Improving soil carbon estimates by linking conceptual pools against measurable carbon fractions in the DAYCENT Model Version 4.5

Shree R.S. Dangal[1,*], Christopher Schwalm[1], Michel A. Cavigelli[2], Hero T. Gollany[3], Virginia L. Jin[4] & Jonathan Sanderman[1]

[1]Woodwell Climate Research Center, 149 Woods Hole Road, Falmouth, MA 02540, USA

[2]US Department of Agriculture - Agricultural Research Service, Sustainable Agricultural Systems Laboratory, Beltsville Agricultural Research Center, Beltsville, MD 20705, USA

[3]US Department of Agriculture - Agriculture Research Service, Columbia Plateau Conservation Research Center, Pendleton, OR 97810, USA

[4]US Department of Agriculture - Agricultural Research Service, Agroecosystem Management Research Unit, University of Nebraska-Lincoln, NE 68583, USA

Correspondence to: Shree R.S. Dangal (shree.dangal@unl.edu)
[*]*Current Address*: School of Natural Resources, University of Nebraska-Lincoln, NE 68583

Key points:

1. The modified model overestimated measured SOC values at long term research sites but better approximated derived SOC values from other data products when calibrated to carbon (C) fraction compared to the default model.
2. Model modifications led to larger absolute and relative losses of SOC compared to the default model during 1895-2005.
3. Under the RCP8.5 scenario, projected SOC losses with the modified model were 33% and 29% larger for croplands and grasslands, respectively, compared to the default model.

Abstract

Terrestrial soil organic carbon (SOC) dynamics play an important but uncertain role in the global carbon (C) cycle. Current modeling efforts to quantify SOC dynamics in response to global environmental changes do not accurately represent the size, distribution and flux of C from the soil. Here, we modified the Daily Century (DAYCENT) biogeochemical model by parameterizing conceptual SOC pools with C fraction data, followed by historical and future simulations of SOC dynamics. Results showed that simulations using modified DAYCENT (DC_{mod}) led to better initialization of SOC stocks and distribution compared to default DAYCENT (DC_{def}) at long-term research sites. Regional simulation using DC_{mod} demonstrated higher SOC stocks for both croplands (34.86 vs 26.17 MgC ha^{-1}) and grasslands (54.05 vs 40.82 MgC ha^{-1}) compared to DC_{def} for the contemporary period (2001-2005 average), which better matched observationally constrained data-driven maps of current SOC distributions. Projection of SOC dynamics to land cover change (IPCC AR4 A2 scenario) under IPCC AR5 RCP8.5 climate scenario showed absolute SOC loss of 8.44 and 10.43 MgC ha^{-1} for grasslands and croplands, respectively, using DC_{mod} whereas, SOC losses were 6.55 and 7.85 MgC ha^{-1} for grasslands and croplands, respectively, using DC_{def}. The projected SOC loss using DC_{mod} was 33% and 29% higher for croplands and grasslands compared to DC_{def}. Our modeling study demonstrates that initializing SOC pools with C fraction data led to more accurate representation of SOC stocks and individual carbon pool, resulting in larger absolute and relative SOC losses due to agricultural intensification in the warming climate.

1. Introduction

Soil is the largest terrestrial reservoir of organic carbon (C), storing about 1500 Pg C in the top 100 cm (Batjes, 2016; Nachtergaele et al., 2012). Any small changes in the magnitude, distribution and forms of terrestrial soil organic carbon (SOC) may lead to large release of C to the atmosphere (Sulman et al., 2018), with significant impact on food security and the global climate system (Lal, 2004). Given that changes in SOC represent one of the largest uncertainties in the global C budget (Ciais et al., 2014), accurate quantification of the distribution and forms of SOC can help to constrain the global C budget and provide key insights on the underlying processes related to SOC protection and cycling (Stockmann et al., 2013).

Changes in SOC stocks at any given time depend on the balance between organic matter inputs via plant production, additions of manure and compost, and outputs via decomposition, erosion and hydrologic leaching of various C compounds (Davidson and Janssens, 2006; Jobbágy and Jackson, 2000). Although higher organic matter inputs to the soil generally correlate with high SOC (Sanderman et al., 2017a), the biological stability of SOC is ultimately determined by the interactions among the soil physicochemical environment (soil moisture, temperature, pH and aeration), soil mineralogy, and the accessibility of the organic matter to microbes and enzymes (Schmidt et al., 2011). Current understanding of the SOC dynamics indicates that the soil physicochemical environment plays an important role in determining the C efflux from soil and that the efflux rates are modified by substrate availability and the affinities of enzymes for the substrates (Six et al., 2002). However, the extent to which different physicochemical characteristics of soil control the stabilization and cycling of SOC is still debated (Carvalhais et al., 2014; Doetterl et al., 2015; Rasmussen et al., 2018). Additionally, the complex molecular structure of C substrates and their sensitivity to climatic and environmental constraints add further complexity in understanding SOC dynamics at different spatial and temporal scales (Davidson and Janssens, 2006).

Previous studies have shown that the factors affecting the stabilization/destabilization of SOC are numerous and that the changes in SOC over space and time are the result of complex interactions among climatic, biotic and edaphic factors (Rasmussen et al., 2018; Stockmann et al., 2013; Torn et al., 1997; Wiesmeier et al., 2019). For example, Carvalhais et al. (2014) have shown that climate, particularly temperature, strongly controls SOC turnover. Doetterl et al. (2015) found that geochemical characteristics such as base saturation, soil texture, silica content and pH also play a dominant role by altering the adsorption and aggregation of SOC. In addition, other studies indicate that soil nitrogen (N) availability affects SOC change due to constraints on microbial activity and plant productivity (Grandy et al., 2008; Janssens et al., 2010; Sinsabaugh et al., 2005). These findings have led to the view that the accumulation and decomposition of organic matter in soil is ultimately determined by the interactions among climate, vegetation type, topography and lithology.

Biogeochemical models commonly rely on capturing SOC heterogeneity associated with the complex interactions among climatic, biotic and edaphic factors by defining a number of distinct SOC pools with different potential turnover rates (Tian et al., 2015; Todd-Brown et al., 2014). The potential turnover rates of distinct soil pools are modified by climatic factors such as soil moisture and temperature, soil chemical factors such as pH and oxygen availability and the mechanism that facilitates C protection via organo-mineral interactions and aggregation, often loosely represented by clay content (Trumbore, 1997). Each of these pools is conceptual in nature, implying that the turnover times of these pools cannot be determined by chemical and physical fractionation (Paul et al., 2001). As a result, there is increasing need and effort to link the conceptual pools with some measurable data to determine the turnover rates of SOC pools in the biogeochemical models.

In current biogeochemical models, there is a general agreement that the soil organic matter (SOM) contains at least three C pools: an active pool dominated by root exudates and the rapidly decomposable components of fresh plant litter, with mean residence time (MRT) ranging from days to years (Hsieh, 1993); a slow pool dominated by decomposed organic material, often of microbial origin, with MRT ranging from years to centuries (Torn et al., 2013); and a passive pool dominated by stabilized organic matter with MRT of several hundred to thousands of years (Czimczik and Masiello, 2007). Changes in the size and relative abundance of these pools are strongly influenced by climate, soil type and land use (Sanderman et al., 2021). Therefore, accounting for accurate distribution of SOC into different pools is paramount to quantify the current SOC stocks and examine the vulnerability of SOC to future environmental changes.

Relating these conceptual pools with SOC partitioned into laboratory defined fractions, such as particulate-, mineral associated- and pyrogenic-forms of C (POC, MOAC and PyC, respectively), can help to constrain the turnover rate of different pools in biogeochemical models. For example, Skjemstad et al. (2004) related POC, MOAC and PyC approximated using a combination of physical size fractionation and solid-state ^{13}C-NMR spectroscopy with resistant plant material (RPM), humic (HUM) and inert organic material (IOM) pools in the Rothamsted carbon (RothC) model to predict changes in SOC in response to changes in soil type, climate and management. However, RothC does not explicitly simulate plant growth and plant response to dynamic changes in climate and other environmental factors (Zimmermann et al., 2007). In addition, the plant material is loosely partitioned into decomposable and resistant forms with large uncertainties in their respective sizes (Cagnarini et al., 2019). Unlike RothC, ecosystem models such as

Century, DeNitrification-DeComposition (DNDC) and Agricultural Production Systems sIMulator (APSIM) integrate the effects of climate, land use change and land management practices by simulating plant physiology and soil biogeochemistry, and explicitly consider the effects of climate, land use and land management on three conceptual soil C pools with different turnover rates (Hartman et al., 2011; Ogle et al., 2010).

In this study, we modified, calibrated and evaluated the version 4.5 of the Daily Century model (hereafter, DAYCENT) to improve the representation of SOC dynamics by linking conceptual pools of active, slow and passive SOC against estimates of the measurable POC, MOAC and PyC fractions, respectively. We then simulated the response of SOC to climate and land use change during the historical and future period using the default (hereafter, DC_{def}) and modified (hereafter, DC_{mod}) DAYCENT model in the US Great Plains ecoregion. The objectives of this study were to 1) modify the DC_{def} model to link active, slow and passive pools of organic C to soil C fractions; 2) calibrate and evaluate DC_{mod} performance by comparing the distribution of C in active, slow and passive pools against C fractions predicted at seven long-term research sites; 3) evaluate the differences between the DC_{mod} and DC_{def} in simulating contemporary SOC stocks and their distribution by comparing against other existing data products in the US Great Plains region; and 4) project the SOC change in response to climate and land cover change through 2100. We hypothesize that (i) calibrating the conceptual pools to C fraction data in the DAYCENT model leads to more accurate initialization of equilibrium pool structure (Skjemstad et al., 2004), thereby allowing a better comparison of measured and simulated SOC in response to climate, land use and management (Basso et al., 2011); (ii) conversion of native vegetation to any agricultural use significantly alters the distribution of SOC among the various soil pools (Guo and Gifford, 2002), but the rate and extent of SOC change depend on the intensity of agricultural use (Lal, 2018; Page et al., 2014), with larger losses from models that allocate more C to active and slow pools; and (iii) land use under a warming climate would result in larger absolute and relative losses of SOC from the model that derive more SOC from the active pool due to rapid decomposition of fresh organic matter induced by warming (Crowther et al., 2016).

2. Materials and methods

2.1 The DAYCENT Model

The DAYCENT Version 4.5 is a daily time step version of the Century biogeochemical model that simulates the dynamics of C and N of both managed and natural ecosystems (Del Grosso et al., 2002; Parton et al., 1998). The exchange of C and N among the atmosphere, vegetation and soil is a function of climate, land use, land management and other environmental factors. The vegetation pool simulates potential plant growth at a weekly time step limited by water, light and nutrients. The DAYCENT model consists of multiple pools of SOM and simulates turnover as a function of the amount and quality of residue returned to the soil, the size of different soil pools and a series of environmental limitations. The type and timing of management events including tillage, fertilization, irrigation, harvest and grazing activities can affect plant production and SOM retention.

The DAYCENT model was originally developed from the monthly CENTURY model version 4.0. The CENTURY 4.0 is a general FORTRAN model of the plant-soil ecosystem that simulates carbon and nutrient dynamics of different types of terrestrial ecosystems (grasslands, forest, crops and savannas). CENTURY 4.0 primarily focused on simulation of soil organic matter dynamics of agro-ecosystems (Metherell et al., 1994). Earlier development of the CENTURY focused on simulation of soil organic matter dynamics of grasslands, forest and savanna ecosystems (Parton et al., 1988; Sanford Jr et al., 1991).

The first DAYCENT model was developed in FORTRAN 77 and C from CENTURY 4.0 to simulate the exchanges of C, water, nutrients, and gases (CO_2, CH_4, N_2O, NOx, N_2) among the atmosphere, soil and plants at a daily time step (Del Grosso et al., 2001; Kelly et al., 2000; Parton et al., 1988). The submodels used in DAYCENT are described in detail by Del Grosso et al. (2001), which includes submodels for plant productivity, soil organic matter decomposition, soil water and temperature dynamics, and trace gas fluxes. Other model developments while transitioning from CENTURY 4.0 to DAYCENT included dynamic carbon allocation and changes in growing degree days routine that triggers the start and end of growing season based on phenology (soil surface temperature, air temperature, and thermal units).

