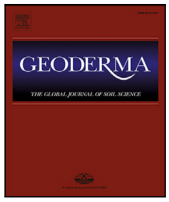
Geoderma 437 (2023) 116565

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Is macroporosity controlled by complexed clay and soil organic carbon?

Aaron N. Koop [a], Daniel R. Hirmas [b,*], Sharon A. Billings [c], Li Li [d], Alejandro Cueva [e], Xi Zhang [f], Hang Wen [g], Attila Nemes [h], Lígia F.T. Souza [i], Hoori Ajami [j], Alejandro N. Flores [k], Aoesta K. Rudick [l], Annalise Guthrie [m], Lola M. Klamm [m], Micah Unruh [m], Pamela L. Sullivan [n]

[a] Department of Geography and Atmospheric Science, University of Kansas, Lawrence, KS, USA
[b] Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, USA
[c] Department of Ecology and Evolutionary Biology and Kansas Biological Survey and Center for Ecological Research, University of Kansas, Lawrence, KS, USA
[d] Department of Civil and Environmental Engineering, Pennsylvania State University, State College, PA, USA
[e] Departamento de Ciencias de la Sustentabilidad, El Colegio de la Frontera Sur, Unidad Villahermosa, Tabasco, Mexico
[f] Red River Research Station and School of Plant, Environment, and Soil Sciences, Louisiana State University-Agricultural Center, Bossier City, LA, USA
[g] School of Earth System Science, Institute of Surface-Earth System Science, Tianjin University, Tianjin, China
[h] Division of Environment and Natural Resources, Norwegian Institute of Bioeconomy Research and Faculty of Environment and Natural Resources, Norwegian University of Life Sciences, Ås, Norway
[i] Department of Earth Sciences, Memorial University of Newfoundland, A1C5S7, St. John's, NL, Canada
[j] Department of Environmental Sciences, University of California Riverside, Riverside, CA, USA
[k] Department of Geosciences, Boise State University, Boise, ID, USA
[l] Kansas Biological Survey and Center for Ecological Research, University of Kansas, Lawrence, KS, USA
[m] Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, USA
[n] College of Earth, Ocean, and Atmospheric Science, Oregon State University, OR, USA

ARTICLE INFO

Handling Editor: Morgan Cristine L.S.

Keywords:
Organo-mineral complexes
Soil structure
Effective porosity
Continental-scale macroporosity

ABSTRACT

Multi-scale evidence of rapid, climate-induced soil structural changes occurring at yearly to decadal timescales is mounting. As a result, it has become increasingly important to identify the properties and mechanisms controlling the development and maintenance of soil structure and associated macroporosity. This is especially relevant since macroporosity has disproportionate effects on saturated hydraulic conductivity (K_{sat}) which strongly influences water storage and flux, thus, affecting the water cycle. In this study, we use decision trees and piecewise linear regression to assess the influence of soil and climate properties on effective porosity (EP; a proxy of macroporosity) in both surface and subsurface horizons under varying land-use and management practices. Data from 1,491 pedons (3,679 horizons) spanning five ecoregions representing bioclimate (e.g., potential vegetation) across the conterminous US demonstrate that, at a continental scale, EP in surface (A) and subsurface (B) horizons is strongly dependent on the complexed fraction of the total mass of soil organic carbon (SOC) and clay; a combined fraction that we refer to as complexed organic carbon and clay (COCC). EP showed a slight positive response to COCC in A horizons but increased steeply with increasing COCC in B horizons. This is because the smaller values of COCC in B horizons reflect a larger pool of clay that has a greater potential to accommodate and complex additions of SOC promoting stronger organo-mineral bonds and the concomitant development and maintenance of soil structure in these horizons. In contrast, larger values of COCC in A horizons reflect conditions where all or most of the clay fraction is effectively complexed with SOC resulting in a larger pool of non-complexed soil organic matter with varying contrasting effects on macroporosity that ultimately mute the response of EP to increases in COCC. In surface horizons, indirect factors such as mean annual precipitation and land use were important predictors of EP, whereas COCC was more influential in controlling EP within the subsoil. The EP-COCC relationship also holds within ecoregions but its effect is mitigated by soil and climate interactions suggesting that the effect of climate on this relationship is indirect and complex. Plowed surface horizons and horizons underlying plowed layers showed greater homogenization (due to disturbance effects reducing heterogeneity in the soil) as well as a reduction in the magnitude and rate of change of EP as a function of COCC compared to undisturbed horizons. Our findings suggest that the complexed fraction of clay and SOC is important for controlling macroporosity and K_{sat} at ecoregion scales and that the EP-COCC relationship may be an important framework for understanding and predicting future land use- and climate-induced changes in soil hydraulic properties.

https://doi.org/10.1016/j.geoderma.2023.116565
Received 16 August 2022; Received in revised form 7 June 2023; Accepted 7 June 2023
Available online 26 June 2023

1. Introduction

Emerging evidence at plot, hillslope, and continental scales indicates that soil structure (i.e., the arrangement of soil particles and pores) is changing faster than previously thought – potentially on decadal timescales – in response to shifts in precipitation regimes (Hirmas et al., 2018; Caplan et al., 2019). These structural changes alter soil macroporosity and saturated hydraulic conductivity (K_{sat}) – properties in soil that control water storage and flux (e.g., > 70% of water flux through soil can be controlled by macropores; Watson & Luxmoore, 1986) – and, thus, the water cycle. Specifically, more humid conditions appear to promote a reduction in both macroporosity and K_{sat}, while drier conditions promote an increase in these properties (Hirmas et al., 2018). The mechanisms governing these rapid responses of soil structure to changing precipitation regimes remain elusive. Biotic processes are the most likely explanation for changes in soil structure on such short timescales (Sullivan et al., 2022). However, given the suite of plant and microbial dynamics governing the depth distribution of water, organic carbon, CO_2 fluxes (root and microbe), organic acid production, oxygen availability, and physical mixing via bioturbation by macrofauna, determining which specific biologically-controlled processes are most responsible for these changes and how they are coupled to alterations of soil fabric is difficult (Six et al., 2004; Wen et al., 2021; Sullivan et al., 2022; Li et al., 2022).

Since macroporosity strongly depends on particle-size distribution (PSD) and soil organic carbon (SOC) content (Nemes et al., 2005), mechanisms involved in biophysical processes responsible for soil structural changes and alterations of soil material also appear likely to operate in the context of soil texture-SOC interactions. While it is known that macroporosity and K_{sat} exhibit a positive relationship (Ahuja et al., 1984; Watson and Luxmoore, 1986), soil organic matter (SOM) and K_{sat} exhibit a more complex relationship (i.e., both positive and negative relationships) (Nemes et al., 2005; Araya and Ghezzehei, 2019). In general, increases in SOM have been linked to increased K_{sat} due to the influence of SOM on soil aggregation (and stability) and associated pore-size distribution (Hudson, 1994; Saxton and Rawls, 2006). This positive association is attributed to increases in SOM content promoting soil aggregation and lower bulk density, which leads to greater porosity and higher K_{sat} (Nemes et al., 2005). Explanations for the negative association between high SOM and K_{sat} include: (1) the retention of water by SOM reducing hydraulic conductivity by allowing less water to flow freely, and (2) the effect of SOM on soil structure that may allow more aggregated material to replace larger cracks and clods resulting in more tortuous and thin pathways for water to flow in addition to the potential for mobile organic colloids to clog soil pores (Nemes et al., 2005). Other studies have attributed negative relationships between SOM and K_{sat} to SOM causing transient sub-critical soil water repellency and inhibiting water flow due to reduced wettability (Wang et al., 2009, 2013; Jarvis et al., 2013; Jorda et al., 2015). These studies illustrate the varying effects of SOM on hydraulic properties.

Recent work has shown that soil texture influences the direction of the association between SOM and K_{sat}. For example, using a machine learning approach, Araya and Ghezzehei (2019) predicted increasing K_{sat} with increasing SOC content for all soil textures except the two coarsest classes (i.e., loamy sand and sand). Their trends suggest that SOC-induced aggregation increases the relative proportion of large pores (inter-aggregate pores such as biopores and macropores) in fine and medium textured soils but increases fine intra-aggregate pores in coarse-textured soils thereby shrinking larger pores and reducing K_{sat} with decreases in K_{sat} at SOC ≥ 3% predicted for loamy sand and sand textures (Araya and Ghezzehei, 2019). Similar effects of SOC in reducing K_{sat} of coarser soils while increasing that of finer textured soils were found by Rawls et al. (2004) and Nemes et al. (2005) where soils with 50% sand and clay contents ranging between approximately 25 to 45% resulted in lower predicted K_{sat} for soils with higher SOC (5%) compared to those with lower SOC (1% ≤ SOC ≤ 3%). Both Araya and Ghezzehei (2019) and Nemes et al. (2005) note the importance of better understanding this complex relationship between SOC content, soil texture, and K_{sat}.

In this study, our overall goal is to examine the degree to which complexation of clay and SOC is responsible for broad-scale changes in macroporosity and to understand how climate, land use, vegetation, and depth interact to shape the effects of these organo-mineral associations. Specifically, we use data both from across the conterminous US and within five large US ecoregions to explore the linkage between soil aggregation and macroporosity. We use effective porosity (EP) – calculated as the difference between total porosity and field capacity water content – as a proxy for macroporosity because this information is readily obtainable in many datasets and because K_{sat} can be estimated from EP using a form of the Kozeny-Carman equation proposed by Rawls et al. (1998). In well-drained soils across a climatic gradient, we hypothesized that lower SOC in arid environments would lead to a strong positive association with EP because any addition of SOC in these settings will increase the propensity for organo-mineral complexation and aggregation and enhance the proportion of inter-aggregate macropores. In humid environments, the relationship between organo-mineral complexation, aggregation, and macroporosity is dampened by the infilling of macropores by microaggregates and particulate organic matter. In soils that are disturbed by agricultural practices, we hypothesized that mechanical disaggregation and reductions in SOC and organo-mineral complexes would reduce EP relative to undisturbed soils due to direct land-use effects. Because macroporosity can have disproportionate effects on K_{sat} (Giménez and Hirmas, 2017), understanding the properties and processes controlling soil structural development and macroporosity may provide a foundation for projecting soil ecosystem function.

