Global stocks and capacity of mineral-associated soil organic carbon

Katerina Georgiou1,2✉, Robert B. Jackson2,3,4, Olga Vindušková5,6, Rose Z. Abramoff7,8, Anders Ahlström9, Wenting Feng10, Jennifer W. Harden2,11, Adam F. A. Pellegrini12,13, H. Wayne Polley14, Jennifer L. Soong15,16, William J. Riley17 & Margaret S. Torn17,18

Soil is the largest terrestrial reservoir of organic carbon and is central for climate change mitigation and carbon-climate feedbacks. Chemical and physical associations of soil carbon with minerals play a critical role in carbon storage, but the amount and global capacity for storage in this form remain unquantified. Here, we produce spatially-resolved global estimates of mineral-associated organic carbon stocks and carbon-storage capacity by analyzing 1144 globally-distributed soil profiles. We show that current stocks total 899 Pg C to a depth of 1 m in non-permafrost mineral soils. Although this constitutes 66% and 70% of soil carbon in surface and deeper layers, respectively, it is only 42% and 21% of the mineralogical capacity. Regions under agricultural management and deeper soil layers show the largest undersaturation of mineral-associated carbon. Critically, the degree of undersaturation indicates sequestration efficiency over years to decades. We show that, across 103 carbon-accrual measurements spanning management interventions globally, soils furthest from their mineralogical capacity are more effective at accruing carbon; sequestration rates average 3-times higher in soils at one tenth of their capacity compared to soils at one half of their capacity. Our findings provide insights into the world's soils, their capacity to store carbon, and priority regions and actions for soil carbon management.
Soil organic carbon (SOC) is an integral component of terrestrial ecosystems and plays an important role in ecosystem resilience and productivity. Soil organic matter contains nutrients that support plant growth and yield, retains water and reduces runoff, and resists erosion. Globally, SOC contains more carbon than the atmosphere and vegetation combined, climate- and land-use-induced changes to soil may alter SOC cycling and drive large terrestrial carbon sinks or sources; indeed, human land-use and land-cover change have resulted in a significant net loss of soil carbon over the past two centuries. Improved soil management practices that promote soil carbon sequestration, especially in stable carbon pools, are needed to reverse this trajectory and mitigate climate change.

Field observations suggest that more than half of soil organic carbon is chemically or physically associated with soil minerals. (Supplementary Fig. 1). These interactions limit microbial access to otherwise decomposable substrates and, consequently, mineral-associated organic carbon (MOC) can have turnover times up to 1000 times longer (reaching 100–10,000 years) than particulate organic carbon (POC) at the same depth. Thus, increasing MOC may be a key to lasting carbon sequestration efforts. However, despite its unique role, it is still unclear how much MOC presently exists, and how much could be accrued, in soils globally across depths and geographic regions. The capacity of soils to store MOC and other persistent forms of SOC will influence the long-term trajectory of the terrestrial carbon sink. Furthermore, it is unknown whether the proximity of a given soil to its carbon-storage capacity will influence the rate of C accrual, and thus the efficacy of soil C sequestration efforts. These knowledge gaps hinder climate change mitigation pathways and soil management initiatives, and limit long-term projections of Earth system models.

Soil organic matter decomposition, and its response and feedback to climate, depends on the physico-chemical form of the organic matter. To improve projections of long-term soil-climate feedbacks, it is imperative to predict the amount, distribution, and dynamics of MOC. However, little attention has been afforded to the explicit representation and parameterization of mineral-organic associations in Earth system models. Organic matter and minerals form associations via myriad mechanisms, including ligand exchange, hydrophobic interactions, and cation bridging, and the appropriate—representative yet tractable—mathematical formulation of MOC dynamics in these models is still the subject of research. Nevertheless, existing and candidate approaches require a means to constrain the maximum mineralogical capacity of MOC. This MOCmax is a property of the soil mineralogy. A data-driven approach to convert readily measured soil mineralogical variables to MOCmax is therefore needed to achieve advances in modeling and to enable robust estimates of carbon sequestration capacity.