The first formal version DAYCENT 4.5 (Hartman et al., 2011) was developed from Del Grosso et al. (2002), with a focus on simulation of trace gas fluxes for major crop types in the US Great Plains region. Hartman et al. (2011) focused on calibrating and validating crop yield and trace gas fluxes for all the major crop types in 21 representative counties in the US Great Plains region.

The SOM sub-model consists of active, slow and passive pools with different turnover times. The active pool has a short (1-5 yr) turnover time and consists of live microbes and microbial products. The slow pool has an intermediate turn over time (20-50 yr) and contains physically protected organic matter and stabilized microbial products. The passive pool has a long turnover time (400-2000 yr) with physically and chemically stabilized SOC. In DAYCENT, the turnover of the active, slow and passive pools are simulated as a function of potential decomposition rates of respective pools modified by soil temperature, moisture, clay content, pH and cultivation effects. Changes in SOC are simulated for the top 20 cm of the soil.

In this study, we modified the DAYCENT and developed a methodology to calibrate the size of the conceptual soil pools by comparing it with carbon fraction data at long term research sites.

First, we developed measurable carbon fraction data using a combination of diffuse reflectance spectroscopy and a machine learning model (section 2.2). Second, we modified the DAYCENT model to link conceptual active, slow, and passive pools with the carbon fraction data (section 2.3 & 2.4). Third, we parameterized the DAYCENT by tuning the potential decomposition rates (*k*) such that the size of the active, slow and passive soil pools match with the POC, MAOC and PyC, respectively at the long-term research sites (section 2.5). Fourth, we calibrated both the default and modified DAYCENT using input data developed in section 2.3 against observed total SOC at the long-term research sites (section 2.6), followed by model validation (section 2.7) and historical and future simulations (section 2.8).

2.2 Development of carbon fraction datasets to match with soil carbon pools

To link the SOC pools in DAYCENT with measurable C fractions, we used seven long-term research sites located in the United States (Cavigelli et al., 2008; Gollany, 2016; Ingram et al., 2008; Liebig et al., 2010; Schmer et al., 2014; Sindelar et al., 2015; Syswerda et al., 2011), which span a range of climatic, land use and land management gradients (Table 1). Six of seven research sites are part of Long-Term Agroecosystem Research (LTAR) network focused on sustainable intensification of agricultural production. The remaining site is part of Columbia Plateau Conservation Research Center (CPCRC) Long-Term Experiment (LTE). At each site, we predicted the POC, MAOC and PyC fractions using a diffuse reflectance mid-infrared (MIR) spectroscopy-based model as detailed in Sanderman et al. (2021). The predictive models for the C fractions were developed from a database of fully fractionated soil samples using a combination of physical size separation and solid-state ^{13}C NMR spectroscopy (Baldock et al., 2013b) of Australian (Baldock et al., 2013a) and US origin (Sanderman et al., 2021). All samples for model development were scanned using a Thermo Nicolet 6700 FTIR spectrometer with Pike AutoDiff reflectance accessory located at the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia. The soil samples from all the long-term research sites were scanned using a Bruker Vertex 70 FTIR equipped with a Pike AutoDiff reflectance accessory located at Woodwell Climate Research Center in the United States. For all samples, spectra were acquired on dried and finely milled soil samples. Since the SOC fraction model and the soil samples were scanned using different instruments, we developed a calibration transfer routine to account for the differences in spectral responses between the CSIRO (primary) and Woodwell (secondary) instruments by scanning a common set of 285 soil samples. The calibration transfer routine was developed using piecewise direct standardization (PDS) as described in Dangal & Sanderman (2020).

Table 1. General attributes of the LTAR, LTER and CPCRC-LTE sites used for DAYCENT parameterization and calibration

Site Name	Sampling Location	Lon	Lat	T_{avg} (ºC)	Annual Precip. (mm)	Elev (m)	Land use	Data Avail.	Reference
Lower Chesa. Bay	Beltsville, MD	-76.9	39.1	12.8	1110	41	CS	1996-2016	Cavigelli et al. 2008
CPCRC-NTLTE	Pendleton, OR	-118.4	45.4	10.6	437	456	WW-FA	2005-2014	Gollany 2016
Cent. Plains Exp. Ran.	Cheyenne, WY	-104.9	41.2	8.6	425	1930	C3-C4 Gra.	2004-2013	Ingram et al. 2008
Northern Plains	Mandan, ND	-100.9	46.8	4	416	593	C3-C4 Gra.	1959-2014	Liebig et al 2010
Platte/High Plains Aq.	Lincoln, NE	-96.5	40.9	11	728	369	CC,CS	1998-2011	Sindelar et al 2015
Platte/High Plains Aq.	Mead, NE	-96.0	41.0	9.8	740	349	CC	2001-2015	Schmer et al. 2014
Kellogg Bio. Station	H. Corners, MI	-85.4	42.4	9.7	920	288	CSW-Gra.	1989-2017	Syswerda et al. 2011ǂ

CS: Corn-Soya; WW: Winter Wheat; FA: Fallow; CC: Continuous Corn, SC: Soya-Corn, CSW: Corn-Soya-Wheat, Gra.: Grass
ǂH. Corners, MI is a LTER & LTAR site; CPCRC-NTLTE: Columbia Plateau Conservation Research Center No-Till Long-Term Experiment.

For estimating C fractions of the prediction set (i.e., soil spectra of seven long-term research sites), we used a local memory based learning (MBL) approach that fits a unique target function corresponding to each sample in the prediction set (Dangal et al., 2019; Ramirez-Lopez et al., 2013). The MBL selects spectrally similar neighbors for each sample in the prediction sets to build a unique SOC fraction model for each target sample. The spectrally similar neighbors were optimized by developing a soil C fraction model using a range of spectrally similar neighbors and selecting the neighbors that produce the minimum root mean square error based on local cross validation. Before developing the soil C fraction model, the spectra of both the calibration and prediction sets were baseline transformed. Following baseline transformation, spectral outliers were detected using F-ratios (Hicks et al., 2015). The F-ratio estimates the probability distribution function of the spectra and picks samples that fall outside the calibration space as outliers (Dangal et al., 2019). Observation data used for building the soil C fraction model were square root transformed before model development and later back-transformed when estimating the goodness-of-fit. The performance of predictive models is shown in Table S1.

The predicted soil C fractions for the seven long-term research sites were then converted into C fraction stocks using the relationship between C fraction (%), bulk density (BD; g/cm^3) and the depth (cm) of soil samples. Since the BD data were not available for all long-term research sites for different crop rotation and grazing intensities, we predicted BD using methods similar to those described above. The only difference was that the samples used to develop the BD model were based on a much larger database of soil spectra scanned at the Kellogg Soil Survey Laboratory (KSSL) in Lincoln, USA (Dangal et al., 2019). Before predicting BD, the calibration transfer, as documented in Dangal & Sanderman (2020), between the KSSL and Woodwell soil spectra were developed and the local modeling approach (i.e., MBL) was used to make final prediction for samples with missing laboratory BD. Calibration transfer between the spectrometers at the Woodwell (secondary instrument) and KSSL (primary instrument) laboratory was necessary to improve prediction of BD (R^2 = 0.46-0.64 and RMSE = 0.26-0.50) (Dangal and Sanderman, 2020). One of the technical challenges associated with the comparison of simulated pool sizes against diffuse reflectance spectroscopy-based predictions of POC, MOAC and PyC at long-term research sites was the absence of laboratory data on C fractions to validate the MIR based predictions. To address this shortcoming, we first compared the sum of the MIR based predictions of POC, MOAC and PyC against observation of total SOC available at these sites (Figure S1). When comparing the total SOC against MIR based predictions, we did not limit the comparison to 20 cm, but allowed it across the full soil depth profile based on the availability of SOC data at the seven long-term research sites. Additionally, the laboratory data used for model comparison were available at multiple depths of up to 60 cm often without a direct measurement for the 0-20 cm depth necessitating an approximation of the 0-20 cm stock. For example, when soils were collected from 0-15 and 15-30 cm, we estimated the 20 cm SOC stock by adding 1/3 of the 15-30 cm SOC stock to the entire 0-15 cm SOC stock.

2.3 Input datasets for driving the DAYCENT model

The US Great Plains region was delineated using the Level I ecoregions map (Omernik and Griffith, 2014) available through the Environmental Protection Agency (https://www.epa.gov/eco-research/ecoregions-north-america). The datasets for driving the DAYCENT were divided into two parts: 1) dynamic datasets that include time series of daily climate (precipitation, maximum and minimum temperature), annual land cover land use change (LCLUC) and land management practices (irrigation, fertilization and cropping system, tillage intensity) and 2) static datasets that include information on soil properties (soil texture, pH and bulk density) (Sanderman et al., 2021), and topography maps (Jarvis et al., 2008). For the historical period (1895-2005), we used a combination of VEMAP and PRISM (1895-1979) and Daymet (1980-2005) (Daly and Bryant, 2013; Kittel et al., 2004; Thornton et al., 2012). The VEMAP datasets are available at a daily time step and a coarser spatial resolution (0.5º x 0.5º), while the PRISM datasets are available at a monthly time step and a finer spatial resolution (10 km × 10 km). We interpolated the PRISM data at a daily time step by using the daily trend from the VEMAP datasets such that the monthly precipitation totals and monthly average temperature matches the monthly climate from the PRISM data. For the future (2006-2100), we used the Intergovernmental Panel on Climate Change (IPCC) 5th assessment report (AR5) RCP4.5 and RCP8.5 climate scenarios available at a spatial resolution of 1/16º x 1/16º.

Table 2. Default and modified decomposition (k) parameters used in the DAYCENT to simulate the size of different carbon pools

Pools	Default	Modified *k* (yr^{-1})				
	k (yr^{-1})	grid search	N	Optimized	Absolute	Relative (%)
Active	7.30	(3,12)	301	3.50	-3.80	-52
Slow	0.20	(0.10,0.30)	201	0.14	-0.06	-30
Passive	0.0045	(0.001,0.0085)	351	0.0075	0.003	+67

For annual LCLUC, we used spatially explicit datasets available at a resolution of 250m × 250m for the historical (1938-2005) and future (2006-2100) periods under the IPCC 4th assessment report (AR4) A2 scenario (Sohl et al., 2012). We used only the A2 land cover scenario because there was not much difference in the trajectories of land cover change through 2100. For the period 1895-1937, we backcasted the proportional distribution of croplands and grasslands by integrating the Sohl et al. (2012) data with HYDE v3.2 data (Klein Goldewijk et al., 2017). We estimated the fractional distribution of croplands and grasslands by calculating the total number of pixels dominated by each land cover type at 250m resolution within each 1/16 ^{o} grid cell (Figure S2a). Irrigation and fertilization data are based on census of agriculture statistics (Falcone and LaMotte, 2016). All datasets were interpolated/aggregated to a common resolution of 1/16^{o} x 1/16^{o} (approximately 7km x 7km at the equator).

Cropping systems and crop rotation are based on county level data for the US Great Plains region available through Hartman et al. (2011), which were merged with tillage type and intensity data (Baker, 2011) to write 24 unique schedule files that describe grid-specific cropping system and crop management practices. The 24 unique schedule files include sequences of time blocks, with each block describing a unique set of crop types, crop rotation, tillage type, tillage intensity, fertilization, irrigation and residue removal (Hartman et al., 2011). Using these schedule files, we developed an unsupervised classification algorithm (K-means) to create 24 unique clusters as a function of long-term average climate (precipitation, minimum- and maximum-temperatures), land forms, land cover type and elevation. We then assigned all the grid cells to one of the 24 unique clusters to create a spatially explicit dataset on cropping system and crop rotation. While developing the unsupervised classification algorithm, the eastern part of the US Great Plains region dominated by corn (*Zea mays* L.) - soybean (*Glycine max* (L.) Merr.) rotation was underrepresented. To address this shortcoming, we used randomly selected grid points from the CropScape data (https://nassgeodata.gmu.edu/CropScape/) available through the USDA National Agricultural Statistics Service in the unsupervised classification algorithm. Additionally, cropping systems classified using the unsupervised algorithm was verified against current CropScape data allowing for realistic representation of cropping systems. The distribution of schedule files representing different crop rotation and crop types used to build the unsupervised classification is shown in Figure S2b and the spatial distribution of crop rotations based on the unsupervised classification is shown in Figure S3.