2. Methods

2.1. Datasets

In this study, we acquired field-based pedon information and laboratory data collected through the United States Department of Agriculture–Natural Resources Conservation Service (USDA–NRCS) National Cooperative Soil Survey (NCSS) Characterization Database (https://ncsslabdatamart.sc.egov.usda.gov). We selected pedons distributed across broad geographic extents in the US including soils with large and widespread taxonomic diversity (i.e., Entisols, Inceptisols, Mollisols, Aridisols, Alfisols, and Ultisols); poorly drained soils (i.e., "very poorly drained", "poorly drained", and "somewhat poorly drained" soil drainage classes) and soils with lithologic discontinuities were removed. These pedons were grouped into United States Environmental Protection Agency (US EPA)/United States Geological Survey (USGS) level I ecoregions (Omernik and Griffith, 2014). Following Hirmas et al. (2018), soil samples were also grouped into surface layers (A horizons) that occurred in the upper 25 cm and subsurface layers (B horizons) that occurred within 25–100 cm of the land surface. Additionally, horizon data were grouped into Ap (plowed) and Bp horizons (defined here as B horizons that occur within a profile containing a shallower Ap horizon) resulting in four categories of horizons (i.e., A, Ap, B, and Bp). In this study, Ap horizons exhibited evidence of plowing (e.g., platy soil structure) at the time of sampling, but were not necessarily under current agricultural production. To minimize

* Corresponding author.
E-mail address: dhirmas@ttu.edu (D.R. Hirmas).

effects from confounding variables, morphological horizons containing concentrations of aggregating agents, such as carbonate (e.g., Bk), or horizons that were indurated (e.g., Bqm) were not considered in this study. Calcium carbonate equivalent data was not ubiquitous across our sites and was therefore not included in these analyses. In addition, mean annual precipitation (MAP) and mean annual temperature (MAT) products from the Parameter-elevation Regressions on Independent Slopes Model (PRISM; PRISM Climate Group and Oregon State University, 2021) were merged with the NCSS data. Ultimately, we selected level I ecoregions that had all four categories of horizons (i.e., A, Ap, B, and Bp) which included Eastern Temperate Forests (ETF), Great Plains (GP), Mediterranean California (MC), North American Deserts (NAD), and Northwestern Forested Mountains (NWFM). These ecoregions represent changing bioclimatic conditions (e.g., potential vegetation) across the conterminous US. In addition, MAP and MAT were also included in these analyses since steep climatic gradients can occur within individual ecoregions.

To better understand soil, climate, land use, and depth controls on macroporosity at a continental scale and across ecoregions, measured, calculated, and/or derived properties from A, Ap, B, and Bp horizons as well as MAP and MAT were included in the analyses. These included PSD, SOM (which was converted from SOC using the standard SOM:SOC mass ratio of 1.72), field-capacity volumetric water content (FC), complexed clay (CC), complexed SOC (COC), mass fraction of water-stable aggregates (WSA), total porosity (TP), and EP.

CC and COC were calculated following Dexter et al. (2008):

$$\mathrm{CC} = \mathrm{IF}\left[n\mathrm{SOC} < \mathrm{clay}\right] \mathrm{THEN}\,[n\mathrm{SOC}]\,\mathrm{ELSE}\,\left[\mathrm{clay}\right] \tag{1}$$

$$\mathrm{COC} = \mathrm{IF}\left[\mathrm{SOC} < \frac{\mathrm{clay}}{n}\right] \mathrm{THEN}\,[\mathrm{SOC}]\,\mathrm{ELSE}\,\left[\frac{\mathrm{clay}}{n}\right] \tag{2}$$

where it was assumed that n g of clay complex with 1 g of SOC with n assumed to be 10 following Dexter et al. (2008). Units for CC and COC are g 100^{-1} g^{-1} (Dexter et al., 2008).

We note that Dexter et al. (2008) found that COC (calculated with a SOC:clay ratio of 1:10) controlled soil physical properties and behavior rather than the total amount of SOC. In contrast, Johannes et al. (2017) found that total SOC controls physical properties rather than COC, although they did observe that SOC:clay ratio decreases with decreasing soil structure quality. As a result, they proposed a SOC:clay ratio > 1:8 as "very good" and yielding field optimal structure, 1:10 < SOC:clay < 1:8 as "good", 1:13 < SOC:clay < 1:10 as structural "improvement suggested", and < 1:13 leading to poor structural conditions. Johannes et al. (2017) note that a SOC:clay ratio of 1:10 is a reasonable goal even under tillage. These studies used datasets from Poland and France (Dexter et al., 2008) and Switzerland (Johannes et al., 2017) with soil properties considerably different than the US data in our study; however, both highlight the importance of a SOC:clay ratio of 1:10. In addition, we tested the n value by plotting Spearman correlation coefficients between COC and specific volume for a range of n values and found no meaningful trend in the data; our analysis depended instead on a consistent partitioning of the potentially complexable mass (i.e., clay and SOC) into complexed and non-complexed pools (see Fig. A.1 and Section 2.2 Variable transformations and statistical analyses). As a result, in this study, we used a SOC:clay ratio of 1:10 in the calculation of CC and COC.

We combined CC and COC into a single variable representing the complexed fraction of clay and SOC, abbreviated "complexed organic carbon and clay" (COCC):

$$\mathrm{COCC} = \frac{\mathrm{CC} + \mathrm{COC}}{\mathrm{clay} + \mathrm{SOC}} \tag{3}$$

Low to high values of COCC represent an increasing fraction of SOC complexed with clay and increasingly porous soil structure (*sensu* Dexter et al., 2008; Czyż and Dexter, 2016). Values of COCC refer to the mass fraction of complexable material (i.e., clay and SOC) predicted to be in organo-mineral complexes.

Czyż and Dexter (2016) estimate that the effective density of organo-mineral complexes is very low with a mean value of 0.17 ± 0.04 g mL^{-1} in arable soils suggesting that these complexes are extremely porous with open structures perhaps in the form of fibers or chains. By considering organo-mineral complexes as a separate phase in soils, they can account for observed reductions in bulk density with increasing SOM contents (Dexter et al., 2008; Czyż and Dexter, 2016). Thus, we consider COCC (i.e., the complexed fraction of clay and SOC) to incorporate both solid phase and pore space components of soil structure.

Values of WSA were estimated using an artificial neural network pedotransfer function (Rivera and Bonilla, 2020). TP for each horizon was derived using the dry bulk density measured at a water content corresponding to a matric potential of −33 kPa. EP was calculated as the difference between TP and FC and considered to be a proxy of macroporosity representing the volume fraction of the largest pores in the soil (Rawls et al., 1998; Hirmas et al., 2018). For reference, see Table 1 for frequently used terms and variables, their abbreviations, and brief descriptions.

2.2. Variable transformations and statistical analyses

Prior to statistical analyses, distributions of all of the variables were visually inspected and transformed to reduce skewness and approximate a normal distribution as needed. Given the compositional and bounded nature of the data used in this study, we transformed and standardized variables as shown in Table 2. A centered log ratio (clr) transformation was applied to compositional data of mixtures with two or more components as follows:

$$\mathrm{clr}(x_{ij}) = \ln \frac{x_{ij}}{GM(x_i)} \tag{4}$$

where x_{ij} refers to the value of the jth component of the ith sample and GM is the geometric mean taken across all components of the ith sample.

In order to reduce zero values in the data, several variables were amalgamated including sand and coarse fraction as well as CC and COC (COCC; Eq. (3)). We used a modified Aitchison procedure (Pawlowsky-Glahn and Egozcue, 2006) to replace the remaining zero values to prepare the compositional variables for a centered log ratio (clr) transformation. Conceptually, COCC can be thought of as the complexed fraction of the complexable mass (i.e., clay and SOC) in the soil. Statistical values calculated using transformed variables were back transformed to ease interpretation of results. For compositional data involving mixtures of two components, we used the following inverse of the clr transformation:

$$\mathrm{clr}^{-1}(\hat{x}_t) = \frac{\exp(2\hat{x}_t)}{1 + \exp(2\hat{x}_t)} \tag{5}$$

where $\hat{x}_t$ is the statistic to be back transformed calculated in clr-transformed space. For mixtures with more than two components, the average geometric mean of each sample was used in equation Eq. (4) to estimate the value of the statistic on the original scale of each component. The final variables selected as predictors of EP were sand and coarse fraction (SCF), clay, SOM, COCC, WSA, MAP, and MAT.

The selected predictor variables were analyzed with decision trees (DTs) to assess their relative importance in explaining differences in EP both within individual ecoregions and across the full spatial domain represented by the samples in this study. Transformed variables were used in the DTs similar to other soil studies involving machine learning methods using compositional data (e.g., Zhang et al., 2020). Prior to the DT analyses, we detected and removed outliers using an adjusted Mahalanobis distance (Korkmaz et al., 2014) across each of the five

Table 1
Frequently used terms and variables, their abbreviations, and brief descriptions. For more details (e.g., text, equations, citations) see the entire Methods section. In particular, see Omernik and Griffith (2014) for ecoregion information, Dexter et al. (2008) and Czyż and Dexter (2016) as well as Eqs. (1), (2), and (3) for COCC information, Rivera and Bonilla (2020) for WSA information, and Rawls et al. (1998) for EP information.

Terms and variables	Abbreviations	Descriptions
A horizons	A	Surface layers occurring within the upper 25 cm of the land surface that were described as having morphological properties associated with topsoil
B horizons	B	Subsurface layers occurring within 25–100 cm of the land surface that were described as having morphological properties associated with subsoil
Ap horizons	Ap	Surface layers occurring within the upper 25 cm of the land surface exhibiting evidence of plowing
Bp horizons	Bp	Subsurface layers occurring within 25–100 cm of the land surface in a profile containing a shallower Ap horizon
Eastern Temperate Forests	ETF	
Great Plains	GP	
Mediterranean California	MC	
North American Deserts	NAD	
Northwestern Forested Mountains	NWFM	
Sand and coarse fraction	SCF	Sand (< 2 mm) and coarse fraction (2-76 mm) converted to a whole soil basis and amalgamated
Clay	Clay	Clay (< 2 mm) converted to a whole soil basis
Soil organic matter	SOM	Converted from soil organic carbon (SOC) using the standard SOM:SOC mass ratio of 1.72 and converted to a whole soil basis
Complexed organic carbon and clay	COCC	Low to high values of COCC represent an increasing fraction of SOC complexed with clay and increasingly porous soil structure
Water-stable aggregates	WSA	Mass fraction of water-stable aggregates
Effective porosity	EP	Calculated as the difference between total porosity and field capacity and considered to be a proxy of macroporosity
Mean annual precipitation	MAP	
Mean annual temperature	MAT	

Table 2
Information on the transformation of variables included in this study. Prior to transformation, CC and COC were amalgamated to form a new single complexed clay and SOC variable (COCC; Eq. (3)). This amalgamation was also conducted for sand and coarse fraction (SCF). In addition, zero values were replaced using a modified Aitchison procedure following Pawlowsky-Glahn and Egozcue (2006) before applying the centered log ratio (clr) transformation as indicated below. All of the variables below were standardized prior to analyses.

Variable	Reason for transform	Transform
SCF, clay, SOM, COCC	Compositional data	Modified Aitchison + clr
WSA, EP	Compositional and bounded data	Logit
MAP, MAT	Bounded data	Log

ecoregions and each of the four respective horizon types. Variance inflation factors (VIFs) were used to assess and remove multicollinearity between variables; a VIF ≥ 10 was used as an indicator of collinearity between predictor variables as recommended by Logan (2010). We added two categorical variables – horizon type and land use – with the levels A/B and plowed/non-plowed, respectively. We incorporated both categorical variables and continuous variables in the DT analyses. The rpart and rpart.plot R packages (Therneau et al., 2019; Milborrow, 2021) were used to generate DTs and trees were pruned using the default complexity parameter (i.e., 0.01). This parameter represents the overall threshold by which R^2 must increase at each step/node of the splitting process to prevent overfitting (Therneau et al., 2019; Milborrow, 2021). The complexity parameter at each step/node of the DT represents a minimization of the standard deviation of the errors calculated from cross-validation predictions generated from a set of cost-complexity prunings (Therneau et al., 2019). Thus, the variable occurring at the first split is the most important and holds more explanatory power (higher proportion of the R^2) and each subsequent split a lower proportion of the R^2 until the DT is pruned and stops splitting into steps/nodes. We used piecewise linear regression in the segmented R package (Muggeo, 2003, 2008) to understand the relationship between EP and COCC.