Here we synthesized MOC observations from 1144 soil profiles spanning diverse biomes, soil types, and climates worldwide. Our synthesis included soils with a wide range of clay plus silt mineral content (CS; 1.5–100%), mean annual temperature (MAT; −2.9–29 °C), and mean annual precipitation (MAP; 79–3806 mm yr−1), as well as different vegetation types and land-uses (Supplementary Figs. 2 and 3). We leveraged these observations, with insights from theory and process-based models, and demonstrated that MOCmax can be inferred as an emergent property from readily measured soil mineralogical variables. We explored the variability of observed MOC and used a machine learning approach to elucidate the role of environmental variables, including climate and vegetation, in driving the observed departures from mineralogical saturation. We categorized sites into natural/less-managed (forest and grassland) and intensively managed (cropland) ecosystems to further investigate the effects of vegetation type and management on the degree of MOC undersaturation across soil depths. Finally, to explore how this undersaturation affects the sequestration efficacy of soils over decadal timescales, we examined 103 carbon-accrual measurements spanning management interventions globally.

Results and discussion
Carbon capacity of low- and high-activity mineral soils. Studies have presented conflicting results on the importance of clay plus silt content (CS) as a single linear predictor of MOC, with many of these analyses focusing on the prediction of bulk SOC, which contains additional pools of non-mineral-associated organic carbon (e.g., POC). As we present here, theory and model insights, as well as our extensive global data analysis, all suggest that there is no universal linear relationship between MOC (or SOC) and CS (Supplementary Figs. 4–6). Rather, this relationship depends, to a first-degree, on the combined effect of the C loading on minerals (g C m−2 mineral) which is a function of environmental conditions and management practices, and the effective mineral area on which C can bind (m2 g−1 mineral) which is a function of the type of mineral.

While it may be difficult for soils to reach MOCmax in practice, the proximity of a given soil to its maximum carbon capacity is an important factor for determining its effectiveness in sequestering additional carbon (see Supplementary Fig. 9 for a conceptual schematic) and, consequently, underestimating MOCmax can have significant implications for estimates of carbon deficit and potential accrual.

Vegetation and management controls on carbon under- saturation across depths. Our analysis suggests that many soils
are substantially below their mineralogical capacity (Fig. 1). This mineralogical undersaturation may be attributed to environmental (e.g., climate and plant C inputs) limitations on MOC storage and decomposition, management practices that result in MOC losses, or both. To explore where and why soils contain less MOC than they could based on their mineralogy alone, we calculated the mineralogical C saturation for a given soil as the difference from its mineralogical carbon capacity (MOCmax)—was calculated for each measurement in our observational synthesis. a %C saturation grouped by managed (n = 573) and natural (n = 862) ecosystems across depths (see “Methods”). b %C saturation by ecosystem (crop, forest, and grassland) and depth (surface < 30 cm and deep > 30 cm). Only categories with >5 observations for both surface (crop n = 425; forest n = 242; grassland n = 355) and deep (crop n = 21; forest n = 53; grassland n = 136) soils are shown. Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), 1.5× interquartile range (whiskers), and means (diamonds).
in surface soils (0–30 cm), grasslands were closer to C saturation in deeper horizons (30–120 cm) than forests were \((p < 0.005;\) Fig. 2b). This pattern confirms that root profiles and rates of C inputs influence the achieved MOC levels. This is an additional reason to expect that deeper-rooted vegetation may be more effective in sequestering C, and should thus be examined for restoring degraded lands and selecting cover crops\(^{41,46}\).

