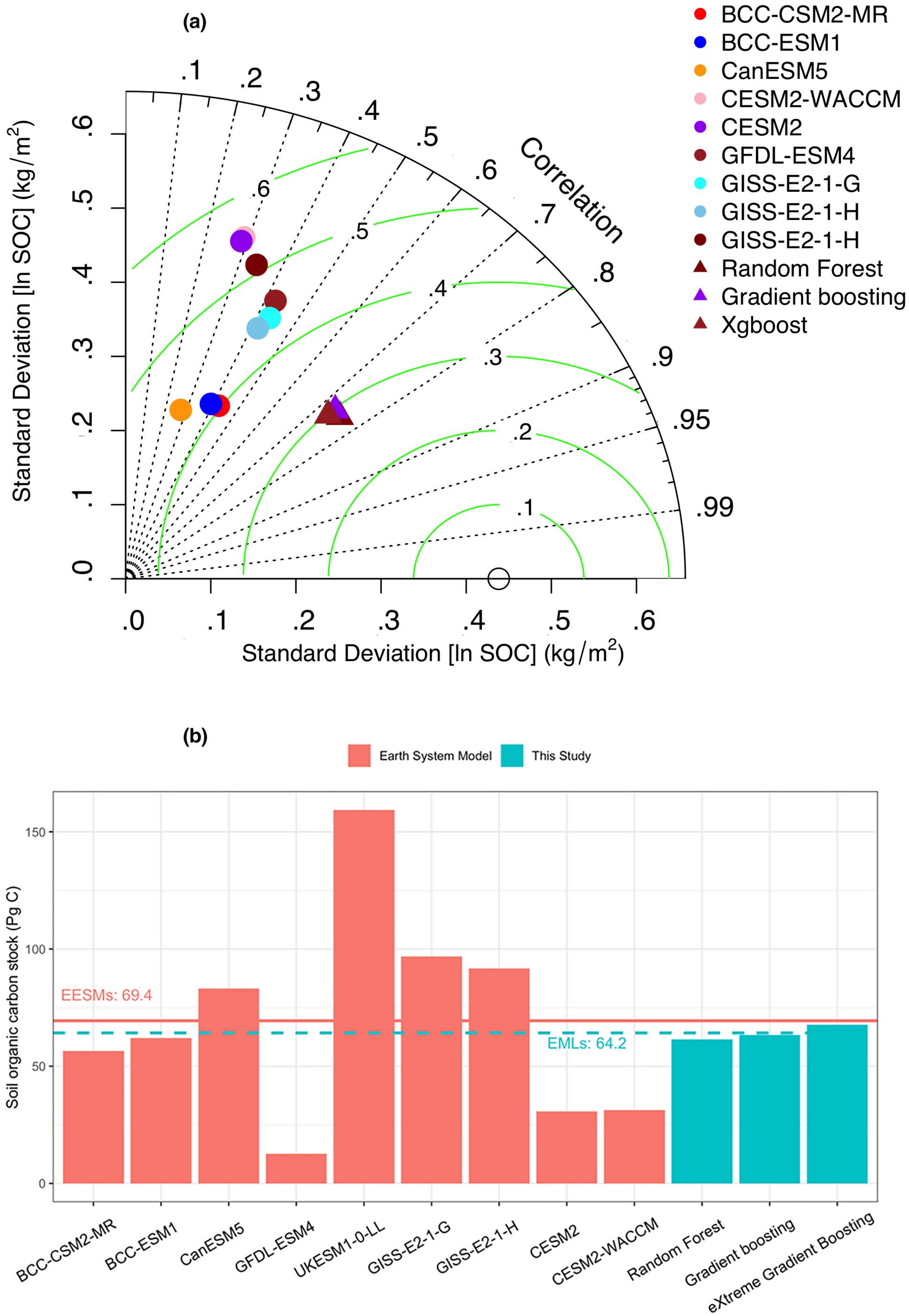
(a)
BCC-CSM2-MR
BCC-ESM1
CanESM5
CESM2-WACCM
CESM2
GFDL-ESM4
GISS-E2-1-G
GISS-E2-1-H
GISS-E2-1-H
Random Forest
Gradient boosting
Xgboost
Correlation
Standard Deviation [ln SOC] (kg/m^2)
Standard Deviation [ln SOC] (kg/m^2)
(b)
Earth System Model
This Study
Soil organic carbon stock (Pg C)
EESMs: 69.4
EMLs: 64.2
BCC-CSM2-MR
BCC-ESM1
CanESM5
GFDL-ESM4
UKESM1-0-LL
GISS-E2-1-G
GISS-E2-1-H
CESM2
CESM2-WACCM
Random Forest
Gradient boosting
eXtreme Gradient Boosting