3. Results

3.1. Continental-scale

A total of 1491 pedons fit our criteria across the five ecoregions and were used in the DTs (Fig. 1). The number of horizons with complete cases for all variables included in the DT analyses (i.e., EP, SCF, clay, SOM, COCC, WSA, MAP, and MAT) ranged from 220 to 1314 across the ecoregions in this study (Table 3). These locations captured a wide range in MAP and MAT, spanning 310–1114 mm and 6.8–15.8 °C, respectively (Table 3).

EP was predicted by clay, COCC, ecoregion, and land use at a continental scale across all of the sites in this study (Fig. 2; $R^2 = 0.484$). Clay occurred at the first, second, and third split levels all of which showed a negative relationship with EP. For the bulk soil properties of the first DT split, higher clay (left branch) was represented by an elevated mean clay content (36.1%) and a lower mean SOC content (0.661%) across horizons, while lower clay (right branch) was represented by a lower mean clay content (17.0%) and an elevated mean SOC content (1.21%) across horizons. COCC was the second factor that emerged in predicting EP, and exhibited a strong positive relationship with EP across horizons that contained a higher clay content. To better understand this

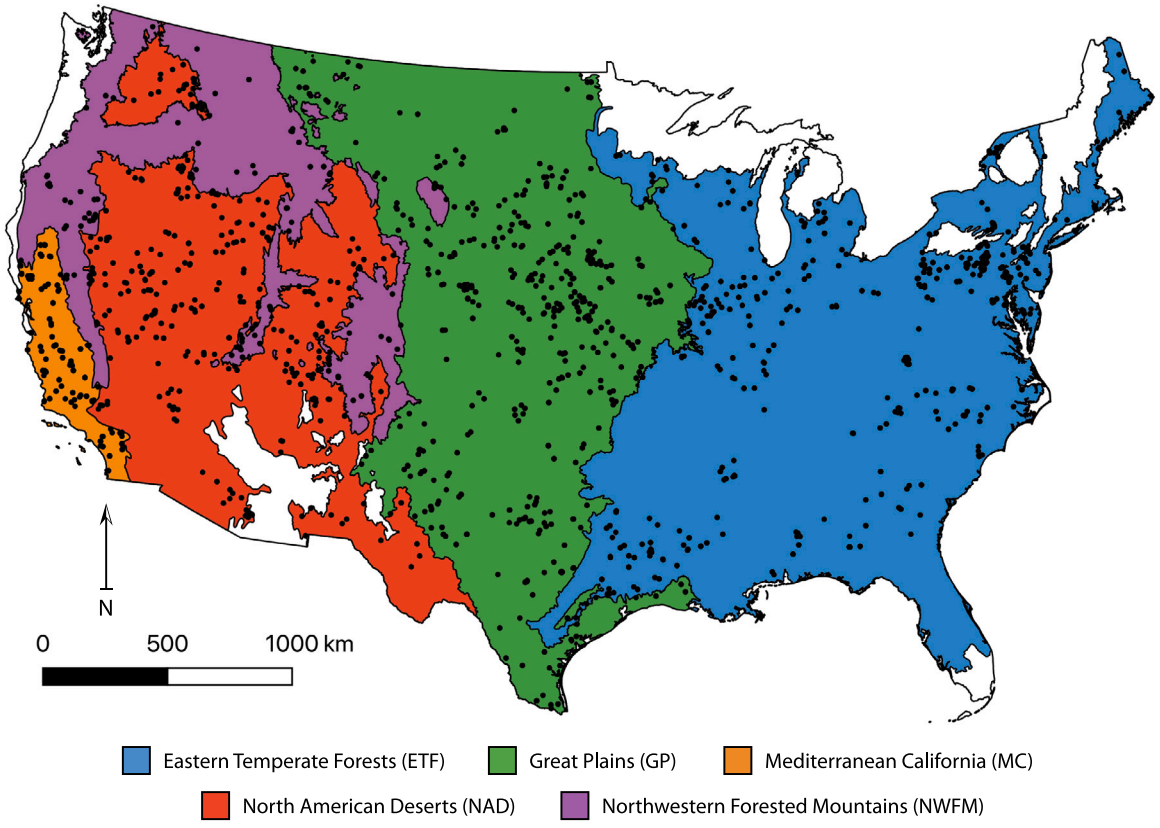


Fig. 1. Spatial distribution of pedon locations used in DT analyses across the Eastern Temperate Forests (ETF; blue), Great Plains (GP; green), Mediterranean California (MC; orange), North American Deserts (NAD; red), and Northwestern Forested Mountains (NWFM; purple) ecoregions of the conterminous US. In this study, each ecoregion represented areas under similar bioclimates or with similar potential vegetation.

Table 3
Number of pedons and horizons (arising from the pedons) used in the DT analyses across the Eastern Temperate Forests (ETF), Great Plains (GP), Mediterranean California (MC), North American Deserts (NAD), and Northwestern Forested Mountains (NWFM) ecoregions. Each ecoregion represented areas characterized by similar bioclimatic conditions or with similar potential vegetation. Also included below are MAP and MAT ± one standard deviation for the pedon locations across each of the ecoregions.

Ecoregion	Number of pedons	Number of horizons	MAP (mm)	MAT (°C)
ETF	479	1314	1114 ± 167	12.2 ± 3.82
GP	503	1259	661 ± 217	11.5 ± 4.26
MC	92	220	490 ± 278	15.8 ± 2.21
NAD	304	596	310 ± 125	10.4 ± 3.98
NWFM	113	290	638 ± 299	6.84 ± 3.05

relationship, we directly regressed EP and COCC (Fig. 3a, b) using a piecewise linear regression and estimated a breakpoint value of −0.250 corresponding to 37.8% of the total clay and SOC being complexed.

The breakpoint value of 37.8% COCC in Fig. 3a and b generally corresponds to the clay and SOC differences observed in the first split of the DT in Fig. 2. Fig. 4 shows that this value also corresponds to differences between A and B horizons with the former having lower clay content and higher SOC and the latter enriched in clay and lower in SOC. EP was predicted by horizon type first, and then by SCF, MAP, SOM, and land use across all sites in this study (R^2 = 0.379; Fig. 4). With clay removed, both A and B horizons split by SCF in the DT with values of SCF positively correlated to EP (Fig. 4).

A horizon EP was predicted by clay, MAP, land use, SCF, SOM, and ecoregion (Fig. 5a; R^2 = 0.397). Similarly, EP in B horizons was predicted by clay, MAP, SCF, and ecoregion; however, COCC was an important predictor while land use was not (Fig. 5b; R^2 = 0.449). At the first split level, both A and B horizons split by clay with higher clay resulting in decreased EP and lower clay resulting in increased EP (Fig. 5a and b). At the second split level, MAP and land use were important but indirect predictors of EP across A horizons (Fig. 5a). However, a more direct influence of COCC on EP was observed for B horizons (Fig. 5b).

3.2. Ecoregions and land use

Ecoregion and land use occurred at the third split level of the continental-scale DT (right branch; Fig. 2). ETF, MC, and NWFM ecoregions corresponded to decreased EP, whereas the GP and NAD ecoregions corresponded to increased EP (following a second split level of higher clay). Plowed horizons showed lower EP compared to non-plowed horizons (following a second split level of lower clay). We

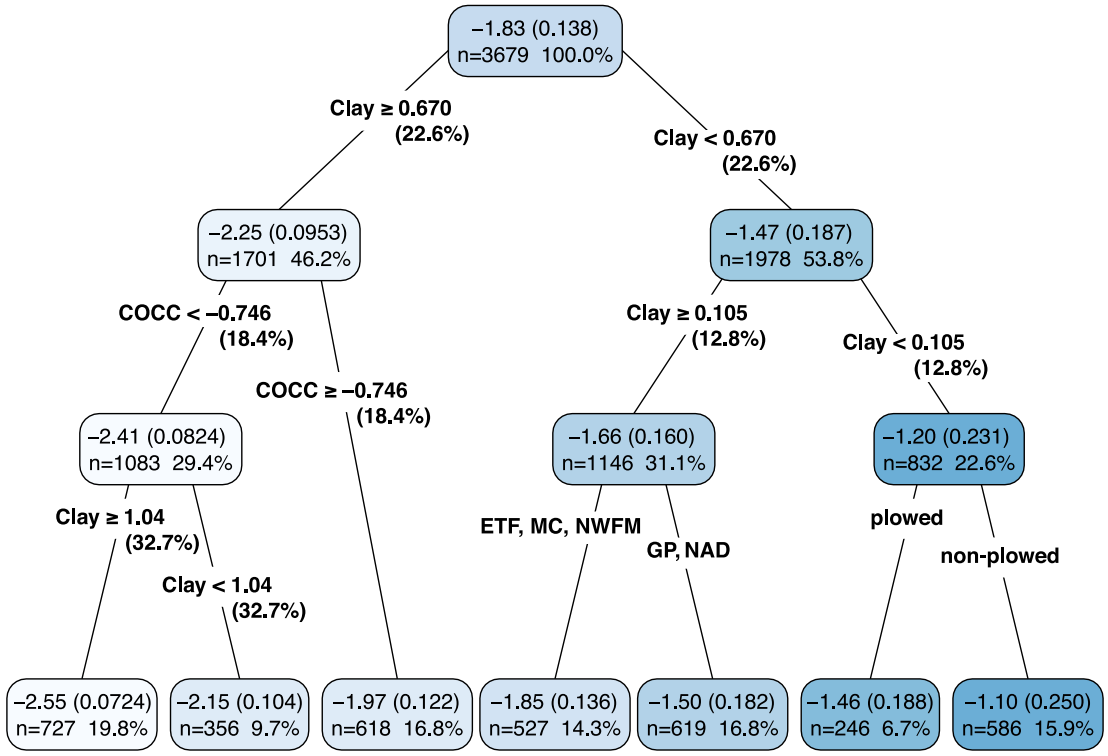


Fig. 2. Pruned DT displaying predicted EP at a continental scale across the conterminous US. All categorical variables (i.e., ecoregion, horizon type, and land use) and continuous (transformed) variables (i.e., SCF, clay, SOM, COCC, WSA, MAP, and MAT) were included as predictors. The first values displayed in the blue-shaded nodes correspond to the mean EP (mean transformed EP; back transformed values in parentheses) of the data subset, whereas the second and third values refer to the number (n) of horizons and the percentage of all horizons that fall in each subset, respectively. These three values are determined by the criteria specified in the previous splits with lighter shades of blue indicating lower mean EP and darker shades of blue indicating higher mean EP. Back transformed values for continuous variables are in parentheses.

found a higher breakpoint value separating the response of EP to COCC in A and B horizons (39.4%) than the breakpoint value separating Ap and Bp horizons (33.3%) (Fig. 3c and d, respectively).