2.4 Linking DAYCENT conceptual pools with C fractions

The SOC dynamics in the DAYCENT consists of the first-order kinetic exchanges among conceptual pools (active, slow, and passive) defined by empirical turnover rates (Parton et al., 1987). However, a major impetus for quantifying these pools comes from the fact that the size and distribution of SOC in the different pools cannot be directly linked with experimental data. Here, we developed a methodology to link the conceptual active, slow and passive pools to spectroscopy-based estimates of POC, MAOC and PyC fractions. The rate of decomposition across POC, MAOC and PyC are consistent with the potential turnover rates assigned to the active, slow, and passive pools in soil C models (Baldock et al., 2013b). As a result, we modified the potential turnover rates in the DAYCENT model such that the absolute difference between the simulated SOC and predicted C fractions was minimized (see section 2.5 below). When matching the soil pools with C fraction data, we compared the sum of belowground structural, metabolic and active pool SOC to POC, slow pool SOC to MAOC, and passive pool SOC to PyC. Details on matching the conceptual pools with C fraction data are provided in Figure S4.

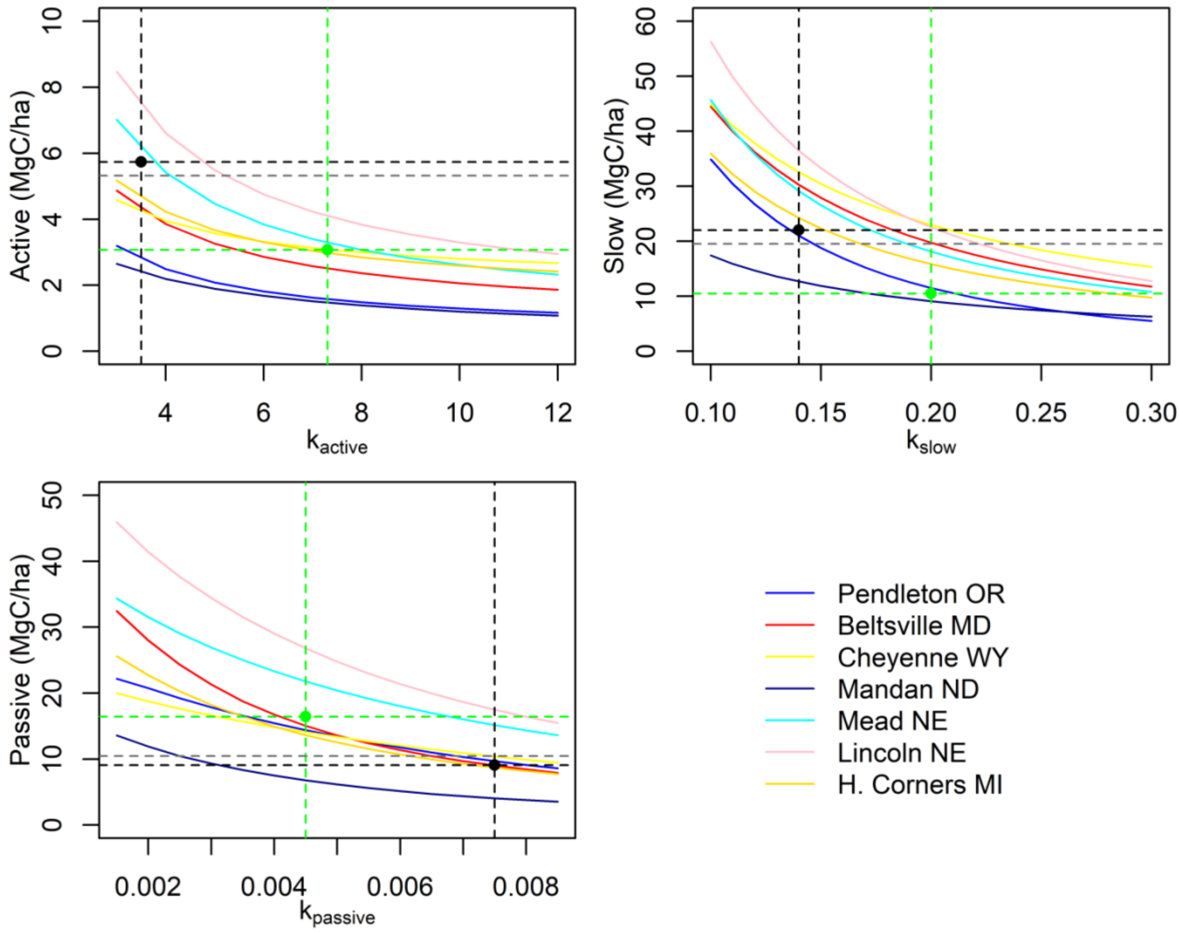

Figure 1. Parameterization of k_{active}, k_{slow} and $k_{passive}$ using carbon fractions predicted across long term research sites. The dashed black line represents the potential decomposition rates (*k*) that is optimized when the absolute difference between the DC_{mod} simulated SOC in different pools and the predicted C fractions is minimum. The dashed green line represents the size of different soil SOC pools using the default *k* value based on DC_{def} model. The dashed grey line is the average POC (i.e. active), MAOC (i.e. slow) and PyC (i.e. passive) predicted using the combination of diffuse reflectance spectroscopy and machine learning at seven long term research sites {Citation}.

2.5 Model parameterization

In this study, we performed a grid search to parameterize the potential decomposition rates for respective soil pools by running the DAYCENT at seven long-term research sites (Figure 1; Table

2), and compare the simulated SOC in active, slow, and passive pools with the POC, MAOC and PyC fractions. In the current DAYCENT model, total SOC is defined as follows:

$$SOC_{total} = SOC_{strc} + SOC_{metab} + SOC_{active} + SOC_{slow} + SOC_{passive} \quad (1)$$

Where,

SOC_{strc} = structural SOC pool

SOC_{metab} = metabolic SOC pool

SOC_{active} = active SOC pool

SOC_{slow} = slow SOC pool

$SOC_{passive}$ = passive SOC pool

Each of the above SOC pool has a specific potential decomposition rates that determines the time (ranging from years to centuries) until decomposition. Plant material is transferred to the active, slow and passive pools from aboveground and belowground litter pools and three dead pools. Total C flow (CF_{act}) out of the active pool is a function of potential decomposition rates modified by the effect of moisture, temperature, pH, and soil texture.

$$CF_{act} = k_{act} \times SOC_{act} \times bg_{dec} \times clt_{act} \times text_{ef} \times anerb_{dec} \times pH_{eff} \times dtm \quad (2)$$

Where,

CF_{act} = the total amount of C flow out of the active pool (g C m^{-2})

k_{act} = intrinsic decomposition rate of the active pool (yr^{-1})

SOC_{act} = SOC in the active pool (g C m^{-2}).

bg_{dec} = the effect of moisture and temperature on the decomposition rate (0-1)

clt_{act} = the effect of cultivation on the decomposition rate for crops (0-1) for the active pool

$text_{ef}$ = the effect of soil texture on the decomposition rate (0-1)

$anerb_{dec}$ = the effect of anaerobic conditions on the decomposition rate (0-1)

pH_{eff} = the effect of pH on the decomposition rate (0-1)

dtm = the time step (fraction of year)

The respiratory loss when the active pool decomposes is calculated as:

$$CO_{2(act)} = CF_{act} \times p1CO_2 \tag{3}$$

Where,

$CO_{2(act)}$ = respiratory loss from the SOC_{act} pool (g C m^{-2})

$p1CO_2$ = scalar that control respiratory CO_2 loss computed as a function of intercept and slope parameters modified by soil texture

The C flow from active to passive pool is then computed as:

$$CF_{act2pas} = CF_{act} \times fps1s3 \times (1 + animpt \times (1 - anerb)) \tag{4}$$

Where,

$CF_{act2pas}$ = C flow from the active to the passive pool (g C m^{-2})

$fps1s3$ = impact of soil texture on the C flow (0-1)

$animpt$ = the slope term that controls the effect of soil anaerobic condition on C flows from active to passive pool (0-1)

$anerb$ = effect of anaerobic condition on decomposition computed as a function of soil available water and potential evapotranspiration rates

The C flow from active to the slow pool is then computed as the difference between total C flow out of the active pool, respiratory CO2 loss, C flow from active to passive pool and C lost due to leaching. Mathematically,

$$CF_{act2slo} = CF_{act} - CO_{2(act)} - CF_{act2pas} - C_{leach} \tag{5}$$

Where,

C_{leach} = C lost due to leaching calculated as a function of leaching intensity (0-1) and soil texture

Likewise, total C flow (CF_{slo}) out of the slow pool is a function of potential decomposition rates modified by the effect of moisture, temperature, pH, and soil texture.

$$CF_{slo} = k_{slo} \times SOC_{slo} \times bg_{dec} \times clt_{slo} \times anerb_{dec} \times pH_{eff} \times dtm \quad (6)$$

k_{slo} = intrinsic decomposition rate of the slow pool (yr^{-1})

SOC_{slo} = SOC in the slow pool ($g\ C\ m^{-2}$).

clt_{slo} = the effect of cultivation on the decomposition rate for crops (0-1) for the slow pool

The respiratory loss when the slow pool decomposes is calculated as:

$$CO_{2(slo)} = CF_{slo} \times p2CO_2 \quad (7)$$

Where,

$CO_{2(slo)}$ = respiratory loss from the SOC_{slo} pool ($g\ C\ m^{-2}$)

$P2CO_2$ = parameter that controls decomposition rates of the slow pool (0-1)

The C flow from slow to passive pool is then computed as:

$$C_{slo2pas} = CF_{slo} \times fps2s3 \times (1 + animpt \times (1 - anerb)) \quad (8)$$

Where,

$fps2s3$ = impact of soil texture on decomposition (0-1)

The C flow from slow to active pool is then computed as a difference between total C flow out of the slow pool, respiratory CO2 loss and total C flow from slow to passive pool. Mathematically,

$$CF_{slo2act} = CF_{act} - CO_{2(slo)} - CF_{slo2pas} \quad (9)$$

Likewise, total C flow (CF_{pas}) out of the passive pool is a function of potential decomposition rates modified by the effect of moisture, temperature and pH.

$$C_{pas} = k_{pas} \times SOC_{pas} \times bg_{dec} \times clt_{pas} \times pH_{eff} \times dtm \quad (10)$$

Where,

k_{pas} = intrinsic decomposition rate of the passive pool (yr^{-1})

SOC_{pas} = SOC in the slow pool (g C m^{-2}).

clt_{pas} = the effect of cultivation on the decomposition rate for crops (0-1) for the passive pool

The CF_{pas} is either lost through respiratory processes or transferred to the active pool using the following equation:

$$CO_{2(pas)} = CF_{pas} \times p3co2 \tag{11}$$

$$CF_{pas2act} = CF_{pas} \times (1 - p3co2)) \tag{12}$$

Where,

$CO_{2(pas)}$ = respiratory loss from the passive SOC pool (g C m^{-2})

$p3co_2$ = parameter that control decomposition rates of passive pool (0-1)

$CF_{pas2act}$ = C flow from passive to active pool (g C m^{-2})

Since DAYCENT is a donor-controlled model and changes in organic matter are primarily driven by a top down approach, we first parameterize the active soil pool by comparing the simulated SOC in the active pool against POC predicted using diffuse reflectance spectroscopy. During the parameterization process, we varied the potential decomposition rates (k_{active}) by running the model to equilibrium under native vegetation for 2000 years. We then used site history at seven long-term research sites to create schedule files and simulate the effects of historical cropping systems, land use change, land management and grazing practices on the active SOC. The potential decomposition rates for the active soil pool were optimized when the absolute difference between the average of SOC in the active pool and the POC for the top 20 cm across all sites was minimum. We repeated the above process for parameterizing the slow- and passive-carbon pools by comparing it with MOAC and PyC, respectively. Similar to the active pool, we performed a grid search using the existing parameters based on the default model that controls the potential decomposition rates (k_{slow} and $k_{passive}$) of the slow- and passive-pools. We then optimized the parameter by using the potential decomposition rates that provides the minimum difference in the absolute values across all sites.