FIGURE 3 Comparison of baseline soil organic carbon (SOC) stock simulated by nine earth system models (ESMs) and estimates from this study. (a) The Taylor diagram shows prediction quality of machine learning (ML) and ESM models for SOC stocks. (b) Total SOC stock; EESMs = ensemble mean of ESM models; EMLs = ensemble results from ML models

under the SSP5 8.5 emission scenario (Figure 5). The difference between the ML and the ESM projections of SOC change indicates the need for robust benchmarking of ESM projections of baseline SOC stocks using observed environmental controller datasets (Y. Luo et al., 2016; Mishra et al., 2017). The results from both ML and ESMs indicate vulnerability of SOC stocks at lower latitudes of the continental United States. Given the higher productivity of bioenergy crops in southern states (Gautam et al., 2020; Mishra et al., 2013), the bioenergy landscape in the vulnerable areas can help sequester soil carbon and to buffer the future impact of climate change under high-emission scenarios.

To better understand the underlying environmental factors causing agreement and disagreement between the ML and ESM model predictions, we plotted the environmental controllers for each region separately (Supporting Information Figure S3). The spatial locations where the ESM and ML models agree on SOC sequestration had lower mean annual temperature and precipitation, and higher elevation, compared to locations of disagreement and emissions (Supporting Information Figure S3). The locations where both ESM and ML models agree on net SOC decomposition had higher precipitation, NPP, and initial SOC concentrations compared to locations where the models either disagreed or both predicted net sequestration. There was some difference in predicted future SOC distribution depending on ESM climate forcing and ML models (Supporting Information Figures S4–S6), highlighting the need for multimodal ensemble analysis to better represent the prediction uncertainty. Future projection of decadal changes of SOC stocks using ML models showed higher carbon loss in the high-emission scenario because of changes in climatic factors. These ML model-predicted trajectories of SOC change represent the assumption of shared socioeconomic pathways and relative concentration pathways, where the atmospheric concentration of CO_2 changes with change in urbanization, forest cover, gross domestic product (GDP), education, and population, and starts diverging post-2050 between the two future scenarios (Riahi et al., 2017; Van Vuuren et al., 2011).

The ensemble mean ML model prediction showed more carbon loss under the high-emission scenario across different land cover types of the continental United States. The SOC stocks from forest, cultivated land, wetland, and other herbaceous land cover types changed by −282, 65, −181, and 352 Tg C, respectively, by 2100 under the SSP2 4.5 emission scenario (Supporting Information Table S3). The spatial distribution of the magnitude of SOC change is presented in Supporting Information Figure S2. Our projections based on ML models are consistent with the historic pattern, which shows net SOC loss over time over the continental United States (Sanderman et al., 2017). Under the SSP5 8.5 emission scenario, the SOC stocks from forest, cultivated land, wetland, and other herbaceous land cover types decreased by −357, −409, −266, and −384 Tg C, respectively, by 2100 (Supporting Information Table S3). Our results based on the ensemble ML model showed that the loss of SOC roughly doubles in forest and wetlands. For all the landuse type SOC loss increased under high-emission scenarios compared to low-emission scenarios. In contrast to the ML results, the ensemble of ESMs showed no significant change in the continental US surface SOC stocks (Supporting Information Table S3). ESM projections suggested future increases in SOC stocks in forest and herbaceous land cover types and decreases in SOC stocks from cultivated and wetlands land cover types (Supporting Information Table S3). The spatial patterns of SOC change from ESMs provided some contrasting results compared to our ML projections (Supporting Information Figure S3). For example, the ESM projections showing increases in SOC stocks at lower latitudes and in dry regions (south-west) are physically unexplainable.