Land use appears to be driving changes in EP and COCC to varying degrees across ecoregions. In general, both EP and COCC decreased from A to Ap and B to Bp horizons across ecoregions. This trend was more pronounced from A to Ap horizons as exhibited by differences in mean EP for the ETF, GP, MC, NAD, and NWFM (Fig. 3c and d). Likewise, differences in mean COCC also exhibited this trend across surface horizons of the ETF, GP, MC, and NWFM although EP values in A and Ap horizons of NAD were more comparable to each other [i.e., A: 0.542 (74.7%); Ap: 0.545 (74.8%)] (Fig. 3c and d). This trend between non-plowed and plowed horizons was less pronounced in B and Bp horizons as exhibited by differences in mean EP for the GP and NWFM with the exception of NAD values. However, B and Bp horizon EP values in ETF did not show considerable differences while EP values in MC increased from B to Bp horizons (Fig. 3c and d). Differences in mean COCC showed a slight decreasing trend from B to Bp horizons of the ETF, GP, MC, NAD, and NWFM (Fig. 3c and d). We note that smaller sample sizes for MC Ap and Bp horizons (n = 25 and 37, respectively) as well as NWFM Ap and Bp horizons (n = 11 and 14, respectively) should be taken into account when interpreting these results. In general, Ap and Bp centroids were grouped more closely compared to A and B centroids (Fig. 3c and d).

The first split between right and left branches of all ecoregion DTs largely represented A and B horizons, respectively (Fig. 6). This ecoregion behavior was similar to the behavior observed in the DTs using data across the conterminous US (Figs. 2 and 4). Overall, the ability of the DTs to predict EP across the ecoregions varied from an $R^2 = 0.399$ (ETF) to $R^2 = 0.771$ (MC) (Fig. 6; Table 4).

Table 4
Relationships of variables with EP (+ and − signs indicate the direction of the relationship) and R^2 values for the ETF, GP, MC, NAD, and NWFM ecoregion DTs displayed in Fig. 6a, b, c, d, and e, respectively.

Ecoregion	Variable	Relationship with EP	R^2
ETF	Clay	−	0.399
	WSA	+	
	SCF	+	
GP	Clay	−	0.615
	COCC	+	
	MAT	−	
	MAP	−	
	SCF	+	
	Plowing	−	
MC	Clay	−	0.771
	MAP	−	
	COCC	+	
	MAT	−	
NAD	Clay	−	0.501
	COCC	+	
	MAP	−	
	SCF	+	
	Plowing	−	
NWFM	COCC	+	0.712
	WSA	+	
	SOM	−	
	SCF	+	

Although there is a significant relationship between EP and COCC across the conterminous US, differences in what variables control this relationship emerge when comparing ecoregions. At the first split level, all of the ecoregions, except NWFM, split on clay content with higher

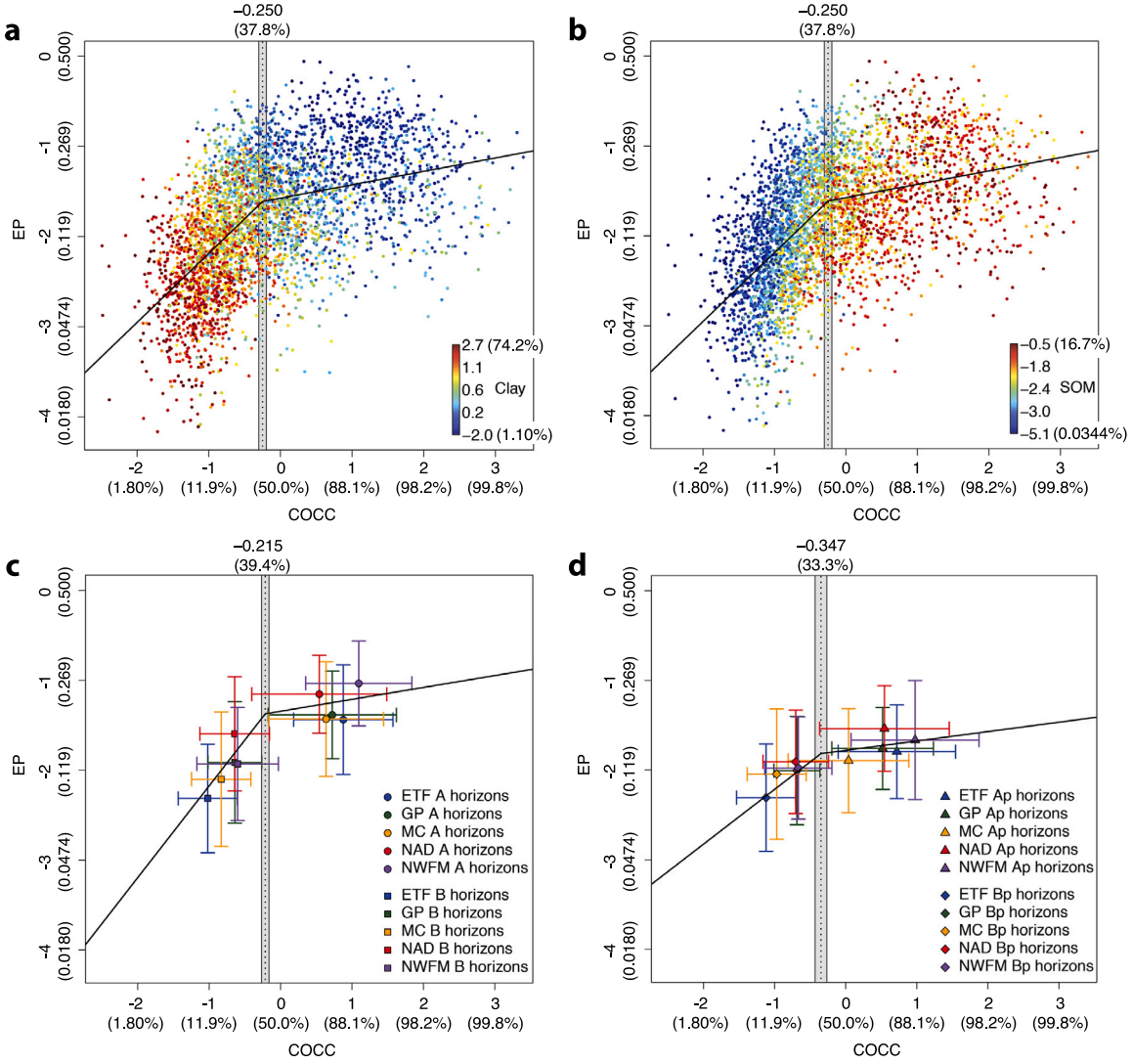


Fig. 3. (a) and (b) Piecewise linear regression of EP against COCC colored by clay and SOM, respectively, across all ecoregions and horizon types at a continental scale ($R^2 = 0.339$). The shaded bar in each panel represents one standard error distance around the breakpoint value. Values provided above this shaded bar at the top of each panel in parentheses are the back transformed values for COCC. Centroids and error bars (± one standard deviation) are shown in (c) for non-plowed A (circles) and B horizons (squares) and in (d) for plowed Ap (triangles) and Bp horizons (diamonds) across the ETF (blue), GP (green), MC (orange), NAD (red), and NWFM (purple) ecoregions. All of the variables in this figure were transformed as shown in Table 2. Back transformed values for EP and COCC and raw (untransformed) values for the clay and SOM color ramps are in parentheses.

and lower values observed in the left and right branches largely corresponding to B and A horizons, respectively (Fig. 6a, b, c, and d). COCC occurred at the first split for NWFM where lower values on the left branch were associated with B horizons, whereas higher values on the right branch corresponded to A horizons (Fig. 6e). WSA emerged as the next most important predictor of EP for the NWFM (right branch; Fig. 6e) and also occurred at the second split level of the ETF DT (right branch; Fig. 6a). Where MAP was more limited or where steeper climatic gradients occurred (i.e., GP, MC, NAD; Table 3), climate (i.e., MAP, MAT) tended to have an important and negative relationship with EP (Fig. 6b, c, and d; Table 4). Additionally, COCC occurred at the second split level of the GP DT (left branch; Fig. 6b) similar to where it occurred in the DTs shown in Figs. 2 and 5b derived using the full spatial extent of the US data.

4. Discussion

4.1. Continental-scale relationships

Given that soil texture is known to be a primary control of EP (Rawls et al., 1982, 1998; Nemes et al., 2005; Hirmas et al., 2018), it is not surprising that clay occurred at the first split level and showed a negative relationship with EP in the continental-scale DT (Fig. 2). Lessivage is a ubiquitous process and likely to be responsible for this clay separation between the two branches as fines translocate from surface A horizons and accumulate within subsurface B horizons (Buol et al., 2011; Turk et al., 2012; Koop et al., 2020). The differences in SOC content between the two branches also suggest this first split is governed by horizon development as A horizons accumulate more humified organic matter compared to B horizons (Buol et al., 2011; Turk et al., 2012). Thus, these processes of soil development appear to be important in controlling EP at a broad scale. Furthermore, these processes of soil development create soil structural conditions characterized by an integration of both the spatial distribution of mineral and organic compounds and the related spatial configuration of pore space (*sensu* Vogel et al., 2022). Thus, sorption of organic compounds on mineral surfaces is key to explaining the stability of soil as a heterogeneous porous matrix (Vogel et al., 2022).

4.1.1. Relationships between EP and COCC

As previously mentioned, COCC exhibited a strong positive relationship with EP across horizons containing a higher clay content.

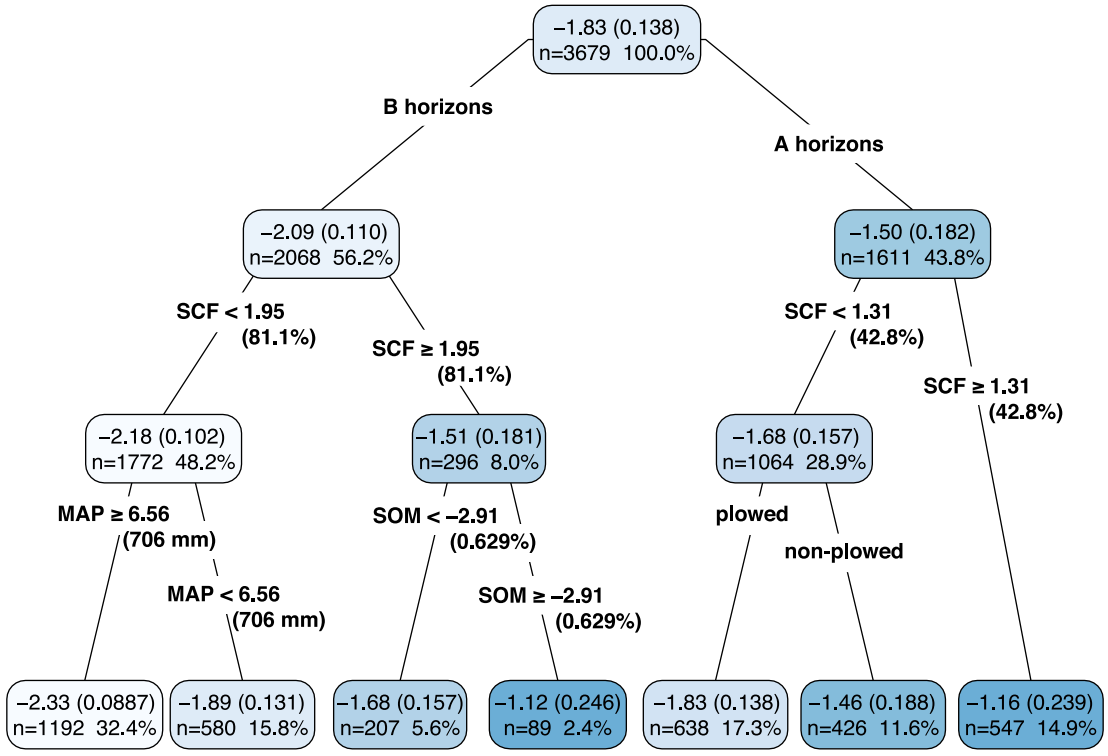


Fig. 4. Pruned DT showing predicted EP (mean transformed EP) at a continental scale across the conterminous US. All categorical variables (i.e., ecoregion, horizon type, and land use) and continuous (transformed) variables except clay and COCC (i.e., SCF, SOM, WSA, MAP, and MAT) were included as predictors. Back transformed values for continuous variables are in parentheses. Refer to Fig. 2 caption for general guidance on DT interpretation.