2.6 Model calibration and simulation procedure

The DAYCENT model has been well calibrated across a range of climatic, environmental, and land use gradients for different crop and grassland types. Details of the calibration procedure can be found in Hartman et al. (2011). Briefly, adjustment of key model parameters that control plant growth and SOM changes were made by changing the schedule files at each point in time. For example, transitioning to higher yielding corn varieties occurred in 1936, while the short and semi-dwarf wheat varieties were introduced in the 1960s. During the calibration process, model parameters that control the maximum photosynthetic rate and grain to stalk ratio were adjusted within realistic limits to account for improvement in crop varieties. Additionally, adjustments in the schedule files were made to account for residue removal in early years, while residues were retained in later years, thereby increasing nutrient input to the soils. These calibration strategies have allowed to better capture crop dynamics in the US Great Plains region (Hartman et al., 2011). Model simulation begins with the equilibrium run starting from year zero to year 1894 by repeating daily climate data from 1895-2005 and native vegetation without disturbance or land use change. Following the equilibrium run, we performed a historical simulation to quantify the effects of land use history, land management practices, and climate change on the evolution of SOC during 1895-2005. Finally, we performed future simulations using two climate scenarios (RCP4.5 and RCP8.5) and A2 LCLUC, with land management practices (i.e. irrigation, fertilization, tillage practices, and crop rotation) held at 2005 levels during 2006-2100.

2.7 Model validation at site and regional scales

The performance of the calibrated model was assessed by comparing simulated SOC in the active, slow, and passive pools against predictions of POC, MAOC and PyC, respectively, at the seven long-term research sites. In the validation procedure, we ran the model at these sites using plant growth and soil parameters determined from model calibration, but with changing climate, environmental, and land use data based on the land use history of the respective sites. For all the sites, we compared the distribution of SOC in different pools and evaluated model performance using linear regression and the goodness-of-fit statistics (bias, R^2, RMSE).

We also compared the distribution of SOC simulated using DAYCENT against the machine learning model-based predictions of POC, MAOC, and PyC for the US Great Plains ecoregion (Sanderman et al., 2021). Additionally, we compared simulated total SOC against two other SOC maps for the contemporary period (Hengl et al., 2017; Ramcharan et al., 2018) .

2.8 Historical and future changes in SOC stocks

To quantify the effect of the new parameterization scheme linking measurable soil C pools with conceptual active, slow, and passive pools from the DAYCENT, we designed two scenarios. In the first scenario, we ran the model using the default (DC_{def}) and the modified (DC_{mod}) model that links conceptual pools with C fraction during the historical period (1895-2005) to quantify the differences in SOC across different pools associated with different parameterization. In the second scenario, we performed future simulations to understand if the different model structures (DC_{def} versus DC_{mod}) result in different effects of climate and LCLUC on SOC stocks. We used the IPCC AR5 RCP8.5 and RCP4.5 climate scenarios and the IPCC AR4 A2 LCLUC scenarios to quantify the effects of future climate and LCLUC change on SOC stocks. The RCP8.5 corresponds to the pathway that tracks current global trajectories of cumulative CO_2 emissions (CO_2 levels reaching

960 ppm by 2100) with the assumption of high population growth and modest rates of technological change and energy intensity improvements (Riahi et al., 2011; Schwalm et al., 2020). The RCP4.5 is a modest emission scenario with CO_2 levels reaching 540 ppm by 2100 under the assumption of shift toward low emission technologies and the deployment of carbon capture and geologic storage technology (Thomson et al., 2011). The A2 land cover scenario emphasizes rapid population growth and economic development, and resembles closely to the RCP8.5 scenario. We used the AR4 for LCLUC because Sohl et al. (2012) data were available at high resolution and allowed for smoother transition between land cover types when moving from historical to future A2 LCLUC scenarios. The purpose of the second scenario is to better understand the response of SOC to future climate and LCLUC and examine the effect of the new model modification on the projected change in total SOC through 2100.

3. Results and Discussion

By quantifying the size and distribution of conceptual SOC pools of ecosystem models using a combination of diffuse reflectance spectroscopy and machine learning, we were able to modify DAYCENT by relating the conceptual active, slow and passive pools with measurable POC, MAOC and PyC fractions (section 3.1). Model modification led to more accurate representation of the magnitude and distribution of SOC (section 3.2) and was necessary to accurately quantify the legacy effect of previous land use under a changing climate and reproduce current SOC stocks compared to the default model (section 3.3). Projection of future SOC change show that the default model underestimates the SOC loss in response to climate and land cover change by 31% and 29% for croplands and grasslands, respectively (section 3.4). Overall, our results demonstrate that relating the pools sizes from the ecosystem model with C fraction data is necessary to better initialize SOC pool and simulate SOC response to climate and land use into the future.

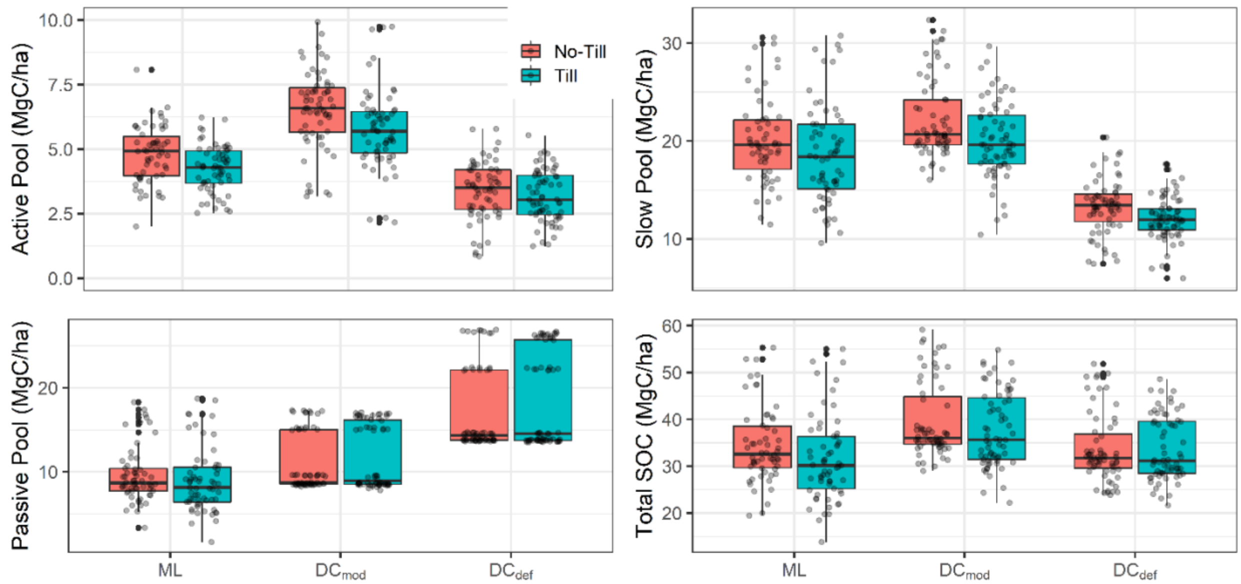

Figure 2. Comparison of the machine learning (ML) and DAYCENT simulated SOC using the modified (DC_{mod}) and default (DC_{def}) models at long-term research sites with a known cropping history. The black dots in the boxplot represent the SOC at the various sites plotted by adding a random value such that they do not overlap with each other.

3.1 Model evaluation of total SOC and the distribution of SOC at long-term research sites

The modified model (DC_{mod}) linking conceptual soil pools to measurable C fractions showed better representation of the distribution of C stocks across different pools compared to the default model (DC_{def}) (Figures 2 & 3). When the mean SOC at these sites were compared to DC_{mod} and DC_{def} simulated SOC, DC_{mod} had better fit (R^2 = 0.52) and lower RMSE (8.49 Mg C ha^{-1}) compared to DC_{def} (R^2 = 0.40; RMSE = 8.93 Mg C ha^{-1}) (Figure S5). The mean SOC based on observation for these sites was 38.96 Mg C ha^{-1}, which is comparable to the sum of predicted C fractions (37.07 Mg C ha^{-1}) and simulated SOC using DC_{mod} (42.30 Mg C ha^{-1}) and DC_{def} (36.60 Mg C ha^{-1}) models. The DC_{mod} simulated SOC was higher than observation and machine learning based SOC by 9 and 12%, respectively, while DC_{def} showed under-predicted SOC by 6% compared to observation. Although DC_{mod} showed a tendency toward over-prediction, assessment of the distribution of SOC demonstrated that DC_{mod} was able to better simulate the distribution of SOC in soil pools compared to DC_{def}. The DC_{mod} simulated the highest proportion of C in the slow (56%) pool followed by the passive (30%) and active (14%) pools, which is comparable to the machine learning model-based estimates of MAOC (57%), PyC (29%) and POC (14%), respectively. Unlike DC_{mod}, DC_{def} model simulated the highest proportion of C in passive (53%), followed by slow (39%) and active (8%) pools (Table S2).

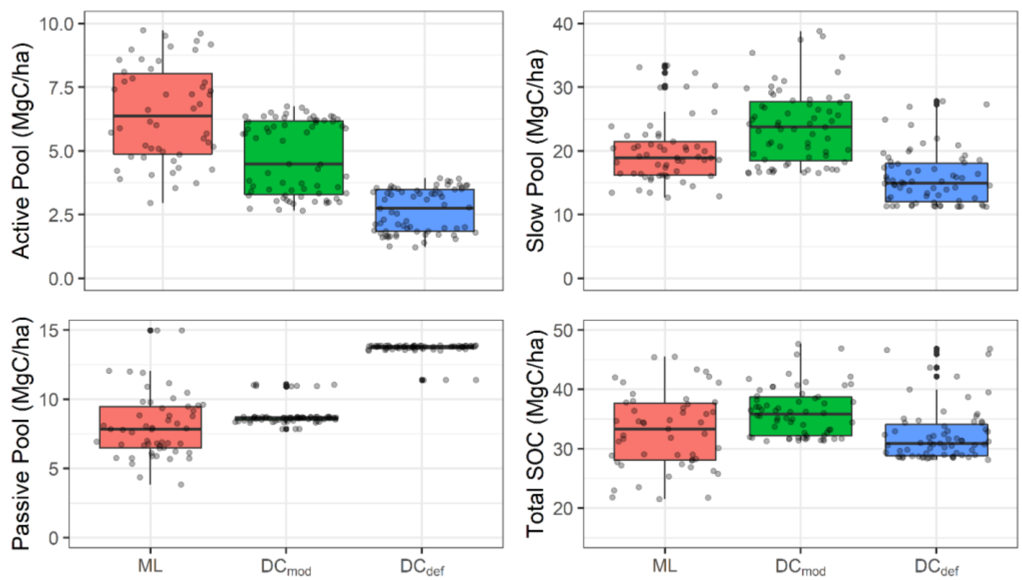

Figure 3. Comparison of the machine learning (ML) and DAYCENT simulated SOC using the modified (DC_{mod}) and default (DC_{def}) models across different pools at two long-term research sites dominated by grasslands with a known grazing history. The black dots in the boxplot represent the SOC across different sites plotted by adding a random value such that they do not overlap with each other.

Evaluation of the model performance (DC_{mod}) for grasslands and croplands showed that the modified model (DC_{mod}) outperformed the default model (DC_{def}) with better model fit ($R^2 = 0.60$), lower bias (-1.94 Mg C ha^{-1}) and lower RMSE (6.7 Mg C ha^{-1}) for grasslands (Figure S6). The DC_{mod} also produced better model fit for croplands (R^2 = 0.48), but higher bias (-5.84 Mg C ha^{-1}) and RMSE (8.86 Mg C ha^{-1}) compared to the default (DC_{def}) model (bias = -0.82 and RMSE = 7.45 Mg C ha^{-1}). The DC_{mod} was able to better represent the distribution of C in the active, slow and passive pools for both grasslands and croplands, while DC_{def} showed large discrepancies when representing the distribution of SOC for croplands (Table S2).

The results of this exercise demonstrate that optimizing the model parameters to initialize the conceptual SOC pools by matching with C fraction data can reproduce the distribution of SOC (Figures 2 & 3), building confidence in the modeling of SOC stocks, and their pool distribution (Lee and Viscarra Rossel, 2020; Luo et al., 2016). A common approach to initializing soil C pools is based on the use of soil C steady-state conditions, which is primarily achieved by running the model over a long period of 100 to 10000 years under native vegetation. However, this approach has shown large uncertainty in the estimation of contemporary SOC partly due to differences in parameter values used to determine the initial SOC stocks, which vary many fold across models (Tian et al., 2015; Todd-Brown et al., 2014). Additionally, the size and distribution of the soil C pools are constrained by model structure and parameter values producing large differences in initial conditions, which ultimately propagates into uncertainties in historical and future projection of SOC change (Ogle et al., 2010; Shi et al., 2018). Relating these conceptual pools to measurable C fractions by optimizing parameters that control decomposition rates can help to constrain initial pool size and reduce uncertainties related to initial SOC stocks across different models (Christensen, 1996; Luo et al., 2016; Zimmermann et al., 2007). Results of this study show that tuning the potential decomposition rates within reasonable range (Figure 1) can effectively capture the distribution of SOC among different pools without significantly altering the magnitude of total SOC (Figures 2 & 3).