Our study has some limitations and assumptions. Given the uncertainty in the projection of environmental variables in ESMs, we constrained some dynamic predictor variables (e.g., potential evapotranspiration, PET) to historic levels for future projections. Our assumption was based on the comparison of the observed and ESM baseline PET. The space-for-time substitution approach that we used may produce some unrealistic changes in SOC, when the future substituted environmental factors data are outside the range considered by ML models. The limitations of the space-for-time approach also include the assumption that the current SOC stocks are under steady states and the soil carbon will reach to new steady state under future warming. Further, this approach assumes that the observed interactions between SOC and environmental factors can be extrapolated to the future (Heuvelink et al., 2021).

In summary, results from our study indicate that SOC change is sensitive to environmental factors that have not historically been incorporated into ESMs or lost due to spatial averaging, which may have resulted in systematic underestimation of potential SOC loss across the continental United States. Using these factors to estimate current high-resolution SOC stocks and predict the future trajectory of SOC change can provide valuable results which are relevant for policy maker. The results generated indicate that the temporal and spatial representation of SOC stocks in ESMs is relatively static and governed by relatively few environmental factors in comparison to the ensemble ML approach. Results from our data model integration approach show loss of soil carbon in future decades and higher loss under a high-emission scenario. We observed bias in the ESMs prediction for projecting the magnitude and direction of SOC change over space and time compared to observations. ML proved to be a useful approach for benchmarking these results and identifying environmental factors that should be incorporated to improve future ESM predictions. Future research efforts should be directed toward achieving more realistic representation of environmental controls of SOC stocks in ESMs using hybrid machine and process models, and reducing the uncertainty in understanding the impacts of future climate extremes on the SOC stocks of different land cover types.