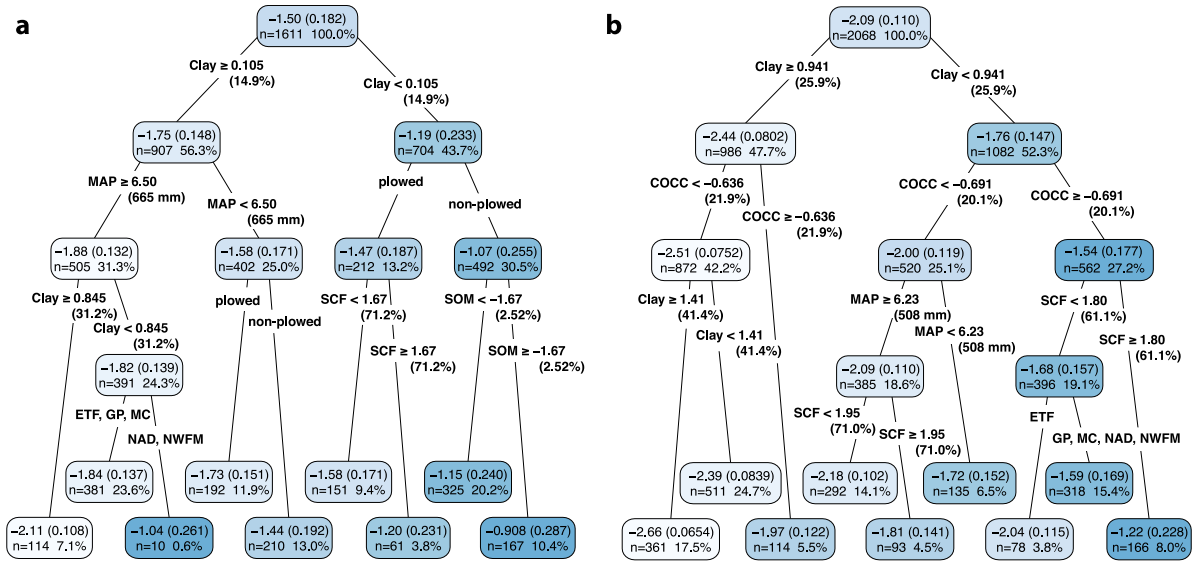


Fig. 5. Pruned DTs displaying predicted EP (mean transformed EP) at a continental scale across the conterminous US for (a) A horizons and (b) B horizons. For both horizon types, categorical variables (i.e., ecoregion and land use) and continuous (transformed) variables (i.e., SCF, clay, SOM, COCC, WSA, MAP, and MAT) were included as predictors. Back transformed values for continuous variables are in parentheses. Refer to Fig. 2 caption for general guidance on DT interpretation.

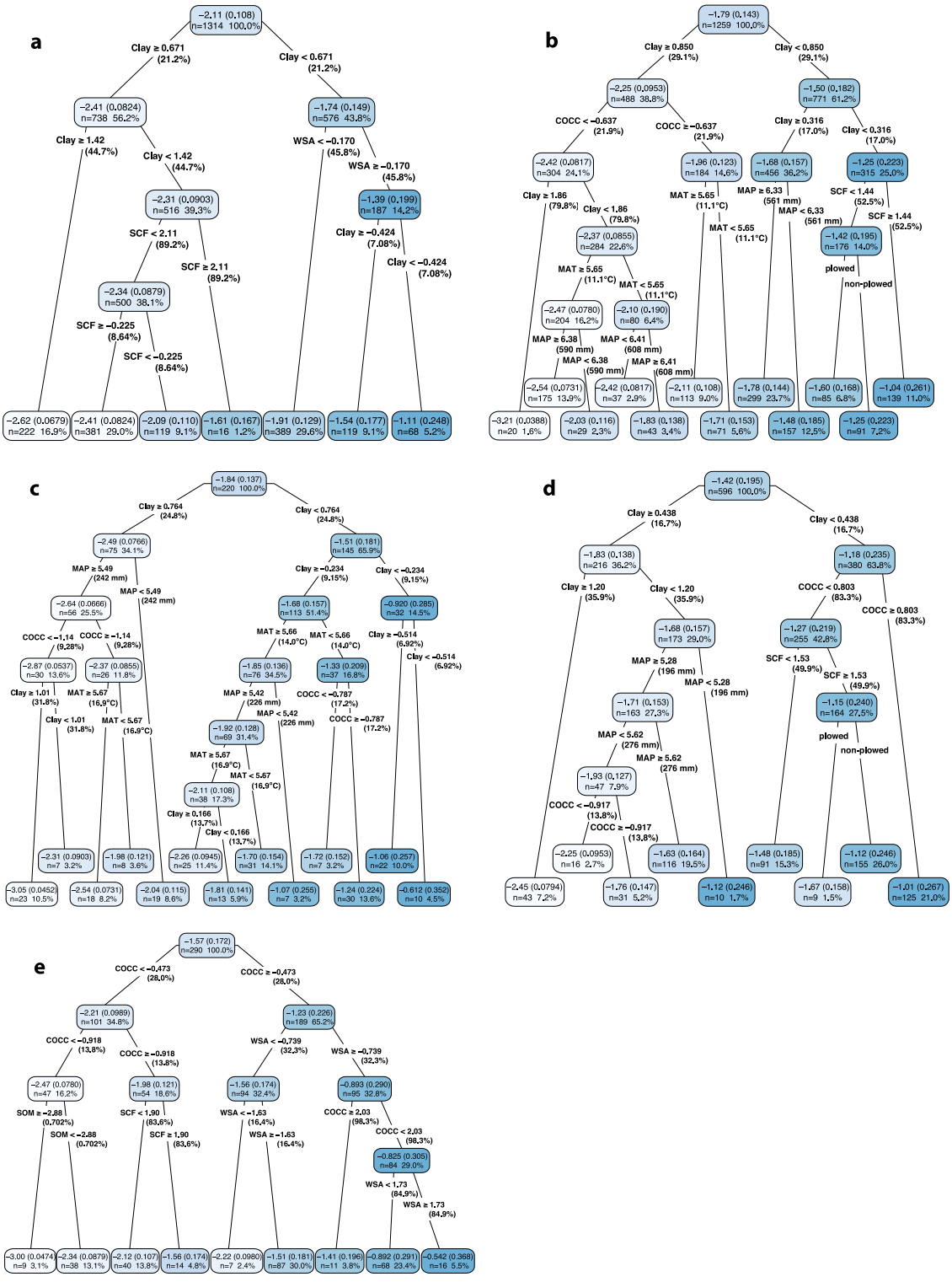


Fig. 6. Pruned DTs showing predicted EP (mean transformed EP) for (a) ETF, (b) GP, (c) MC, (d) NAD, and (e) NWFM ecoregions of the conterminous US. For each ecoregion, categorical variables (i.e., horizon type and land use) and continuous (transformed) variables (i.e., SCF, clay, SOM, COCC, WSA, MAP, and MAT) were included as predictors. Back transformed values for continuous variables are in parentheses. Refer to Fig. 2 caption for general guidance on DT interpretation.

Specifically, EP was more sensitive to changes in COCC below the breakpoint (higher clay and lower SOC content), and less sensitive above the breakpoint (Fig. 3a and b). Below the breakpoint, any additional SOC and/or reduction in clay (such that a greater fraction of the clay is complexed) will cause a steep increase in EP. However, above the breakpoint, when all of the clay is already complexed, any additional organic matter does not cause the same increase in EP. Several studies have shown that organo-mineral complexes (i.e., chemically protected organic matter bound to soil minerals) have a propensity to saturate with SOC likely due to the finite amount of mineral specific surface area onto which SOC can be stabilized (Hassink, 1996; Chung et al., 2008; Gulde et al., 2008; Stewart et al., 2008; Feng et al., 2014). Thus, EP appears to be controlled by the percentage of clay surface area that is complexed with SOC. When that surface area is not yet saturated,

EP can change more readily in response to SOC additions compared to when the surface area is saturated.

Samples below and above the COCC breakpoint in Fig. 3a and b have, in general, an effective mineral surface area (represented by the clay fraction) that is either partially or completely complexed with SOC, respectively (*sensu* Dexter et al., 2008). Here, we refer to partial complexation as "undersaturated" (i.e., values of COCC lower than the breakpoint) and complete complexation as "oversaturated" (i.e., COCC values above the breakpoint). The term "oversaturated" refers to the excess SOC not complexed with soil clay. Additional inputs of SOC in undersaturated conditions lead to increasing complexation and strengthening of organo-mineral bonds and concomitant development and/or maintenance of soil structure which increases EP (Dexter et al., 2008). However, in oversaturated conditions, the response of EP to increased inputs of SOC is much more muted. In addition, declines in SOC stabilization and storage efficiency as soils approach saturation may be linked to changes in the type, strength, or turnover time of organo-mineral interactions with increasing SOC inputs (Kleber et al., 2007; Sollins et al., 2009; Feng et al., 2014). For example, as soils approach SOC saturation, weaker organic–organic interactions become relatively more abundant than stronger organo-mineral interactions and it has been hypothesized that the stability and strength of SOC bound to soil minerals decreases with increasing SOC loadings (Kleber et al., 2007; Feng et al., 2014). Thus, while increasing complexation and a strengthening of organo-mineral bonds can occur with these inputs, it is accompanied by an abundance of non-complexed SOM that can have varying effects on macroporosity (including clogging of macropores). Ultimately, non-complexed SOM mutes the sensitivity of EP to changes in COCC and results in the asymptotical behavior of the EP-COCC relationship.

The EP-COCC relationship can also be viewed in the context of global mineral-associated organic carbon stocks and carbon-storage capacity where deeper soil layers and areas under agricultural management exhibit the largest undersaturation of mineral-associated carbon with sequestration efficiency over years to decades (Georgiou et al., 2022). Soils further from their mineralogical capacity are more effective at sequestering carbon (Georgiou et al., 2022). Thus, subsoils are often more suited to long-term carbon sequestration than topsoils (Button et al., 2022). The strength of the EP-COCC relationship is also likely related to mineral affinities for SOC (Slessarev et al., 2022).