While tuning the parameters that control potential decomposition rates, active, and slow pools were adjusted by -3.8 yr^{-1} (-52% compared to default rate) and -0.06 yr^{-1} (-30%) respectively, and passive pool was increased by 0.003 yr^{-1} (67%) to match with C fractions data at the long-term research sites. These modifications were done such that the model was able to simulate total SOC and their distribution under current climatic, and land use conditions while also allowing to capture the legacy effect of previous land use, crop rotation, and tillage practices. It is important to note that other soil C models use C fraction data obtained under land use of varying intensities to run the model to steady state (Zimmermann et al., 2007), although soils under continuous use are in a transient state (Wieder et al., 2018). The rate and direction of SOC change can be modified by environmental factors, previous land use, and current management practices (e.g., intensity, cropping systems and fertilization/irrigation), which ultimately determine a new equilibrium or transient state (Chan et al., 2011; Van Groenigen et al., 2014). Here, we run the model to steady state conditions, and calibrated the SOC stocks to current land use and management practices by matching with C fractions data at all the sites.

3.2 Model evaluation of SOC stocks and their distribution at the regional scale

Evaluation of the model performance at the regional level by comparing model simulations to three data-driven SOC maps showed that the default (DC_{def}) model under-predicts SOC stocks for the contemporary period (2001-2005 average). The modified (DC_{mod}) model was better able to reproduce the spatial pattern as observed in the data driven estimates of SOC (Figure 4). The DC_{mod} simulated contemporary SOC stocks of 34.86 Mg C ha^{-1} were closer to the estimates based on three data-driven models (32.38 – 39.19 Mg C ha^{-1}) (Figure S7). The DC_{def} simulated SOC stocks of 26.17 Mg C ha^{-1}, which is lower than the machine learning based predictions by 19-33%.

Interestingly, both DC_{def} and DC_{mod} were not able to reproduce the high C stocks in the northeastern Great Plains although data driven modeling shows large SOC stocks.

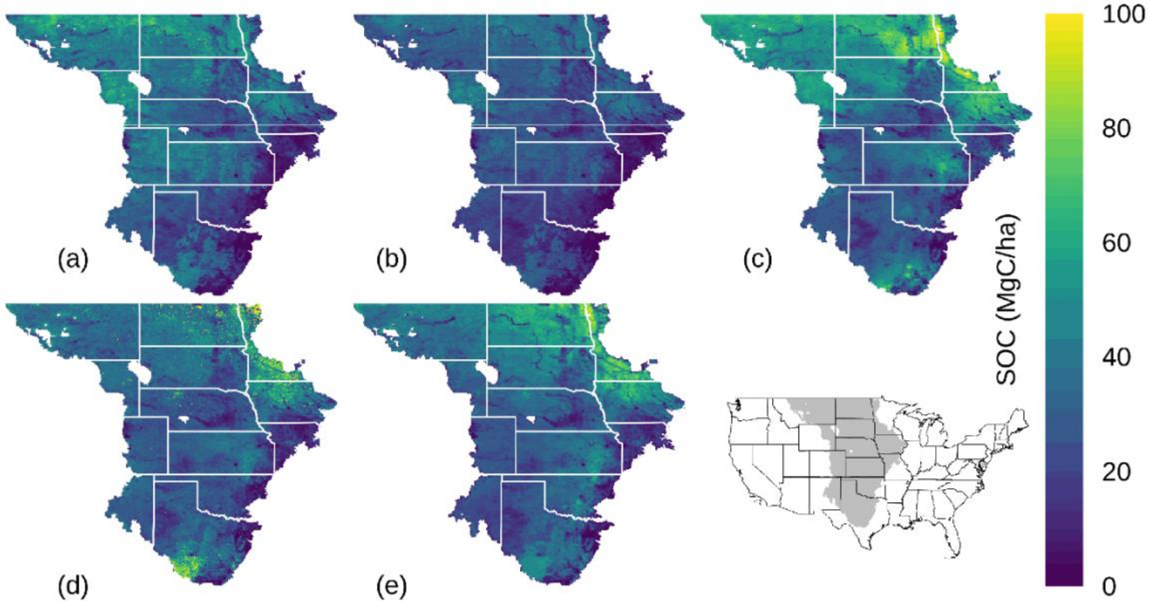

Figure 4. Spatial pattern of SOC change during the contemporary period: modified (DC_{mod}) (a), default (DC_{def}) (b), Sanderman et al. (2021) (c), Ramcharan et al. (2018) (d), and Hengl et al. (2017) (e). Data-driven SOC maps were scaled by cropland and grassland distribution maps before comparing against DAYCENT-simulated SOC.

Evaluation of the model performance using a scatterplot shows that calibration of active, slow, and passive pools was necessary to produce unbiased estimates of SOC despite having slightly higher RMSE values than the default model when compared to the different SOC data sets (Figure 5). Among the three data driven models, Sanderman et al. (2021) also provided prediction of POC, MAOC, and PyC in the US Great Plains region. Comparison of the distribution of SOC across different pools indicate that the DC_{mod} was able to reproduce SOC in the slow/MAOC, and passive/PyC pools but under-predicted the size of the active/POC pool (Figure S8).

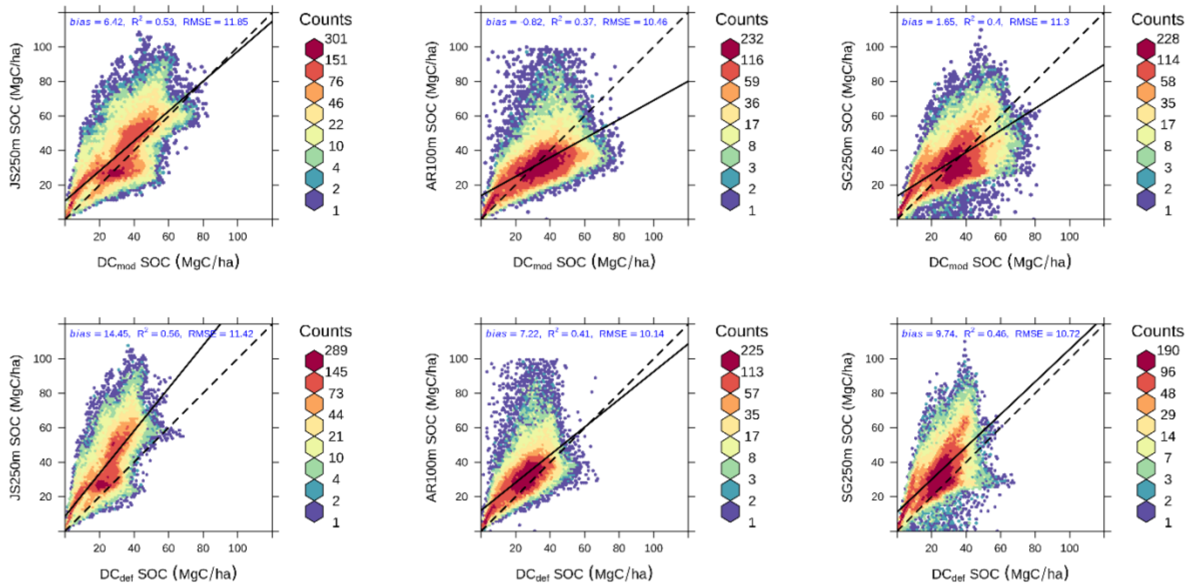

Figure 5. Scatter plots of the comparison of DAYCENT simulated SOC (DC_{mod} & DC_{def}) against Sanderman et al. (2021) – JS250m, Ramcharan et al. (2018) – AR100m, and Hengl et al. (2017) – SG250m.

While the modified (DC_{mod}) model was able to better capture the magnitude and spatial pattern of SOC when compared against data based on machine learning models, the datasets themselves present a few challenges when comparing with the results from this study. First, these datasets were produced using the environmental covariates approach under current climatic and land use conditions, and thus represent SOC dynamics using aggregated climate, land use, and environmental conditions over a certain period. However, in the DAYCENT model, we used annual and daily time series data for climatic and land use conditions to simulate the processes that control SOM retention and stabilization, which could lead to inconsistencies when comparing results between this study and data driven products. Second, outputs based on machine learning models are sensitive to the number of samples used in the training sets. For example, machine learning-based SOC shows higher stocks in the northeastern Great Plains region compared to the DC_{mod} or DC_{def} models (Figure 4). This may be because the region contains thousands of shallow seasonal wetlands with higher SOC stocks averaging between 78 to 109 Mg C ha^{-1} to the depth of 20cm (Tangen and Bansal, 2020). Accounting for the large number of wetlands samples in the training set would likely produce higher SOC stocks in the region. We did not specifically model wetlands SOC and only considered grasslands and croplands, which cover >90% of the land area in the US Great Plains region and as such may have underrepresented these high SOC ecosystems.

3.3 Historical changes in SOC stocks and their distribution

When the baseline SOC (1895-1899 average) values were compared with the current (2001-2005 average) SOC stocks, the modified (DC_{mod}) and default (DC_{def}) models simulated a loss of 1063 Tg C (12%) and 634 Tg C (10%), respectively. On a per unit area basis, DC_{mod} showed higher absolute (17.62 Mg C ha^{-1}) and relative (33%) SOC losses compared to the loss of 10.60 Mg C ha^{-1} (27%) using DC_{def} for croplands. Grasslands showed similar patterns of higher absolute (2.51 Mg C ha^{-1}) and relative (4%) SOC losses using DC_{mod} compared to the loss of 1.06 Mg C ha^{-1} (3%) using DC_{def}. Overall, croplands showed a large and significant loss of C when compared against the baseline SOC using both models, while grasslands showed both losses and gains of SOC during 1895-2005 (Figure 6). The SOC loss from conversion of native vegetation to croplands were on average 14.70 Mg C ha^{-1} and 9.29 Mg C ha^{-1} using DC_{mod} and DC_{def}, respectively. This translates into a relative loss using DC_{mod} that is higher than the loss using DC_{def} by 58% during 1895-2005. For grid cells under native grasslands, DC_{mod} simulated slightly higher average SOC loss (1.96 Mg C ha^{-1}) compared to DC_{def} (1.39 Mg C ha^{-1}).

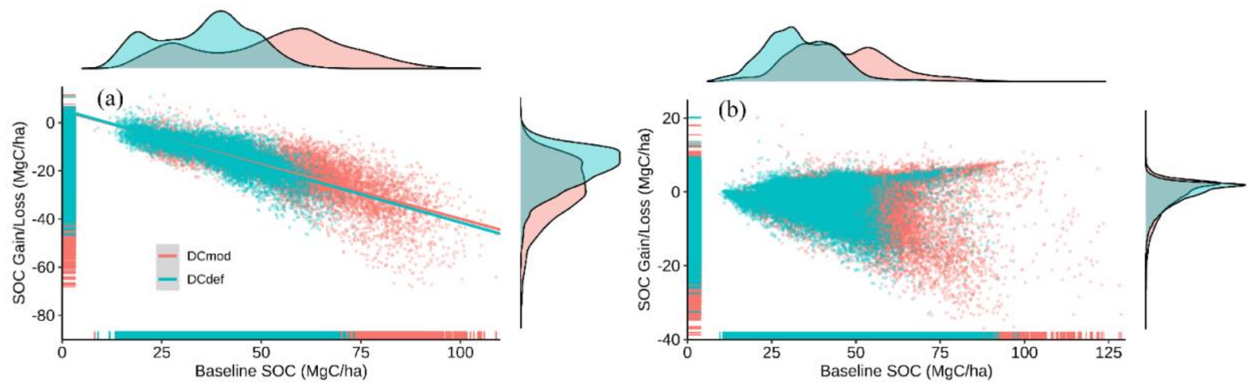

Figure 6. Changes in contemporary (2001-2005 average) SOC after conversion of native vegetation to croplands (a) and under native vegetation (b) as a function of baseline (1895-1899 average) SOC stocks. Negative values are losses while positive values are gains of SOC.