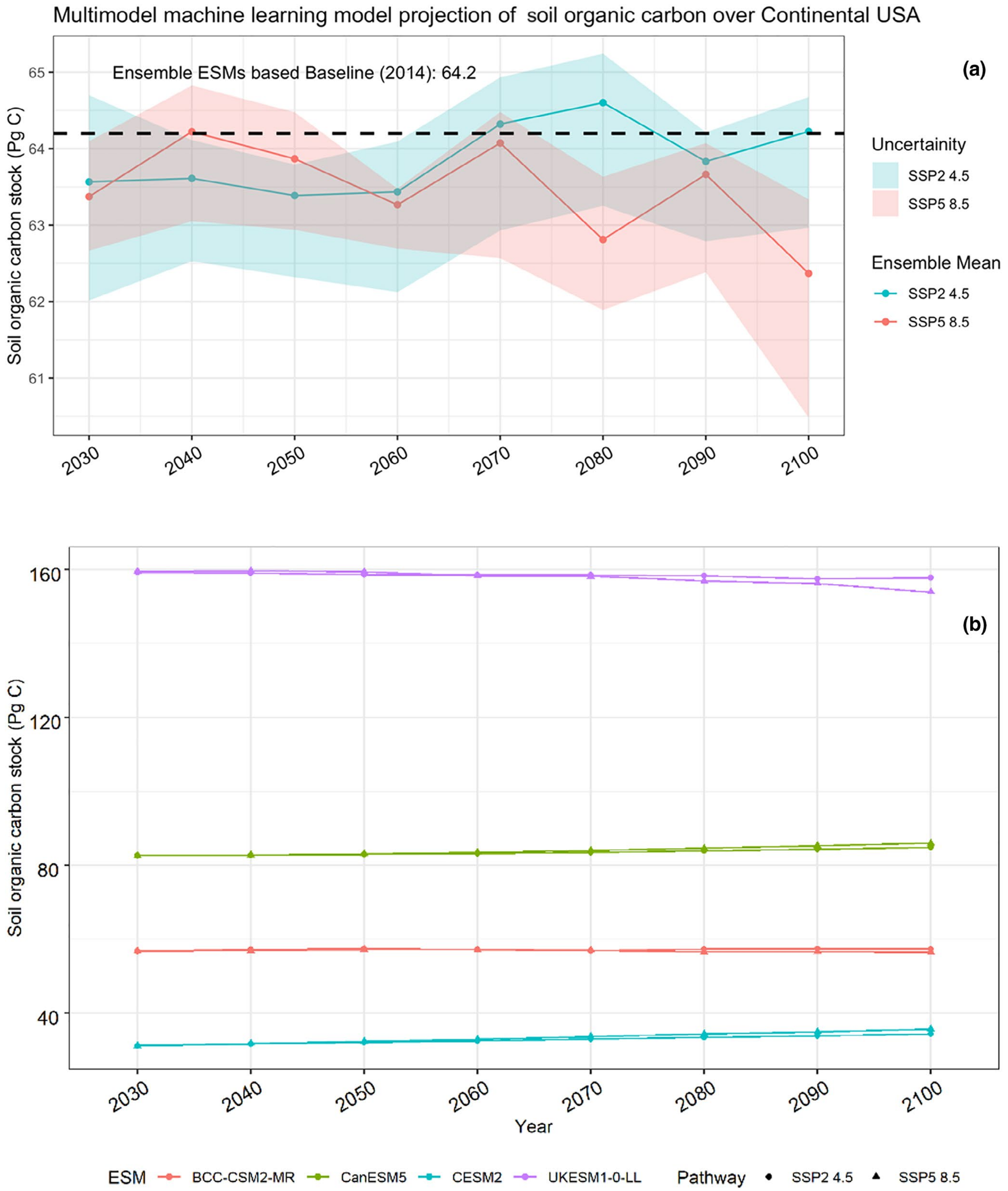

FIGURE 4 Comparison of decadal changes in soil organic carbon (SOC) stocks. (a) Machine learning (ML)-predicted decadal SOC changes. Lines show ensemble mean predictions, and shaded area shows uncertainty ranges as expressed by the difference between the first and third quartile values and (b) earth system model (ESM)-predicted decadal SOC change and its distribution among different ESMs. SSP = shared socioeconomic pathway

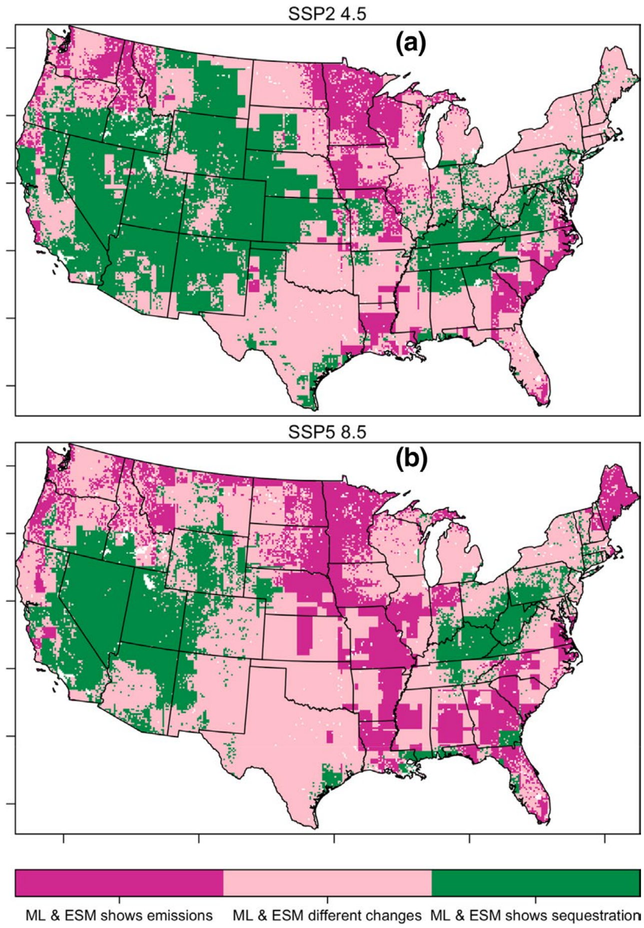

FIGURE 5 Direction of soil organic carbon (SOC) change. Comparison of change in SOC stocks by 2100 between machine learning (ML) and earth system models (ESMs). (a) and (b) compare two emission scenarios, SSP2 4.5 and SSP5 8.5, respectively, where SSP = shared socioeconomic pathway

ACKNOWLEDGMENTS

This work was part of the Department of Energy, Joint BioEnergy Institute (http://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. We gratefully acknowledge the computing resources provided on Bebop; a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory.

CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

S.G., U.M. and C.D.S. designed the study, S.G. conducted the analysis and prepared the manuscript draft, and all other coauthors assisted in discussion of the results and preparation of the manuscript.

DATA AVAILABILITY STATEMENT

The data and R script supporting the results of this study are available in Dryad at: https://datadryad.org/stash/share/CMMTOMgCLP2vA3Lj5a7NK2PBUcUGB1H19gmkGrZsx_M

ORCID

Sagar Gautam https://orcid.org/0000-0002-0828-1631
Corinne D. Scown https://orcid.org/0000-0003-2078-1126
Kabindra Adhikari https://orcid.org/0000-0002-1365-7015

REFERENCES

Adhikari, K., Owens, P. R., Libohova, Z., Miller, D. M., Wills, S. A., & Nemecek, J. (2019). Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. *Science of the Total Environment*, *667*, 833–845. https://doi.org/10.1016/j.scitotenv.2019.02.420

Ahlström, A., Schurgers, G., & Smith, B. (2017). The large influence of climate model bias on terrestrial carbon cycle simulations. *Environmental Research Letters*, *12*(1), 014004. https://doi.org/10.1088/1748-9326/12/1/014004

Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M., & Kirk, G. J. (2005). Carbon losses from all soils across England and Wales 1978–2003. *Nature*, *437*(7056), 245–248.

Bliss, N. B., Waltman, S. W., West, L. T., Neale, A., & Mehaffey, M. (2014). Distribution of soil organic carbon in the conterminous United States. In A. Hartemink, K. McSweeney (Eds), *Soil carbon*. Progress in Soil Science. Springer. https://doi.org/10.1007/978-3-319-04084-4_9

Bouskill, N., Riley, W., & Tang, J. (2014). Meta-analysis of high-latitude nitrogen-addition and warming studies implies ecological mechanisms overlooked by land models. *Biogeosciences*, *11*(23), 6969–6983. https://doi.org/10.5194/bg-11-6969-2014

Bradford, M. A., Carey, C. J., Atwood, L., Bossio, D., Fenichel, E. P., Gennet, S., Fargione, J., Fisher, J. R. B., Fuller, E., Kane, D. A., Lehmann, J., Oldfield, E. E., Ordway, E. M., Rudek, J., Sanderman, J., & Wood, S. A. (2019). Soil carbon science for policy and practice. *Nature Sustainability*, *2*(12), 1070–1072. https://doi.org/10.1038/s41893-019-0431-y

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Burt, R. (Ed.) (2004). *Soil survey laboratory methods manual: soil survey investigations report No. 42 Version 4.0*. United States Department of Agriculture, Natural Resources Conservation Service.

Chadburn, S., Burke, E., Cox, P., Friedlingstein, P., Hugelius, G., & Westermann, S. (2017). An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*, *7*(5), 340–344. https://doi.org/10.1038/nclimate3262

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., & Cho, H. (2015). *Xgboost: Extreme gradient boosting*. R Package Version 0.4-2, 1(4).

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., & Randerson, J. T. (2018). The International

Land Model Benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Conant, R. T., Ryan, M. G., Ågren, G. I., Birge, H. E., Davidson, E. A., Eliasson, P. E., Evans, S. E., Frey, S. D., Giardina, C. P., Hopkins, F. M., Hyvönen, R., Kirschbaum, M. U. F., Lavallee, J. M., Leifeld, J., Parton, W. J., Megan Steinweg, J., Wallenstein, M. D., Martin Wetterstedt, J. Å., & Bradford, M. A. (2011). Temperature and soil organic matter decomposition rates–synthesis of current knowledge and a way forward. *Global Change Biology*, *17*(11), 3392–3404. https://doi.org/10.1111/j.1365-2486.2011.02496.x

Crowther, T. W., Todd-Brown, K. E. O., Rowe, C. W., Wieder, W. R., Carey, J. C., Machmuller, M. B., Snoek, B. L., Fang, S., Zhou, G., Allison, S. D., Blair, J. M., Bridgham, S. D., Burton, A. J., Carrillo, Y., Reich, P. B., Clark, J. S., Classen, A.T., Dijkstra, F. A., Elberling, B., … Bradford, M. A. (2016). Quantifying global soil carbon losses in response to warming. *Nature*, *540*(7631), 104–108. https://doi.org/10.1038/nature20150