4.1.2. Ecoregions and land use trends

The potential for increased clay-SOC complexation due to increased mineral surface area likely dictated the higher clay content branch splitting between ecoregions with lower EP (i.e., ETF, MC, and NWFM) and higher EP (i.e., GP and NAD; these can be seen on the right branch of the second and third split levels in Fig. 2). This ecoregion split could also be attributed to differences in how SOC inputs are stabilized onto mineral surfaces via various organo-mineral bonding reactions (e.g., ligand exchange, cation bridging, H-bonding, van der Waal forces) that depend on factors such as the composition of organic inputs, soil mineralogy, and environmental conditions (Arnarson and Keil, 2000; Gu et al., 1994; Stevenson, 1994; Feng et al., 2014). Decreased clay-SOC complexation potential due to decreased mineral surface area likely had an impact on the lower clay content branch splitting between plowed and non-plowed horizons (seen on the right branch of the second and third split levels in Fig. 2). In this case, the split between plowed horizons (decreased EP) and non-plowed horizons (increased EP) may have involved processes occurring in predominantly the non-complexed SOC pool. Observed losses of SOC induced by cultivation include reduced allocation of organic carbon to soils, reduced belowground allocation of photosynthate, enhanced aggregate disruption and exposure of physically-protected organic carbon, and enhanced rates of decomposition of available organic carbon substrates due to more favorable abiotic conditions (e.g., aeration, temperature, and water content) (Baldock and Broos, 2012). Therefore, disturbance effects such as plowing which reduce SOC in these horizons are likely responsible for the decreased EP. Due to the lower clay content across A horizons, more of the SOC present is not complexed with clay and is therefore subject to more rapid turnover processes associated with the labile non-complexed SOC pool (which is also a factor contributing to decreased EP). Finally, non-plowed horizons would allow for more SOC accumulation and increased EP. The lower clay content across these horizons also corresponds to processes occurring in the non-complexed SOC pool where more of the SOC present is not complexed with clay; however, the larger SOC content in these non-plowed horizons is likely contributing to the overall increase in EP compared to plowed horizons.

Although there was overlap of error bars associated with centroids represented in Fig. 3c and d, disturbance effects such as plowing generally contributed to decreases in both EP and COCC that were more pronounced in surface horizons and less pronounced but nonetheless propagated in subsurface horizons (with a few notable exceptions discussed in the following sections). These surface and subsurface horizon changes occurred despite mineral-bound SOC comprising a large majority of the total SOC and having longer turnover times (Balesdent et al., 1988; Christensen, 1998; Trumbore, 2000; Kahle et al., 2002; Feng et al., 2014). As previously mentioned, Ap and Bp centroids were generally grouped more closely compared to A and B centroids (Fig. 3c and d). This appeared to indicate that disturbance effects lead to homogenization of the EP-COCC relationship for Ap and Bp horizons including a weakening of this relationship for Bp horizons with relatively low COCC values when compared to undisturbed (or less disturbed) conditions across A and B horizons.

4.1.3. Impact of agricultural disturbance on the EP-COCC relationship

Comparing A and B centroids (Fig. 3c) with Ap and Bp centroids (Fig. 3d) provides some indication of disturbance whereby reduction in the magnitude and rate of change of EP as a function of COCC occurs. In general, undisturbed A and B horizons showed an increase in EP with increasing COCC. However, disturbed Ap horizons and B horizons under Ap show a reduction of this positive trend between EP and COCC. When comparing differences between ecoregion median values greater than the breakpoints, COCC and EP from undisturbed horizons were 16.4% and 0.0598 greater than disturbed values, respectively. For soils below the COCC breakpoints, undisturbed COCC and EP values were 1.98% and 0.0116 greater than disturbed values, respectively. Thus, these changes in EP appear to be driven by disturbance effects (e.g., tensile, shear, and/or compressive external forces caused by field operations) which likely impact the amount and stability of COCC with concomitant modifications of soil structure (Dexter, 1988; Ghezzehei, 2012; Horn and Peth, 2012).

Accompanying these general decreases in COCC and EP from undisturbed to disturbed settings were general decreases in WSA. For example, undisturbed WSA values (untransformed) for samples with median COCC values above the breakpoints were 13.4% greater than disturbed values whereas undisturbed soils had WSA values that were 2.85% greater than disturbed soils for median COCC values below the breakpoints. Since stress attenuation is greater for increasingly aggregated soils (under in situ conditions for soils with the same internal parameters) (Horn and Peth, 2012), tensile strength associated with COCC and WSA of undisturbed soils is likely greater and more effective at attenuating stress than tensile strength of COCC and WSA associated with disturbed soils. However, since a combination of shear and compressive stresses generally characterize in situ stress conditions during field operations (e.g., traffic, soil-tool interactions) (Horn and Peth, 2012), a number of other accompanying and confounding factors may impact the EP-COCC relationship. Undisturbed bulk density, ped size, and clay values were $0.0380\ g\ cm^{-3}$, 4.04 mm, and 5.56% less than disturbed values, respectively, for median COCC values greater than the breakpoints. For horizons with COCC less than the breakpoints, undisturbed bulk densities were $0.0690\ g\ cm^{-3}$ greater than disturbed

values, whereas ped size and clay values were 3.91 mm and 5.09% less than disturbed values, respectively (Fig. 3c and d).

Undisturbed surface horizon bulk densities were lower than disturbed soils indicating compression and densification in plowed horizons likely due to traffic during field operations. However, in B and Bp horizons, which were generally undersaturated with respect to COCC, undisturbed bulk densities were greater than those from disturbed horizons. While the reasons for this finding are unclear, this may be due to the effect of shear stresses in disturbed horizons where shearing behavior in aggregated soil depends on the water content and density of the aggregates as well as the geometry of the applied stress system (Dexter, 1988; Ghezzehei, 2012). If the density of the structured soil is above a characteristic critical value, shearing is confined to a well-defined surface or narrow band near the source of the stress where rolling of round particles or alignment of platy clay particles can occur and gross changes in volume take place in only a small proportion of the soil. In contrast, if the density of the structured soil is below the critical value then stress propagates through the whole soil volume resulting in an overall increase in density with disproportionate destruction of larger pores and consequent reductions in air-filled porosity and hydraulic conductivity (Kurtay and Reece, 1970; Hettiaratchi and O'Callaghan, 1980; Dexter, 1988; Ghezzehei, 2012). Therefore, undisturbed bulk densities that were greater than disturbed bulk densities may reflect conditions where density of the structured soil is above the critical value and soil-tool interactions cause shearing that is more likely to be confined near the source of the stress with gross changes in volume taking place in a small proportion of the soil.

Clay contents from undisturbed horizons were less than those in disturbed horizons suggesting that deep-loosening or deep-plowing and/or mixing with other soil material not only led to declines in soil stability and strength (Horn and Peth, 2012), but also brought illuvial clay upward and mixed it throughout the disturbed horizons. Peds were smaller for A and B horizons compared to Ap and Bp horizons likely because of the presence of clods in disturbed horizons. Clods are formed due to compaction by agricultural machinery where boundaries between macroaggregates may be lost or remain as microcracks depending on soil wetness and machinery loads during compaction; clods can also be formed during desiccation of large masses of clayey soil (Ghezzehei, 2012). Thus, clods and their formation processes are particularly important in subsoil horizons which typically do not have aggregates but exhibit similarly complex structure that, in contrast to topsoil (aggregate) structure, can be almost irreversibly damaged when subjected to subsoil compaction (Ghezzehei, 2012). In sum, due to external forces associated with field operations, Ap and Bp horizons not only undergo numerous disturbance effects (e.g., tensile, shear, and/or compressive stresses, clay saturation, subsoil compaction) but through these effects they also become increasingly homogenized and exhibit a reduction in the magnitude and rate of change of EP as a function of COCC (Fig. 3d) when compared to undisturbed A and B horizons (Fig. 3c). Thus, these results generally fall inline with other broad-scale studies suggesting improvement in soil structure and porosity under no-tillage practices especially in the long term (e.g., Mondal and Chakraborty, 2022).

4.2. Effects of A and B horizons on EP

As previously mentioned, lessivage and the accumulation of humified organic matter creates either under- or oversaturated conditions that strongly control EP. These conditions are expressed in subsurface and surface horizons, respectively. As shown in Fig. 4 with clay and COCC removed, EP was predicted by horizon type first, and then by SCF, MAP, SOM, and land use across all sites in this study confirming that undersaturated conditions largely correspond to B horizons (i.e., lower COCC and EP) and oversaturated conditions to A horizons (i.e., higher COCC and EP). With clay removed, both A and B horizons were split by SCF which was positively correlated to EP and where an increase in SCF in A horizons and decrease in B horizons likely indicated the translocation of fines from surface horizons into subsurface layers (Koop et al., 2020).

Although differences between A and B horizons emerged at the second split level in Fig. 5a and b, both horizon types were dependent on the fraction of clay complexed with SOC. Higher MAP in A horizons correspond with decreases in EP likely due to increased weathering and production of secondary minerals which corresponded to higher clay and, thus, a reduced fraction of that clay complexed with SOC (Fig. 5a). Under lower MAP, decreases in weathering and production of secondary minerals are likely responsible for lower clay contents making a greater proportion of that clay subject to complexation with SOC with, correspondingly, greater potential for aggregation and increases in EP. B horizons, however, more directly reflect the EP-COCC relationship at the second split level of the DT (Fig. 5b). Since A horizons are typically oversaturated with respect to COCC, we observed more indirect factors (i.e., MAP, land use) coming to the forefront as important predictors of EP, whereas since B horizons are typically undersaturated, we observe a more direct influence of COCC on EP. In general, clay content in these B horizons was not fully complexed with SOC and, therefore, EP showed a greater response to increased SOC in the subsoil.

4.3. Ecoregion influences on EP

As with the continental-scale DTs (Figs. 2 and 4), all of the ecoregion DTs first split into under- and oversaturated conditions with respect to the proportion of clay complexed with SOC, with the left and right branches largely representing B and A horizons, respectively (Fig. 6). However, EP within the NWFM ecoregion (Fig. 6e) appears to exhibit a propensity to respond to SOC additions for both surface and subsurface horizons. The similarity in the EP-COCC relationship between A and B horizons may be due to the increased occurrence of macroaggregation as represented by WSA emerging as the next strongest predictor of EP (Fig. 6e). WSA also occurred at the second split level of the ETF DT (right branch; oversaturated A horizons) (Fig. 6a). The increased aggregation in forested soils may be caused by both elevated additions of organic matter and increases in acidity. Rivera and Bonilla (2020) observed a positive gradient of macroaggregate stability from arid to humid environments due to increases in SOM and decreases in pH. At lower pH values (< 7), solubility and mobility of cations are higher resulting in the formation of bridges with clay and SOM; in addition, microbial activity under lower pH values increases and promotes macroaggregation and stabilization (Tisdall and Oades, 1982; Bronick and Lal, 2005; Regelink et al., 2015; Wu et al., 2017; Rivera and Bonilla, 2020). Thus, in these forested (or previously forested) environments with higher MAP (Table 3) and SOM and lower pH, WSA is an important predictor that is positively associated with EP (Table 4).