The simulation of total SOC stocks following historical land use under a changing climate is constrained by model parameters that determine the time until decomposition, modified by the interaction of land use intensity with changing climate (Arora and Boer, 2010; Eglin et al., 2010). Land use change can modify total SOC through its effect on individual soil pools, with the POC/active pool more vulnerable to loss compared to the MAOC/slow and PyC/passive pools (Poeplau and Don, 2013). The potential decomposition rates using the modified (DC_{mod}) model were adjusted to match C fraction data such that higher SOC was allocated to rapid and slow cycling pools, which are more vulnerable to loss following land use change and management intensity at decadal to century time scales (Hobley et al., 2017; Sulman et al., 2018). We further compared the historical SOC loss following land use change against other studies to determine the robustness of the new parameterization using DC_{mod}. The SOC loss rate using DC_{mod} are closer to the mean 30 cm loss rate of 17.7 Mg C ha^{-1} (Sanderman et al., 2017b), and relative loss of 42-49% following conversion of forest/pasture to croplands (Guo and Gifford, 2002). However, it is important to note that these previous studies are not directly comparable with the results from this study because of differences in sampling depth, the intensity of land use and the time since disturbance.

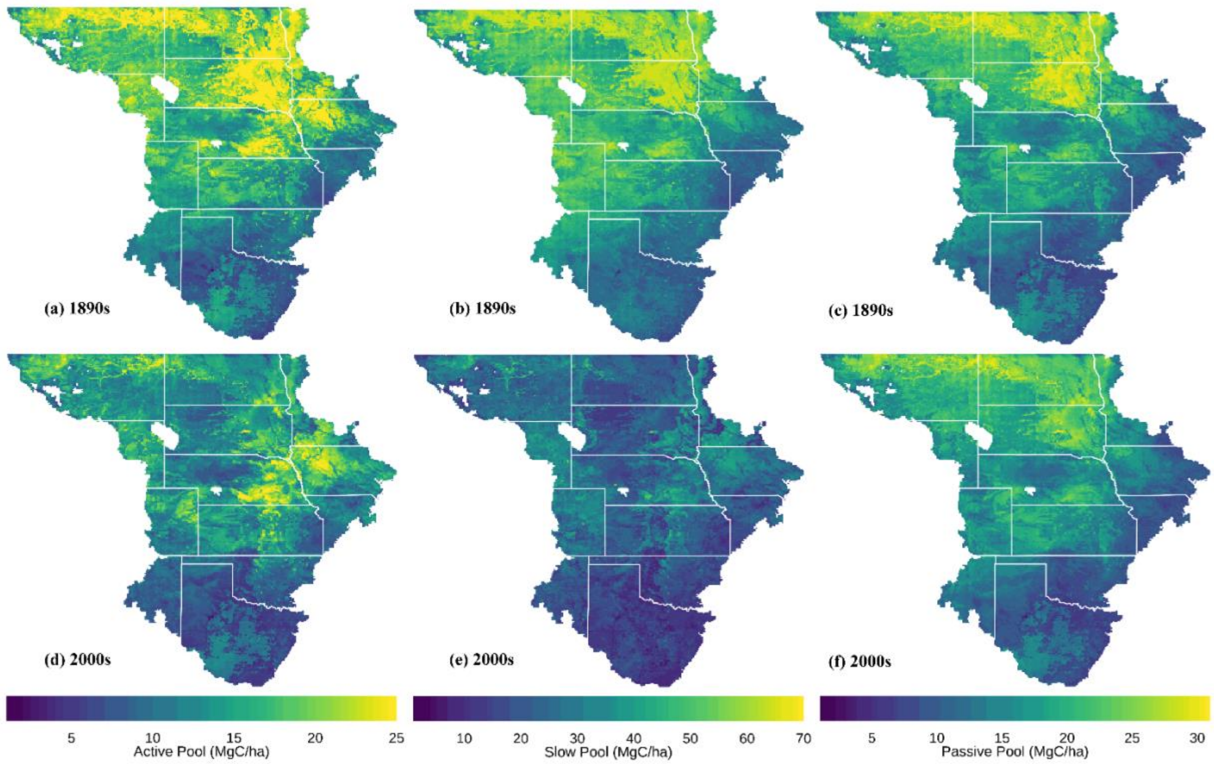

Figure 7. The active, slow, and passive soil pools of SOC stocks (20 cm depth) based on the modified (DC_{mod}) model under native vegetation (1895-1899 average; top maps) and following land cover land use change (2001-2005 average; bottom maps).

Comparison of the total SOC and its distribution in different pools between the two models provided a more nuanced picture of the effect of new parameterization on SOC stocks and the response of SOC to historical land use. The spatial pattern of the SOC stocks showed that the baseline SOC in the active, slow and passive pools simulated by the modified (DC_{mod}) model (Figure 7) were higher than the default (DC_{def}) model (Figure S9). As a result, there were higher SOC losses from the active and slow pools using DC_{mod} compared to DC_{def} (Figure 7, S9). When averaged over all pixels, the cropland SOC loss in the active, and slow, pools were 0.85, 10.09 and gains in the passive pool was 0.34 Mg C ha^{-1}, respectively, using DC_{def}. The DC_{mod} simulated larger SOC loss for all pools with active, slow, and passive pools losing SOC by 1.48, 16.04 and 0.09 Mg C ha^{-1}, respectively. The magnitude of SOC loss from grasslands was lower compared to croplands for all three pools, with the largest SOC loss from the slow pool of 1.45 and 0.49 Mg C ha^{-1} using DC_{mod} and DC_{def} models, respectively. The distribution of SOC to different pools indicated that DC_{def} had 44%, 43% and 13% SOC in the passive, slow, and active pools for croplands, while DC_{mod} had 57% of the total SOC allocated to the slow pool, followed by the passive (23%) and active (20%) pools. For grasslands, both models were consistent in allocating the largest proportion of SOC (59% in default and 70% in modified) to slow pools, followed by passive and active pools.

The differences in the total SOC and their distribution between the models is constrained by the sensitivity of the SOC pools to environmental, climatic, and management factors (Davidson and Janssens, 2006; Dungait et al., 2012; Luo et al., 2016). The SOC stocks in the passive pool are not significantly different between the models at the regional level because the passive pool is less sensitive to environmental, climatic, and management factors, and it has a smaller contribution to total SOC (Collins et al., 2000), the SOC stocks in the passive pool were not significantly different between the models at the regional level. However, the active and slow pools respond strongly to environmental, climatic, and management constraints, which is largely driven by rapidly cycling fresh organic matter input in the active pool, and gradually decomposing detritus in the slow pool (Sherrod et al., 2005). In the DC_{mod}, the potential decomposition rates of the active and slow pools are adjusted, allowing the model to retain more SOC to match with C fraction data. This modification resulted in higher SOC stocks in these pools, which translated into higher total losses despite slower turnover rates relative to DC_{def}. Model modification was necessary not only to match total SOC values but also to simulate the distribution of SOC into the active, slow and passive pools.

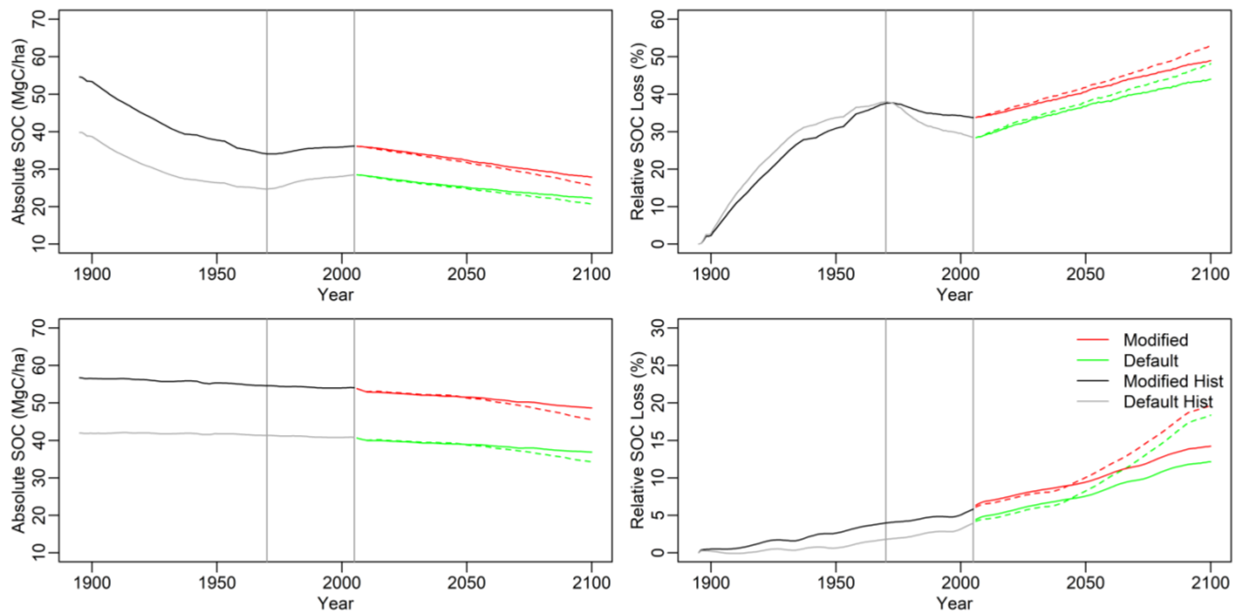

Figure 8. Temporal change in the absolute SOC stocks (20 cm depth) for croplands (a) and grasslands (c) and relative SOC loss compared to the 1895 SOC for croplands (b) and grasslands (d) in response to land use under a changing climate through 2100. The solid and dashed lines after 2006 represent RCP4.5 and RCP8.5 climate scenarios, respectively, both under the A2 land cover change scenario.

3.4 Future changes in SOC stocks and their distribution

Projection of the SOC dynamics in response to land cover change under a changing climate resulted in greater relative changes for both croplands and grasslands using the modified (DC_{mod}) compared to the default (DC_{def}) model (Figure 8). Despite greater rates of loss, by the end of the 21st century, DC_{mod} still simulated higher total SOC stocks compared to DC_{def} model (Table 3). By the end of 21st century, the DC_{mod} simulated total SOC stocks of 2818 and 2563 Tg C for croplands under the RCP4.5 and RCP8.5 scenarios, while the DC_{def} simulated total SOC stocks of

2266 and 2082 Tg C. Native grasslands had higher SOC stocks of 3310 and 3095 Tg C using the DC_{mod} compared to the SOC stocks of 2505 and 2324 Tg C using the DC_{def} under the RCP4.5 and RCP8.5 scenarios, respectively. On a per unit area basis, absolute loss (difference between the 2095s and 2000s) were slightly higher for croplands, with a mean loss rate 10.43 Mg C ha^{-1} compared to 8.44 Mg C ha^{-1} for grasslands using DC_{mod} under the RCP8.5 scenario (Table 3). The DC_{def} also simulated similar trend with slightly higher absolute losses for croplands (7.85 Mg C ha^{-1}) compared to grasslands (6.55 Mg C ha^{-1}) under the RCP8.5 scenario. Relative losses estimated as a percentage of contemporary SOC stocks were higher in croplands (29% for DC_{mod} vs 28% for DC_{def} model) compared to grasslands (16% for both DC_{mod} and DC_{def} model) under the RCP8.5 scenario. Using the DC_{mod}, the SOC loss rate were 33% and 29% higher for croplands and grasslands, respectively, compared to the DC_{def} by the end of the 21st century under the RCP8.5 scenario. While both models simulated total SOC loss over the 21st century, the difference in SOC between models sums to an additional loss of 1252 Tg SOC under the RCP8.5 scenario.