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Flato, G. M. (2011). Earth system models: An overview. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 783–800. https://doi.org/10.1002/wcc.148

Friedlingstein, P., Andrew, R. M., Rogelj, J., Peters, G. P., Canadell, J. G., Knutti, R., Luderer, G., Raupach, M. R., Schaeffer, M., van Vuuren, D. P., & Le Quéré, C. (2014). Persistent growth of CO 2 emissions and implications for reaching climate targets. *Nature Geoscience*, *7*(10), 709–715. https://doi.org/10.1038/ngeo2248

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*, 1189–1232. https://doi.org/10.1214/aos/1013203451

Gautam, S., Mishra, U., Scown, C. D., & Zhang, Y. (2020). Sorghum biomass production in the continental United States and its potential impacts on soil organic carbon and nitrous oxide emissions. *GCB Bioenergy*, *12*(10), 878–890. https://doi.org/10.1111/gcbb.12736

Gonçalves, D. R. P., Mishra, U., Wills, S., & Gautam, S. (2021). Regional environmental controllers influence continental scale soil carbon stocks and future carbon dynamics. *Scientific Reports*, *11*(1), 1–10. https://doi.org/10.1038/s41598-021-85992-y

Gottschalk, P., Smith, J. U., Wattenbach, M., Bellarby, J., Stehfest, E., Arnell, N., Osborn, T. J., Jones, C., & Smith, P. (2012). How will organic carbon stocks in mineral soils evolve under future climate? Global projections using RothC for a range of climate change scenarios. *Biogeosciences*, *9*(8), 3151–3171. https://doi.org/10.5194/bg-9-3151-2012

Guo, Y., Amundson, R., Gong, P., & Yu, Q. (2006). Quantity and spatial variability of soil carbon in the conterminous United States. *Soil Science Society of America Journal*, *70*(2), 590–600. https://doi.org/10.2136/sssaj2005.0162

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science*, *72*(4), 1607–1623. https://doi.org/10.1111/ejss.12998

Hugelius, G., Loisel, J., Chadburn, S., Jackson, R. B., Jones, M., MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M. B., Treat, C., Turetsky, M., Voigt, C., & Yu, Z. (2020). Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proceedings of the National Academy of Sciences*, *117*(34), 20438–20446. https://doi.org/10.1073/pnas.1916387117

Jenny, H. (1983). *Factors of soil formation: A system of quantitative pedology*. McGraw-Hill.

Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, *10*(2), 423–436.

Kelly, R. H., Parton, W. J., Crocker, G. J., Graced, P. R., Klír, J., Körschens, M., Poulton, P. R., & Richter, D. D. (1997). Simulating trends in soil organic carbon in long-term experiments using the century model. *Geoderma*, *81*(1–2), 75–90. https://doi.org/10.1016/S0016-7061(97)00082-7

Kern, J. S. (1994). Spatial patterns of soil organic carbon in the contiguous United States. *Soil Science Society of America Journal*, *58*(2), 439–455. https://doi.org/10.2136/sssaj1994.03615995005800020029x

Kharin, V. V., Zwiers, F., Zhang, X., & Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, *119*(2), 345–357. https://doi.org/10.1007/s10584-013-0705-8

Koven, C. D., Hugelius, G., Lawrence, D. M., & Wieder, W. R. (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

Kumar, S., Nakajima, T., Mbonimpa, E. G., Gautam, S., Somireddy, U. R., Kadono, A., Lal, R., Chintala, R., Rafique, R., & Fausey, N. (2014). Long-term tillage and drainage influences on soil organic carbon dynamics, aggregate stability and corn yield. *Soil Science and Plant Nutrition*, *60*(1), 108–118. https://doi.org/10.1080/00380768.2013.878643

Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *Science*, *304*(5677), 1623–1627.