Where MAP was more limited or where steeper climatic gradients occurred (i.e., GP, MC, NAD; Table 3), MAP and MAT tended to be negatively correlated to EP (Fig. 6b, c, and d; Table 4). Across the conterminous US, SOC generally increases as MAP increases up to 700–850 mm and then SOC content fluctuates as MAP continues to increase (Guo et al., 2006). However, when considering grassland and forest ecosystems with MAP < 1000 mm, SOC generally decreases as MAT increases across elevations < 600 m and slopes < 1° (Guo et al., 2006). The negative correlation between SOC and MAT implies that the relative temperature sensitivity of decomposition is greater than that of net primary productivity although strong interactions between temperature, water availability, and substrate quantity make it difficult to assess the temperature dependence of decomposition without confounding effects (Baldock and Broos, 2012). In general, negative correlations between the climatic variables MAP and MAT and EP likely reflected the positive association between MAP and weathering and production of secondary minerals (and lessivage) (Buol et al., 2011; Schaetzl and Thompson, 2015) and the negative association between MAT and SOC (Guo et al., 2006) which strongly influences the fraction

of clay complexed with SOC. However, other than a second split level in the MC ecoregion where higher MAP likely corresponded to increased illuvial clay accumulation (and, thus, lower EP) and lower MAP corresponded to decreased clay illuviation resulting in higher EP (Fig. 6c), MAP and MAT occurred at lower split levels with less explanatory power. This suggests that climate influences the EP-COCC relationship more indirectly and likely in a complex manner. Additionally, COCC occurred at the second split level of the GP DT (left branch in Fig. 6b; undersaturated B horizons) which was similar to where it occurred in the continental-scale DTs in Figs. 2 and 5b suggesting that the GP ecoregion with its sizable area and steep climatic gradients (north–south MAT; west–east MAP) captured what was occurring in subsurface (B) horizons across the conterminous US. That is, EP exhibits a propensity to respond to SOC additions and/or reductions in clay increasing the proportion of complexed clay. EP responding to both SOC additions and reductions in clay was perhaps best exemplified in COCC occurring at the first split level of the NWFM DT (right branch) and the second split level of the NAD DT (right branch). The steeper increase in EP of surface (A) horizons was likely linked to coarser textures (and SOC additions) that increased the fraction of complexed clay in these ecoregions of the western US (Fig. 6d and e; Table 4).

5. Conclusion

This work highlights that EP (a proxy of macroporosity) of surface (A) and subsurface (B) horizons is strongly dependent on the fraction of clay complexed with SOC (as represented by COCC). In this study, we use the relationship between EP and COCC to distinguish between undersaturated (i.e., low values of COCC; largely B horizons) and oversaturated (i.e., high COCC values; largely A horizons) conditions; these conditions refer to the amount of SOC available to complex with clay. In undersaturated conditions, steeper increases in EP are associated with larger amounts of SOC and/or reductions in clay that result in a greater fraction of complexed clay leading to stronger organo-mineral bonds and the concomitant development or maintenance of soil structure. In oversaturated conditions, the slope describing the EP-COCC relationship was positive but reduced compared to undersaturated conditions. This reduction was likely due to all or most of the clay being already effectively complexed and the varying effects that increasing accumulations of non-complexed SOM can have on macroporosity. Ultimately, these oversaturated conditions mute the sensitivity of EP to changes in COCC and result in the observed asymptotical behavior of the EP-COCC relationship.

Since A horizons typically exhibit oversaturated conditions with higher SOC and lower clay contents, indirect factors (e.g., MAP, land use) are important predictors of EP in surface horizons. However, since B horizons are typically associated with undersaturated conditions (i.e., higher clay content and lower SOC), direct influences from COCC on EP occurs in the subsoil. We found that the EP-COCC relationship was also important for explaining changes in EP within ecoregions but its effect is mitigated by soil and climate interactions. In forested ecoregions (i.e., ETF, NWFM) with higher MAP and SOM and lower pH, increased macroaggregation and stability of these macroaggregates are positively associated with EP. Where MAP is more limited or where steeper climatic gradients occur (i.e., GP, MC, NAD), MAP and MAT tend to be negatively correlated with EP likely due to the positive association between MAP and the production or translocation of secondary minerals through weathering and lessivage and the negative association between MAT and SOC. Thus, climate appears to influence the EP-COCC relationship in an indirect and complex manner.

We found that EP exhibits strong tendencies to positively respond to SOC additions and/or reductions in clay both of which increase the proportion of complexed clay. This may explain why arid climates (e.g., the western US), which tend to have low to moderate clay contents with low to moderate SOC contents, lead to increases in the fraction of complexed clay and concomitant increases in macroporosity and K_{sat}. For humid climates (e.g., the eastern US), which tend to have moderate to high clay contents with moderate to moderately high SOC contents, increases in the proportion of complexed clay and concomitant increases in macroporosity and K_{sat} can also occur. However, our findings suggest that under humid climates, even higher SOC contents will cause a corresponding increase in non-complexed SOM pools that can reduce the response of EP to COCC. This may be especially true for coarser textures in humid environments with high SOM where less clay corresponds to a reduced fraction of complexed clay leading to reductions in macroporosity and K_{sat}. Disturbance, as represented by the Ap and Bp horizons, creates conditions that are more homogenized compared to A and B horizons which reduces the magnitude and rate of change of EP as a function of COCC.

The findings of this work point to the importance of complexed clay and SOC in controlling macroporosity and K_{sat} across ecoregions and suggests that the EP-COCC relationship provides an important framework for understanding and predicting future land use- and climate-induced changes to soil hydraulic properties. This study also highlights the potential for using broad geographic soil datasets to examine relevant drivers of soil property changes. This potential stems from the large number of point samples contained in these datasets that increase the detection sensitivity of signals in the soil above the high-frequency, local variability inherent in many soil properties. In addition, broad geographic datasets allow for exploration across a fuller range of soil property values than local-scale studies and provide an assessment of questions at scales inline with global and regional-scale Earth system models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank the Department of Geography and Atmospheric Science at the University of Kansas for support of this project. Thanks goes to Melanie R. Koop for assistance with creating figures. We also appreciate insights shared by Katie Murenbeeld during research group meetings. We also thank the two anonymous reviewers of this work for their helpful feedback which improved the manuscript. This work is supported by Hatch funds (no. CA-R-ENS-5195-H, project accession no. 1022418) and a Signals in the Soil grant (DRH - no. 2021-67019-34341; SAB - no. 2021-67019-34338; ANF - no. 2021-67019-34340) from the USDA National Institute of Food and Agriculture and by the National Science Foundation under Grant No. 2026874, 2034214 (LL), 2034232 (PLS), 2121760 (DRH/HA), 2121639 (SAB), 2121621 (LL), 2121595 (ANF), and 2121694 (PLS). We acknowledge funding from the CLIMASOIL project of The Research Council of Norway (Project number: 325253). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix

See Fig. A.1.

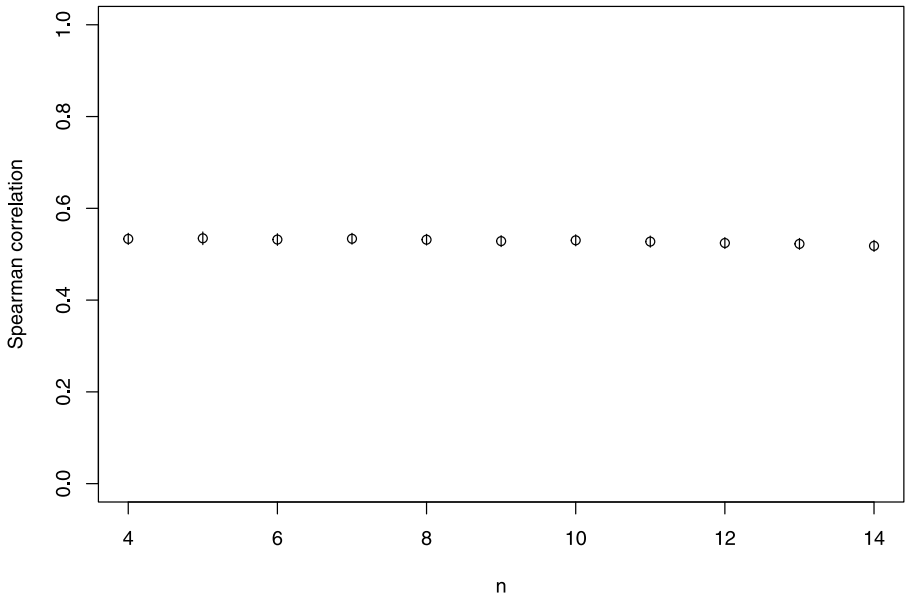


Fig. A.1. Similar to Figure 2 in Dexter et al. (2008), the plot above shows Spearman correlations between complexed organic carbon (COC) and specific volume (1/bulk density) for n values ranging from 4 to 14 for our dataset. Each circle represents the correlation coefficient and each line plus and minus one standard deviation is significant at $p < 0.001$. As shown above, there is no meaningful trend in these correlations across the range of n values tested and we interpret this to mean that our analysis using a SOC:clay ratio of 1:10 ($n = 10$) following Dexter et al. (2008) is not particularly sensitive to this value. Instead, analysis in this study depended on a consistent partitioning of the potentially complexable mass (i.e., clay and SOC) into complexed and non-complexed pools.

References

Ahuja, L.R., Naney, J.W., Green, R.E., Nielsen, D.R., 1984. Macroporosity to characterize spatial variability of hydraulic conductivity and effects of land management. Soil Sci. Am. J. 48 (4), 699–702.

Araya, S.N., Ghezzehei, T.A., 2019. Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. Water Resour. Res. 55 (7), 5715–5737.

Arnarson, T.S., Keil, R.G., 2000. Mechanisms of pore water organic matter adsorption to montmorillonite. Mar. Chem. 71 (3–4), 309–320.

Baldock, J.A., Broos, K., 2012. Soil organic matter. In: Huang, P.M., Li, Y., Sumner, M.E. (Eds.), Handbook of Soil Sciences: Properties and Processes, second ed. CRC Press, Boca Raton, FL, pp. 11–1—11–52.

Balesdent, J., Wagner, G.H., Mariotti, A., 1988. Soil organic matter turnover in long-term field experiments as revealed by C-13 natural abundance. Soil Sci. Am. J. 52 (1), 118–124.

Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. Geoderma 124 (1–2), 3–22.

Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., 2011. Soil Genesis and Classification, 6th Wiley-Blackwell, Hoboken, NJ.

Button, E.S., Pett-Ridge, J., Murphy, D.V., Kuzyakov, Y., Chadwick, D.R., Jones, D.L., 2022. Deep-C storage: Biological, chemical and physical strategies to enhance carbon stocks in agricultural subsoils. Soil Biol. Biochem. 108697.

Caplan, J.S., Giménez, D., Hirmas, D.R., Brunsell, N.A., Blair, J.M., Knapp, A.K., 2019. Decadal-scale shifts in soil hydraulic properties as induced by altered precipitation. Sci. Adv. 5 (9), eaau6635.

Christensen, B.T., 1998. Carbon in primary and secondary organo-mineral complexes. In: Carter, M.R., Stewart, B.A. (Eds.), Structure and Organic Matter Storage in Agricultural Soils. CRC Press, Boca Raton, FL, pp. 97–165.

Chung, H., Grove, J.H., Six, J., 2008. Indications for soil carbon saturation in a temperate agroecosystem. Soil Sci. Am. J. 72 (4), 1132–1139.

Czyż, E.A., Dexter, A.R., 2016. Estimation of the density of the clay-organic complex in soil. Int. Agrophys. 30 (1), 19–23.

Dexter, A.R., 1988. Advances in characterization of soil structure. Soil Tillage Res. 11 (3–4), 199–238.