The turnover rates of SOM are primarily driven by temperature and environmental controls with significant impact on the dynamics of total SOC changes at decadal to century time scales (Knorr et al., 2005). The two model versions used the same climate and environmental data and only differ in the turnover rates of the active, slow, and passive pools. Because the sizes of active, and slow pools in the modified (DC_{mod}) model were larger than the default (DC_{def}) model, simulated absolute and relative losses were higher using the DC_{mod} compared to the DC_{def} for croplands. Larger losses using the DC_{mod} are primarily associated with the legacy effects of management intensity and rising temperatures with larger rates of SOC loss from the active, and slow pools (Crow and Sierra, 2018) of DC_{mod} compared to DC_{def}. Additionally, the size of the passive pool in DC_{def} is larger compared to DC_{mod}, and this pool is less vulnerable to land use intensity and warming climate compared to active and slow pools. Thus, there was a disproportionately larger SOC loss driven by the size of the slow pool and the interaction of climate and management intensity using the DC_{mod} compared to the DC_{def}, which translated into larger absolute and relative losses of SOC. For grasslands, we did not include any management driven changes. Both absolute and relative losses of SOC stocks in the grasslands are primarily driven by the warming climate (Jones and Donnelly, 2004), with active and slow pools losing more SOC stocks using DC_{mod} compared to DC_{def}. Future work should consider the interactive effects of grazing management with climate.

Table 3. DAYCENT (modified and default) simulated absolute changes in total and per unit area soil organic carbon (SOC) during the 2000s, 2045s and 2095s for croplands and grasslands in the US Great Plains region

		Total (TgC)				Per Unit Area (MgC/ha)			
	Time	Default (DC_{def})		Modified (DC_{mod})		Default (DC_{def})		Modified (DC_{mod})	
		RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
Croplands	2000s	2113		2717		28.51		36.17	
	2045s	1988	1938	2588	2513	25.20	24.80	32.41	31.87
	2095s	2266	2082	2818	2563	22.31	20.66	27.91	25.87
Grasslands	2000s	3891		5160		40.82		54.05	
	2050s	3531	3523	4674	4659	38.90	38.80	51.51	51.34
	2095s	2505	2324	3310	3095	36.88	34.27	48.65	45.61
Total	2000s	6004		7877		NA		NA	
(Croplands +	2045s	5519	5461	7262	7172	NA	NA	NA	NA
Grasslands)	2095s	4771	4406	6128	5658	NA	NA	NA	NA

Future land use, management intensity, nitrogen content, and climate interact in different ways to control C flow from soil pools with different mean residence times, which ultimately determine total SOC stocks (Deng et al., 2016; Luo et al., 2017; Sulman et al., 2018). Under a warming climate, SOC formed from fresh organic matter inputs controls the size of the active/POC pool, which is further constrained by the intensity of land use and is more vulnerable to loss (Crow and Sierra, 2018; Lavallee et al., 2020). The active/POC pool also acts as a donor to the slow/MAOC pool with C transfer and rates of SOC accumulation increasingly controlled by temperature (Crow and Sierra, 2018). In the DAYCENT, regardless of model version, the size of the active pool is relatively small as fresh organic matter is either decomposed rapidly or quickly enters the slow pool. Because the slow pool has longer residence times ranging from years to decades, the slow pool is less vulnerable to loss and can accrue C when transfer rates from the active pool exceed the rates of decomposition (Collins et al., 2000; Fontaine et al., 2007). In this study, the rates of decomposition due to rising temperatures had a stronger control on the size of the slow pool compared to the transfer of SOC from the active pool. As a result, the slow pool continued to lose SOC under projected climate changes in the future.

4 Conclusions

In this study, we developed an approach to link conceptual soil pools in biogeochemical models against C fraction data predicted using a combination of diffuse reflectance spectroscopy and machine learning. We then quantified the long-term evolution of SOC change and projected the SOC response to future climate and land cover scenarios using the modified (DC_{mod}) model that has been calibrated to C fraction data. Our results demonstrate that matching the active, slow and passive pools against POC, MOAC and PyC data lead to better representation of total SOC stocks and the distribution of SOC into different pools. With the updated model, the long-term legacy effect of past agricultural management results in larger absolute and relative losses of SOC compared to the default (DC_{def}) model. Projecting the SOC response to climate and land cover change into the future (2005-2100) indicates that the new model modification (DC_{mod}) increases SOC losses by 2100 by 32% and 28% for croplands and grasslands, respectively, under the RCP8.5 scenario compared to using the DC_{def} model.

There are several study limitations that need to be addressed in our future work. First, new modeling efforts should also consider quantifying how changes in aboveground biomass inputs quantity and quality affect SOC dynamics given mixed results in agricultural systems in response to litter inputs (Halvorson et al., 2002; Sanderman et al., 2017a). Second, current models rely on using clay content to modify rates of SOM stabilization and turnover, but recent research has shown that other soil physicochemical properties such as exchangeable calcium and extractable iron and aluminum are stronger predictors of SOM content (Rasmussen et al., 2018). Third, new modeling efforts should constrain model parameters affecting SOC dynamics by integrating them with data-driven modeling and long-term experimental data (Jandl et al., 2014). Finally, given the paucity of data related to C fractions, there is increasing need for measurement and modeling of C fractions across a wide range of environmental and management gradients (Luo et al., 2017). Despite these limitations, we have shown that models calibrated to pool sizes by matching with C fractions can improve long-term SOC predictions by more accurately representing soil C transformations in response to climate, land cover and land use change.

Code and Data Availability:

The DAYCENT model source code is available in Harvard dataverse repository (https://dataverse.harvard.edu/dataverse/daycent45). The new parameterization scheme and scripts for regional model simulation are available in github (https://github.com/whrc/DAYCENT-soil-carbon-pools). Input data for driving the models are freely available online from different sources and have been cited appropriately in the manuscript. Long term ecological data are part of United States Department of Agriculture – Agricultural Research Service and can be requested from the references listed in Table 1.

Author Contributions: S.D., C.S, and J.S designed the study and model development. S.D. performed model improvement, calibration, validation and regional historical and future simulation. All authors contributed to the manuscript.

Competing Interest: The authors declare that they have no conflict of interest.

Acknowledgements

Funding for this research was provided by USDA NIFA award #2017-67003-26481. We thank Melannie D Hartman at Colorado State University for providing access to the DAYCENT model and help with running the model. We also thank staff at the USDA National Soil Survey Center (NSSC) Kellogg Soil Survey Laboratory (KSSL) for providing access to the soil characterization database. This research also used data from the Long-Term Agroecosystem Research (LTAR) network and Columbia Plateau Conservation Research Center (CPCRC), which are both supported by the United States Department of Agriculture**.** The NSF Long-term Ecological Research Program (DEB 1832042) and Michigan State University AgBioResearch provided funding for the data and soil samples from the Kellogg Biological Station.

References

Arora, V.K., Boer, G.J., 2010. Uncertainties in the 20th century carbon budget associated with land use change. Global Change Biology 16, 3327–3348.

Baker, N.T., 2011. Tillage Practices in the Conterminous United States, 1989-2004–datasets Aggregated by Watershed. US Department of the Interior, US Geological Survey Reston, Virginia.

Baldock, J.A., Hawke, B., Sanderman, J., Macdonald, L.M., 2013a. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Research 51, 577–595.

Baldock, J.A., Sanderman, J., Macdonald, L.M., Puccini, A., Hawke, B., Szarvas, S., McGowan, J., 2013b. Quantifying the allocation of soil organic carbon to biologically significant fractions. Soil Research 51, 561–576.

Basso, B., Gargiulo, O., Paustian, K., Robertson, G.P., Porter, C., Grace, P.R., Jones, J.W., 2011. Procedures for initializing soil organic carbon pools in the DSSAT-CENTURY model for agricultural systems. Soil Science Society of America Journal 75, 69–78.

Batjes, N.H., 2016. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma 269, 61–68.

Cagnarini, C., Renella, G., Mayer, J., Hirte, J., Schulin, R., Costerousse, B., Della Marta, A., Orlandini, S., Menichetti, L., 2019. Multi-objective calibration of RothC using measured carbon stocks and auxiliary data of a long-term experiment in Switzerland. European Journal of Soil Science 70, 819–832.

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., 2014. Global covariation of carbon turnover times with climate in terrestrial ecosystems. Nature 514, 213–217.

Cavigelli, M.A., Teasdale, J.R., Conklin, A.E., 2008. Long-term agronomic performance of organic and conventional field crops in the mid-Atlantic region. Agronomy Journal 100, 785–794.

Chan, K.Y., Conyers, M.K., Li, G.D., Helyar, K.R., Poile, G., Oates, A., Barchia, I.M., 2011. Soil carbon dynamics under different cropping and pasture management in temperate Australia: Results of three long-term experiments. Soil Research 49, 320–328.

Christensen, B.T., 1996. Matching measurable soil organic matter fractions with conceptual pools in simulation models of carbon turnover: revision of model structure. Evaluation of soil organic matter models 143–159.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., 2014. Carbon and other biogeochemical cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 465–570.

Collins, H.P., Elliott, E.T., Paustian, K., Bundy, L.G., Dick, W.A., Huggins, D.R., Smucker, A.J.M., Paul, E.A., 2000. Soil carbon pools and fluxes in long-term corn belt agroecosystems. Soil Biology and Biochemistry 32, 157–168.

Crow, S.E., Sierra, C.A., 2018. Dynamic, intermediate soil carbon pools may drive future responsiveness to environmental change. Journal of environmental quality 47, 607–616.

Crowther, T.W., Todd-Brown, K.E., Rowe, C.W., Wieder, W.R., Carey, J.C., Machmuller, M.B., Snoek, B.L., Fang, S., Zhou, G., Allison, S.D., 2016. Quantifying global soil carbon losses in response to warming. Nature 540, 104–108.

Czimczik, C.I., Masiello, C.A., 2007. Controls on black carbon storage in soils. Global Biogeochemical Cycles 21.

Daly, C., Bryant, K., 2013. The PRISM climate and weather system—an introduction. Corvallis, OR: PRISM climate group.

Dangal, S.R., Sanderman, J., 2020. Is Standardization Necessary for Sharing of a Large Mid-Infrared Soil Spectral Library? Sensors 20, 6729.

Dangal, S.R., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Systems 3, 11.

Davidson, E.A., Janssens, I.A., 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature 440, 165–173.

Del Grosso, S., Ojima, D., Parton, W., Mosier, A., Peterson, G., Schimel, D., 2002. Simulated effects of dryland cropping intensification on soil organic matter and greenhouse gas exchanges using the DAYCENT ecosystem model. Environmental pollution 116, S75–S83.

Del Grosso, S.J., Parton, W.J., Mosier, A.R., Hartman, M.D., Brenner, J., Ojima, D.S., Schimel, D.S., 2001. Simulated interaction of carbon dynamics and nitrogen trace gas fluxes using the DAYCENT model. Modeling carbon and nitrogen dynamics for soil management 303–332.

Deng, L., Zhu, G., Tang, Z., Shangguan, Z., 2016. Global patterns of the effects of land-use changes on soil carbon stocks. Global Ecology and Conservation 5, 127–138.

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Pinto, M.C., Casanova-Katny, A., Muñoz, C., Boudin, M., Venegas, E.Z., 2015. Soil carbon storage controlled by interactions between geochemistry and climate. Nature Geoscience 8, 780–783.

Dungait, J.A., Hopkins, D.W., Gregory, A.S., Whitmore, A.P., 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. Global Change Biology 18, 1781–1796.

Eglin, T., Ciais, P., Piao, S.L., Barré, P., Bellassen, V., Cadule, P., Chenu, C., Gasser, T., Koven, C., Reichstein, M., 2010. Historical and future perspectives of global soil carbon response to climate and land-use changes. Tellus B: Chemical and Physical Meteorology 62, 700–718.

Falcone, J.A., LaMotte, A.E., 2016. National 1-kilometer rasters of selected census of agriculture statistics allocated to land use for the time period 1950 to 2012. US Geological Survey Data Release.

Fontaine, S., Barot, S., Barré, P., Bdioui, N., Mary, B., Rumpel, C., 2007. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. Nature 450, 277–280.

Gollany, H., 2016. CQESTR simulation of dryland agroecosystem soil organic carbon changes under climate change scenarios. Synthesis and Modeling of Greenhouse Gas Emissions and Carbon Storage in Agricultural and Forest Systems to Guide Mitigation and Adaptation 6, 59–87.

Grandy, A.S., Sinsabaugh, R.L., Neff, J.C., Stursova, M., Zak, D.R., 2008. Nitrogen deposition effects on soil organic matter chemistry are linked to variation in enzymes, ecosystems and size fractions. Biogeochemistry 91, 37–49.

Guo, L.B., Gifford, R.M., 2002. Soil carbon stocks and land use change: a meta analysis. Global change biology 8, 345–360.

Halvorson, A.D., Wienhold, B.J., Black, A.L., 2002. Tillage, nitrogen, and cropping system effects on soil carbon sequestration. Soil science society of America journal 66, 906–912.