Lloyd, J., & Taylor, J. (1994). On the temperature dependence of soil respiration. *Functional Ecology*, *8*, 315–323. https://doi.org/10.2307/2389824

Lu, M., Zhou, X., Yang, Q., Li, H., Luo, Y., Fang, C., Chen, J., Yang, X., & Li, B. O. (2013). Responses of ecosystem carbon cycle to experimental warming: A meta-analysis. *Ecology*, *94*(3), 726–738. https://doi.org/10.1890/12-0279.1

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., … Zhou, T. (2016). Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochemical Cycles*, *30*(1), 40–56. https://doi.org/10.1002/2015GB005239

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., Ciais, P., Dalmonech, D., Fisher, J. B., Fisher, R., Friedlingstein, P., Hibbard, K., Hoffman, F., Huntzinger, D., Jones, C. D., Koven, C., Lawrence, D., Li, D. J., Mahecha, M., … Zhou, X. H. (2012). A framework for benchmarking land models. *Biogeosciences*, *9*(10), 3857–3874. https://doi.org/10.5194/bg-9-3857-2012

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

Luo, Z., Wang, G., & Wang, E. (2019). Global subsoil organic carbon turnover times dominantly controlled by soil properties rather than climate. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-11597-9

Melillo, J. M., Frey, S. D., DeAngelis, K. M., Werner, W. J., Bernard, M. J., Bowles, F. P., Pold, G., Knorr, M. A., & Grandy, A. S. (2017). Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science*, *358*(6359), 101–105.

Melillo, J. M., McGuire, A. D., Kicklighter, D. W., Moore, B., Vorosmarty, C. J., & Schloss, A. L. (1993). Global climate change and terrestrial net primary production. *Nature*, *363*(6426), 234–240.

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant,

B. P., Martin, M., McConkey, B. G., Mulder, V. L., … Winowiecki, L. (2017). Soil carbon 4 per mille. *Geoderma*, *292*, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

Mishra, U., Drewniak, B., Jastrow, J. D., Matamala, R. M., & Vitharana, U. (2017). Spatial representation of organic carbon and active-layer thickness of high latitude soils in CMIP5 earth system models. *Geoderma*, *300*, 55–63. https://doi.org/10.1016/j.geoderma.2016.04.017

Mishra, U., Gautam, S., Riley, W., & Hoffman, F. M. (2020). Ensemble machine learning approach improves predicted spatial variation of surface soil organic carbon stocks in data-limited northern circumpolar region. *Frontiers in Big Data*, *3*, 40. https://doi.org/10.3389/fdata.2020.528441

Mishra, U., & Riley, W. J. (2014). Active-layer thickness across Alaska: Comparing observation-based estimates with CMIP5 Earth system model predictions. *Soil Science Society of America Journal*, *78*(3), 894–902. https://doi.org/10.2136/sssaj2013.11.0484

Mishra, U., Torn, M. S., & Fingerman, K. (2013). Miscanthus biomass productivity within US croplands and its potential impact on soil organic carbon. *GCB Bioenergy*, *5*(4), 391–399. https://doi.org/10.1111/j.1757-1707.2012.01201.x

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil*, *5*(1), 79–89. https://doi.org/10.5194/soil-5-79-2019

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

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., & Smith, P. (2016). Climate-smart soils. *Nature*, *532*(7597), 49–57.

R Core Team. (2012). *R: A language and environment for statistical computing. 2012*. R Foundation for Statistical Computing.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., … Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Sanderman, J., Hengl, T., & Fiske, G. J. (2017). Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences*, *114*(36), 9575–9580. https://doi.org/10.1073/pnas.1706103114

Schmidt, M. W., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., & Manning, D. A. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, *478*(7367), 49–56.

Schneider, T. (2020). Earth System Modeling 2.0: Toward data-informed climate models with quantified uncertainties. *Bulletin of the American Physical Society*, *65*, 4092.

Smith, J. O., Smith, P., Wattenbach, M., Zaehle, S., Hiederer, R., Jones, R. J., Montanarella, L., Rounsevell, M. D., Reginster, I., & Ewert, F. (2005). Projected changes in mineral soil carbon of European croplands and grasslands, 1990–2080. *Global Change Biology*, *11*(12), 2141–2152. https://doi.org/10.1111/j.1365-2486.2005.001075.x

Sulman, B. N., Moore, J. A. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., Sridhar, B., Hartman, M. D., Wang, G., Wieder, W. R., Bradford, M. A., Luo, Y., Mayes, M. A., Morrison, E., Riley, W. J., Salazar, A., Schimel, J. P., Tang, J., & Classen, A. T. (2018). Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry*, *141*(2), 109–123. https://doi.org/10.1007/s10533-018-0509-z

Sundquist, E. T., Ackerman, K. V., Bliss, N. B., Kellndorfer, J. M., Reeves, M. C., & Rollins, M. G. (2009). Rapid assessment of US forest and soil organic carbon storage and forest biomass carbon sequestration capacity. *US Geological Survey Open-File Report*, *1283*, 15.

Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., & Luo, Y. (2020). Deep learning optimizes data-driven representation of soil organic carbon in earth system model over the conterminous United States. *Frontiers in Big Data*, *3*. https://doi.org/10.3389/fdata.2020.00017

Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., Wu, T., Zhang, Q., & Allison, S. D. (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

Todd-Brown, K., Randerson, J., Post, W., Hoffman, F., Tarnocai, C., Schuur, E., & Allison, S. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, *10*(3), 1717–1736. https://doi.org/10.5194/bg-10-1717-2013

Trumbore, S. E. (1997). Potential responses of soil organic carbon to global environmental change. *Proceedings of the National Academy of Sciences*, *94*(16), 8284–8291. https://doi.org/10.1073/pnas.94.16.8284

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., & Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1), 5–31. https://doi.org/10.1007/s10584-011-0148-z

Varney, R. M., Chadburn, S. E., Friedlingstein, P., Burke, E. J., Koven, C. D., Hugelius, G., & Cox, P. M. (2020). A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. *Nature Communications*, *11*(1), 1–8. https://doi.org/10.1038/s41467-020-19208-8

Wang, B., Gray, J. M., Waters, C. M., Anwar, M. R., Orgill, S. E., Cowie, A. L., & Li Liu, D. (2021). Modelling and mapping soil organic carbon stocks under future climate change in south-eastern Australia. *Geoderma*, *405*, 115442.

Wang, T., Yang, D., Yang, Y., Piao, S., Li, X., Cheng, G., & Fu, B. (2020). Permafrost thawing puts the frozen carbon at risk over the Tibetan Plateau. *Science Advances*, *6*(19), eaaz3513. https://doi.org/10.1126/sciadv.aaz3513

West, L. T., Wills, S., & Loecke, T. D. (2013). *Rapid assessment of US Soil Carbon (RaCA) for climate change and conservation planning: Summary of soil carbon stocks for the conterminous United States*. USDA Technical Report.

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

Wiesmeier, M., Barthold, F., Blank, B., & Kögel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, *340*(1), 7–24. https://doi.org/10.1007/s11104-010-0425-z

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., & Kögel-Knabner, I. (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

Wills, S., Loecke, T., Sequeira, C., Teachman, G., Grunwald, S., & West, L. T. (2014). Overview of the US rapid carbon assessment project: sampling design, initial summary and uncertainty estimates. In *Soil carbon* (pp. 95–104). Springer.

Wu, Z., Dijkstra, P., Koch, G. W., Peñuelas, J., & Hungate, B. A. (2011). Responses of terrestrial ecosystems to temperature and precipitation change: A meta-analysis of experimental manipulation. *Global Change Biology*, *17*(2), 927–942. https://doi.org/10.1111/j.1365-2486.2010.02302.x

Yan, C., Yuan, Z., Shi, X., Lock, T. R., & Kallenbach, R. L. (2020). A global synthesis reveals more response sensitivity of soil carbon flux than pool to warming. *Journal of Soils and Sediments*, *20*(3), 1208–1221. https://doi.org/10.1007/s11368-019-02513-1

Zimov, S. A., Schuur, E. A., & Chapin, F. S. III (2006). Permafrost and the global carbon budget. *Science (Washington)*, *312*(5780), 1612–1613.

BIOSKETCH

Sagar Gautam is a postdoctoral researcher at Sandia National Laboratory, Livermore, California, USA. His major research interest include model based environmental impact assessment of soil, air and water resources. He focuses on integrating big data, climate model outputs and different genera of models for environmental impact assessment.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Gautam, S., Mishra U., Scown C. D., Wills S. A., Adhikari K., & Drewniak B. A. (2022). Continental United States may lose 1.8 petagrams of soil organic carbon under climate change by 2100. *Global Ecology and Biogeography*, *00*, 1–14. https://doi.org/10.1111/geb.13489