Dexter, A.R., Richard, G., Arrouays, D., Czyż, E.A., Jolivet, C., Duval, O., 2008. Complexed organic matter controls soil physical properties. Geoderma 144 (3–4), 620–627.

Feng, W., Plante, A.F., Aufdenkampe, A.K., Six, J., 2014. Soil organic matter stability in organo-mineral complexes as a function of increasing C loading. Soil Biol. Biochem. 69, 398–405.

Georgiou, K., Jackson, R.B., Vindušková, O., Abramoff, R.Z., Ahlström, A., Feng, W., Harden, J.W., Pellegrini, A.F., Polley, H.W., Soong, J.L., Riley, W.J., Torn, M.S., 2022. Global stocks and capacity of mineral-associated soil organic carbon. Nature Commun. 13 (1), 1–12.

Ghezzehei, T.A., 2012. Soil structre. In: Huang, P.M., Li, Y., Sumner, M.E. (Eds.), Handbook of Soil Sciences: Properties and Processes, second ed. CRC Press, Boca Raton, FL, pp. 2–1—2–17.

Giménez, D., Hirmas, D.R., 2017. Macroporosity. In: Lal, R. (Ed.), Encyclopedia of Soil Science, third ed. CRC Press, Boca Raton, FL, pp. 1388–1391.

Gu, B., Schmitt, J., Chen, Z., Liang, L., McCarthy, J.F., 1994. Adsorption and desorption of natural organic matter on iron oxide: Mechanisms and models. Environ. Sci. Technol. 28 (1), 38–46.

Gulde, S., Chung, H., Amelung, W., Chang, C., Six, J., 2008. Soil carbon saturation controls labile and stable carbon pool dynamics. Soil Sci. Am. J. 72 (3), 605–612.

Guo, Y., Gong, P., Amundson, R., Yu, Q., 2006. Analysis of factors controlling soil carbon in the conterminous United States. Soil Sci. Am. J. 70 (2), 601–612.

Hassink, J., 1996. Preservation of plant residues in soils differing in unsaturated protective capacity. Soil Sci. Am. J. 60 (2), 487–491.

Hettiaratchi, D.R.P., O'Callaghan, J.R., 1980. Mechanical behavior of agricultural soils. J. Agric. Eng. Res. 25 (3), 239–259.

Hirmas, D.R., Giménez, D., Nemes, A., Kerry, R., Brunsell, N.A., Wilson, C.J., 2018. Climate-induced changes in continental-scale soil macroporosity may intensify water cycle. Nature 561 (7721), 100–103.

Horn, R., Peth, S., 2012. Mechanics of unsaturated soils for agricultural applications. In: Huang, P.M., Li, Y., Sumner, M.E. (Eds.), Handbook of Soil Sciences: Properties and Processes, second ed. CRC Press, Boca Raton, FL, pp. 3–1—3–30.

Hudson, B.D., 1994. Soil organic matter and available water capacity. J. Soil Water Conserv. 49 (2), 189–194.

Jarvis, N., Koestel, J., Messing, I., Moeys, J., Lindahl, A., 2013. Influence of soil, land use and climatic factors on the hydraulic conductivity of soil. Hydrol. Earth Syst. Sci. 17 (12), 5185–5195.

Johannes, A., Matter, A., Schulin, R., Weisskopf, P., Baveye, P.C., Boivin, P., 2017. Optimal organic carbon values for soil structure quality of arable soils. Does clay content matter? Geoderma 302, 14–21.

Jorda, H., Bechtold, M., Jarvis, N., Koestel, J., 2015. Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity. Eur. J. Soil Sci. 66 (4), 744–756.

Kahle, M., Kleber, M., Jahn, R., 2002. Carbon storage in loess derived surface soils from central Germany: Influence of mineral phase variables. J. Plant Nutr. Soil Sci. 165 (2), 141–149.

Kleber, M., Sollins, P., Sutton, R., 2007. A conceptual model of organo-mineral interactions in soils: self-assembly of organic molecular fragments into zonal structures on mineral surfaces. Biogeochemistry 85 (1), 9–24.

Koop, A.N., Hirmas, D.R., Sullivan, P.L., Mohammed, A.K., 2020. A generalizable index of soil development. Geoderma 360, 113898.

Korkmaz, S., Goksuluk, D., Zararsiz, G., 2014. MVN: An R package for assessing multivariate normality. R J. 6 (2), 151–162.

Kurtay, T., Reece, A.R., 1970. Plasticity theory and critical state soil mechanics. J. Terramech. 7 (3–4), 23–56.

Li, L., Stewart, B., Zhi, W., Sadayappan, K., Ramesh, S., Kerins, D., Sterle, G., Harpold, A., Perdrial, J., 2022. Climate controls on river chemistry. Earth's Future 10 (6), e2021EF002603.

Logan, M., 2010. Biostatistical Design and Analysis using R: A Practical Guide, first ed. Wiley-Blackwell, Hoboken, NJ.

Milborrow, S., 2021. Rpart.plot: plot 'rpart' models: an enhanced version of 'plot.rpart'. R package version 3.1.0. URL http://www.milbo.org/rpart-plot/index.html.

Mondal, S., Chakraborty, D., 2022. Global meta-analysis suggests that no-tillage favourably changes soil structure and porosity. Geoderma 405, 115443.

Muggeo, V.M.R., 2003. Estimating regression models with unknown break-points. Stat. Med. 22 (19), 3055–3071.

Muggeo, V.M.R., 2008. Segmented: An R package to fit regression models with broken-line relationships. R News 8 (1), 20–25.

Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2005. Influence of organic matter on the estimation of saturated hydraulic conductivity. Soil Sci. Am. J. 69 (4), 1330–1337.

Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. Environ. Manag. 54 (6), 1249–1266.

Pawlowsky-Glahn, V., Egozcue, J.J., 2006. Compositional data and their analysis: an introduction. In: Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (Eds.), Compositional Data Analysis in the Geosciences: From Theory to Practice, vol. 264. The Geological Society of London, pp. 1–10,

PRISM Climate Group, Oregon State University, 2021. Parameter-elevation Regressions on Independent Slopes Model. URL http://prism.oregonstate.edu.

Rawls, W.J., Brakensiek, D.L., Saxton, K.E., 1982. Estimation of soil water properties. Trans. Am. Soc. Agric. Eng. 25 (5), 1316–1320.

Rawls, W.J., Giménez, D., Grossman, R., 1998. Use of soil texture, bulk density, and slope of the water retention curve to predict saturated hydraulic conductivity. Trans. ASAE 41 (4), 983–988.

Rawls, W.J., Nemes, A., Pachepsky, Y.A., 2004. Effect of soil organic carbon on soil hydraulic properties. Dev. Soil Sci 30, 95–114.

Regelink, I.C., Stoof, C.R., Rousseva, S., Weng, L., Lair, G.J., Kram, P., Nikolaidis, N.P., Kercheva, M., Banwart, S., Comans, R.N., 2015. Linkages between aggregate formation, porosity and soil chemical properties. Geoderma 247, 24–37.

Rivera, J.I., Bonilla, C.A., 2020. Predicting soil aggregate stability using readily available soil properties and machine learning techniques. Catena 187, 104408.

Saxton, K.E., Rawls, W.J., 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Sci. Am. J. 70 (5), 1569–1578.

Schaetzl, R.J., Thompson, M.L., 2015. Soils: Genesis and Geomorphology, second ed. Cambridge University Press, New York.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro) aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79 (1), 7–31.

Slessarev, E.W., Chadwick, O.A., Sokol, N.W., Nuccio, E.E., Pett-Ridge, J., 2022. Rock weathering controls the potential for soil carbon storage at a continental scale. Biogeochemistry 157 (1), 1–13.

Sollins, P., Kramer, M.G., Swanston, C., Lajtha, K., Filley, T., Aufdenkampe, A.K., Wagai, R., Bowden, R.D., 2009. Sequential density fractionation across soils of contrasting mineralogy: evidence for both microbial-and mineral-controlled soil organic matter stabilization. Biogeochemistry 96 (1), 209–231.

Stevenson, F.J., 1994. Humus Chemistry. Genesis, Composition, Reactions. John Wiley & Sons, New York, NY.

Stewart, C.E., Plante, A.F., Paustian, K., Conant, R.T., Six, J., 2008. Soil carbon saturation: Linking concept and measurable carbon pools. Soil Sci. Am. J. 72 (2), 379–392.

Sullivan, P.L., Billings, S.A., Hirmas, D.R., Li, L., Zhang, X., Ziegler, S., Murenbeeld, K., Ajami, H., Guthrie, A., Singha, K., Giménez, D., Duro, A., Moreno, V., Flores, A., Cueva, A., Koop, A.N., Aronson, E.L., Barnard, H.R., Banwart, S.A., Keen, R.M., Nemes, A., Nikolaidis, N.P., Nippert, J.B., Richter, D., Robinson, D.A., Sadayappan, K., de Souza, L.F.T., Unruh, M., Wen, H., 2022. Embracing the dynamic nature of soil structure: A paradigm illuminating the role of life in critical zones of the anthropocene. Earth-Sci. Rev. 225, 103873.

Therneau, T., Atkinson, B., Ripley, B., 2019. Rpart: recursive partitioning and regression trees. R package version 4.1-15. URL https://cran.r-project.org/package=rpart.

Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. J. Soil Sci. 33 (2), 141–163.

Trumbore, S., 2000. Age of soil organic matter and soil respiration: Radiocarbon constraints on belowground C dynamics. Ecol. Appl. 10 (2), 399–411.

Turk, J.K., Chadwick, O.A., Graham, R.C., 2012. Pedogenic Processes. In: Huang, P.M., Li, Y., Sumner, M.E. (Eds.), Handbook of Soil Sciences: Properties and Processes, second ed. CRC Press, Boca Raton, FL, pp. 30–1—30–29.

Vogel, H.-J., Balseiro-Romero, M., Kravchenko, A., Otten, W., Pot, V., Schlüter, S., Weller, U., Baveye, P.C., 2022. A holistic perspective on soil architecture is needed as a key to soil functions. Eur. J. Soil Sci. 73 (1), e13152.

Wang, Y., Shao, M., Liu, Z., Horton, R., 2013. Regional-scale variation and distribution patterns of soil saturated hydraulic conductivities in surface and subsurface layers in the loessial soils of China. J. Hydrol. 487, 13–23.

Wang, T., Wedin, D., Zlotnik, V.A., 2009. Field evidence of a negative correlation between saturated hydraulic conductivity and soil carbon in a sandy soil. Water Resour. Res. 45 (7).

Watson, K.W., Luxmoore, R.J., 1986. Estimating macroporosity in a forest watershed by use of a tension infiltrometer. Soil Sci. Am. J. 50 (3), 578–582.

Wen, H., Sullivan, P.L., Macpherson, G.L., Billings, S.A., Li, L., 2021. Deepening roots can enhance carbonate weathering by amplifying CO_2-rich recharge. Biogeosciences 18 (1), 55–75.

Wu, X., Wei, Y., Wang, J., Wang, D., She, L., Wang, J., Cai, C., 2017. Effects of soil physicochemical properties on aggregate stability along a weathering gradient. Catena 156, 205–215.

Zhang, M., Shi, W., Xu, Z., 2020. Systematic comparison of five machine-learning models in classification and interpolation of soil particle size fractions using different transformed data. Hydrol. Earth Syst. Sci. 24 (5), 2505–2526.