Hartman, M.D., Merchant, E.R., Parton, W.J., Gutmann, M.P., Lutz, S.M., Williams, S.A., 2011. Impact of historical land-use changes on greenhouse gas exchange in the US Great Plains, 1883–2003. Ecological Applications 21, 1105–1119.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS one 12, e0169748.

Hicks, W., Rossel, R.V., Tuomi, S., 2015. Developing the Australian mid-infrared spectroscopic database using data from the Australian Soil Resource Information System. Soil Research 53, 922–931.

Hobley, E., Baldock, J., Hua, Q., Wilson, B., 2017. Land-use contrasts reveal instability of subsoil organic carbon. Global Change Biology 23, 955–965.

Hsieh, Y.-P., 1993. Radiocarbon signatures of turnover rates in active soil organic carbon pools. Soil Science Society of America Journal 57, 1020–1022.

Ingram, L.J., Stahl, P.D., Schuman, G.E., Buyer, J.S., Vance, G.F., Ganjegunte, G.K., Welker, J.M., Derner, J.D., 2008. Grazing impacts on soil carbon and microbial communities in a mixed-grass ecosystem. Soil Science Society of America Journal 72, 939–948.

Jandl, R., Rodeghiero, M., Martinez, C., Cotrufo, M.F., Bampa, F., van Wesemael, B., Harrison, R.B., Guerrini, I.A., Richter Jr, D. deB, Rustad, L., 2014. Current status, uncertainty and future needs in soil organic carbon monitoring. Science of the total environment 468, 376–383.

Janssens, I.A., Dieleman, W., Luyssaert, S., Subke, J.-A., Reichstein, M., Ceulemans, R., Ciais, P., Dolman, A.J., Grace, J., Matteucci, G., 2010. Reduction of forest soil respiration in response to nitrogen deposition. Nature geoscience 3, 315–322.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological applications 10, 423–436.

Jones, M.B., Donnelly, A., 2004. Carbon sequestration in temperate grassland ecosystems and the influence of management, climate and elevated CO2. New Phytologist 164, 423–439.

Kelly, R.H., Parton, W.J., Hartman, M.D., Stretch, L.K., Ojima, D.S., Schimel, D.S., 2000. Intra-annual and interannual variability of ecosystem processes in shortgrass steppe. Journal of Geophysical Research: Atmospheres 105, 20093–20100.

Kittel, T.G., Rosenbloom, N.A., Royle, J.A., Daly, C., Gibson, W.P., Fisher, H.H., Thornton, P., Yates, D.N., Aulenbach, S., Kaufman, C., 2004. VEMAP phase 2 bioclimatic database. I. Gridded historical (20th century) climate for modeling ecosystem dynamics across the conterminous USA. Climate Research 27, 151–170.

Klein Goldewijk, K., Beusen, A., Doelman, J., Stehfest, E., 2017. Anthropogenic land use estimates for the Holocene–HYDE 3.2. Earth System Science Data 9, 927–953.

Knorr, W., Prentice, I.C., House, J.I., Holland, E.A., 2005. Long-term sensitivity of soil carbon turnover to warming. Nature 433, 298–301.

Lal, R., 2018. Digging deeper: A holistic perspective of factors affecting soil organic carbon sequestration in agroecosystems. Global Change Biology 24, 3285–3301.

Lal, R., 2004. Carbon sequestration in dryland ecosystems. Environmental management 33, 528–544.

Lavallee, J.M., Soong, J.L., Cotrufo, M.F., 2020. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. Global Change Biology 26, 261–273.

Lee, J., Viscarra Rossel, R.A., 2020. Soil carbon simulation confounded by different pool initialisation. Nutrient Cycling in Agroecosystems 116, 245–255.

Liebig, M.A., Gross, J.R., Kronberg, S.L., Phillips, R.L., 2010. Grazing management contributions to net global warming potential: A long-term evaluation in the Northern Great Plains. Journal of Environmental Quality 39, 799–809.

Luo, Y., Ahlström, A., Allison, S.D., Batjes, N.H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E.A., Finzi, A., 2016. Toward more realistic projections of soil carbon dynamics by Earth system models. Global Biogeochemical Cycles 30, 40–56.

Luo, Z., Feng, W., Luo, Y., Baldock, J., Wang, E., 2017. Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions. Global Change Biology 23, 4430–4439.

Metherell, A., Harding, L., Cole, C., Parton, W., 1994. CENTURY soil organic matter model environment, technical documentation, agroecosystem version 4.0 GPSR Technical Report No. 4. Great Plains System Research Unit, USDA-ARS, Fort Collins, CO.

Nachtergaele, F., van Velthuizen, H., Verelst, L., 2012. Harmonized World Soil Database Version 1.2. Food and Agriculture Organization of the United Nations (FAO). International Institute for

Applied Systems Analysis (IIASA), ISRIC-World Soil Information, Institute of Soil Science–Chinese Academy of Sciences (ISSCAS), Joint Research Centre of the European Commission (JRC).

Ogle, S.M., Breidt, F.J., Easter, M., Williams, S., Killian, K., Paustian, K., 2010. Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. Global Change Biology 16, 810–822.

Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environmental management 54, 1249–1266.

Page, K.L., Dalal, R.C., Dang, Y.P., 2014. How useful are MIR predictions of total, particulate, humus, and resistant organic carbon for examining changes in soil carbon stocks in response to different crop management? A case study. Soil Research 51, 719–725.

Parton, W.J., Hartman, M., Ojima, D., Schimel, D., 1998. DAYCENT and its land surface submodel: description and testing. Global and planetary Change 19, 35–48.

Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S., 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Science Society of America Journal 51, 1173–1179.

Parton, W.J., Stewart, J.W., Cole, C.V., 1988. Dynamics of C, N, P and S in grassland soils: a model. Biogeochemistry 5, 109–131.

Paul, E.A., Morris, S.J., Bohm, S., 2001. The determination of soil C pool sizes and turnover rates: biophysical fractionation and tracers. Assessment methods for soil carbon 14, 193–206.

Poeplau, C., Don, A., 2013. Sensitivity of soil organic carbon stocks and fractions to different land-use changes across Europe. Geoderma 192, 189–201.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Science Society of America Journal 82, 186–201.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013. The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma 195, 268–279.

Rasmussen, C., Heckman, K., Wieder, W.R., Keiluweit, M., Lawrence, C.R., Berhe, A.A., Blankinship, J.C., Crow, S.E., Druhan, J.L., Pries, C.E.H., 2018. Beyond clay: towards an improved set of variables for predicting soil organic matter content. Biogeochemistry 137, 297–306.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., Rafaj, P., 2011. RCP 8.5—A scenario of comparatively high greenhouse gas emissions. Climatic change 109, 33–57.

Sanderman, J., Baldock, J.A., Dangal, S.R., Ludwig, S., Potter, S., Rivard, C., Savage, K., 2021. Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy. Biogeochemistry 1–18.

Sanderman, J., Creamer, C., Baisden, W.T., Farrell, M., Fallon, S., 2017a. Greater soil carbon stocks and faster turnover rates with increasing agricultural productivity. Soil 3, 1–16.

Sanderman, J., Hengl, T., Fiske, G.J., 2017b. Soil carbon debt of 12,000 years of human land use. Proceedings of the National Academy of Sciences 114, 9575–9580.

Sanford Jr, R.L., Parton, W.J., Ojima, D.S., Lodge, D.J., 1991. Hurricane effects on soil organic matter dynamics and forest production in the Luquillo Experimental Forest, Puerto Rico: results of simulation modeling. Biotropica 364–372.

Schmer, M.R., Jin, V.L., Wienhold, B.J., Varvel, G.E., Follett, R.F., 2014. Tillage and residue management effects on soil carbon and nitrogen under irrigated continuous corn. Soil Science Society of America Journal 78, 1987–1996.

Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49–56.

Schwalm, C.R., Glendon, S., Duffy, P.B., 2020. RCP8. 5 tracks cumulative CO2 emissions. Proceedings of the National Academy of Sciences 117, 19656–19657.
Sherrod, L.A., Peterson, G.A., Westfall, D.G., Ahuja, L.R., 2005. Soil organic carbon pools after 12 years in no-till dryland agroecosystems. Soil Science Society of America Journal 69, 1600–1608.
Shi, Z., Crowell, S., Luo, Y., Moore, B., 2018. Model structures amplify uncertainty in predicted soil carbon responses to climate change. Nature communications 9, 1–11.
Sindelar, A.J., Schmer, M.R., Jin, V.L., Wienhold, B.J., Varvel, G.E., 2015. Long-term corn and soybean response to crop rotation and tillage. Agronomy Journal 107, 2241–2252.
Sinsabaugh, R.L., Gallo, M.E., Lauber, C., Waldrop, M.P., Zak, D.R., 2005. Extracellular enzyme activities and soil organic matter dynamics for northern hardwood forests receiving simulated nitrogen deposition. Biogeochemistry 75, 201–215.
Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant and soil 241, 155–176.
Skjemstad, J.O., Spouncer, L.R., Cowie, B., Swift, R.S., 2004. Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. Soil Research 42, 79–88.
Sohl, T.L., Sleeter, B.M., Sayler, K.L., Bouchard, M.A., Reker, R.R., Bennett, S.L., Sleeter, R.R., Kanengieter, R.L., Zhu, Z., 2012. Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States. Agriculture, Ecosystems & Environment 153, 1–15.
Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., De Courcelles, V. de R., Singh, K., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture, Ecosystems & Environment 164, 80–99.
Sulman, B.N., Moore, J.A., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., Sridhar, B., Hartman, M.D., Wang, G., Wieder, W.R., 2018. Multiple models and experiments underscore large uncertainty in soil carbon dynamics. Biogeochemistry 141, 109–123.
Syswerda, S.P., Corbin, A.T., Mokma, D.L., Kravchenko, A.N., Robertson, G.P., 2011. Agricultural management and soil carbon storage in surface vs. deep layers. Soil Science Society of America Journal 75, 92–101.
Tangen, B.A., Bansal, S., 2020. Soil organic carbon stocks and sequestration rates of inland, freshwater wetlands: Sources of variability and uncertainty. Science of The Total Environment 749, 141444.
Thomson, A.M., Calvin, K.V., Smith, S.J., Kyle, G.P., Volke, A., Patel, P., Delgado-Arias, S., Bond-Lamberty, B., Wise, M.A., Clarke, L.E., 2011. RCP4. 5: a pathway for stabilization of radiative forcing by 2100. Climatic change 109, 77–94.
Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., Cook, R., 2012. Daymet: Daily surface weather on a 1 km grid for North America, 1980-2008. Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC.
Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D.N., Schwalm, C.R., Michalak, A.M., Cook, R., Ciais, P., Hayes, D., 2015. Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. Global Biogeochemical Cycles 29, 775–792.
Todd-Brown, K.E.O., Randerson, J.T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., Wu, T., 2014. Changes in soil organic carbon storage predicted by Earth system models during the 21st century. Biogeosciences 11, 2341–2356.
Torn, M.S., Kleber, M., Zavaleta, E.S., Zhu, B., Field, C.B., Trumbore, S.E., 2013. A dual isotope approach to isolate soil carbon pools of different turnover times. Biogeosciences 10, 8067–8081.
Torn, M.S., Trumbore, S.E., Chadwick, O.A., Vitousek, P.M., Hendricks, D.M., 1997. Mineral control of soil organic carbon storage and turnover. Nature 389, 170–173.

Trumbore, S.E., 1997. Potential responses of soil organic carbon to global environmental change. Proceedings of the National Academy of Sciences 94, 8284–8291.

Van Groenigen, K.J., Qi, X., Osenberg, C.W., Luo, Y., Hungate, B.A., 2014. Faster decomposition under increased atmospheric CO2 limits soil carbon storage. Science 344, 508–509.

Wieder, W.R., Hartman, M.D., Sulman, B.N., Wang, Y.-P., Koven, C.D., Bonan, G.B., 2018. Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. Global change biology 24, 1563–1579.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., 2019. Soil organic carbon storage as a key function of soils-A review of drivers and indicators at various scales. Geoderma 333, 149–162.

Zimmermann, M., Leifeld, J., Schmidt, M.W.I., Smith, P., Fuhrer, J., 2007. Measured soil organic matter fractions can be related to pools in the RothC model. European Journal of Soil Science 58, 658–667.