**Climate controls on mineral-associated organic carbon.** Climate also affects the achieved MOC in soil and its departure from saturation. However, MOC does not vary linearly with extrinsic factors such as temperature, but rather its response is governed by ‘pedogenic thresholds’ and nonlinearities\(^{47}\). We thus used a machine learning approach to determine the significance of key covariates and reveal emergent relationships of each variable, as well as variable interactions, on MOC (i.e., partial dependence plots; see “Methods”). This approach illuminates not only the overall sensitivity of MOC to individual variables but also the conditional relationship and rate of change with respect to a given variable (i.e., the partial derivative), without imposing a particular (e.g., linear) relationship. Specifically, we optimized a random forest (RF) model using key environmental covariates (see “Methods”, \(R^2 = 0.60\), Supplementary Fig. 11). We then calculated partial dependence relationships of MOC as a function of each variable alone (Supplementary Fig. 12), as well as conditional on interactions between variables (Fig. 3).

Clear predictive relationships emerged for MOC as a function of each variable. We found a linear partial dependence relationship between MOC and CS (Supplementary Fig. 12a), which can be attributed to an underlying increase in total mineral surface area and pore space\(^{23,48}\). This result supports studies that find correlations between MOC and CS, but illustrates the importance of controlling for the variability introduced by environmental covariates. With POC, we observed a saturating relationship in MOC (Fig. 3a), corresponding to an increase in C loading up to an effective saturation of mineral surfaces\(^{48,49}\). Although not all of this POC may be able to sorb onto mineral surfaces, it emerged as a significant variable in our analysis (Supplementary Figs. 11 and 12), likely as an indicator of rapidly decomposing material and the production of microbial necromass and dissolved organic carbon. In contrast, machine learning predictions of the conditional relationship between MOC and aboveground litter inputs showed a weaker dependence, consistent with other findings that aboveground productivity and litterfall are often not good proxies for carbon inputs to mineral soils\(^{42}\).

We observed a decline in MOC with MAT (Fig. 3b and Supplementary Fig. 12c) which is consistent with increased relative desorption with increasing temperatures\(^{30,31}\), and suggests a potential global vulnerability of MOC to warming. This decline with MAT also emerged across different ranges of POC, but interestingly, with significant differences in magnitude (Fig. 3b); low POC regimes (which are furthest from MOC saturation; Fig. 3a) exhibited greater temperature sensitivities compared to high POC regimes (soils closer to saturation; Fig. 3a). Distinct temperature-dependence regimes of saturating sorption curves are corroborated by theory (Supplementary Fig. 6) and experiments\(^{52}\). Specifically, when available carbon concentrations are low, MOC depends on the equilibrium constant \(K_{eq} = f(T)\) (i.e., the ratio of adsorption to desorption rate constants; Supplementary Fig. 6) which is a strong function of temperature, whereas when concentrations are high, MOC approaches \(MOC_{max} = f(T)\) and hence exhibits a weaker temperature dependence. Thus, our findings suggest that restoring degraded soils towards their mineralogical capacity through improved land management may not only contribute to carbon sequestration efforts\(^{53}\) but also impart a greater resiliency of soils to future warming.

Finally, we leveraged a machine learning approach to predict current MOC globally, excluding tundra, peatlands, and deserts (see “Methods”; \(R^2 = 0.79\), Supplementary Figs. 13 and 14). We estimated that MOC stocks total 899 Pg C (5–95% range: 668, 1074 Pg C) to a depth of 1 m, with 448 Pg C (296, 536) in topsoils (0–30 cm) and 451 Pg C (372, 538) in subsoils (30–100 cm) (Fig. 4 and Supplementary Fig. 15; Supplementary Tables 1 and 2). Globally, MOC made up a smaller proportion of total SOC in topsoils (0.66 ± 0.13; mean ± s.d.) compared to in subsoils (0.70 ± 0.17) (0.69 ± 0.15 to a depth of 1 m; Supplementary Fig. 16). Soils had a smaller proportion of MOC to SOC in boreal regions than in tropical and temperate regions, suggesting a larger proportion of non-protected carbon (i.e., POC) at high latitudes that may be vulnerable to warming\(^{54}\) (Supplementary Discussion). Our global estimates of MOC stocks, and as a fraction of total SOC, are to our knowledge the first such spatially- and depth-resolved data products, providing a crucial link for understanding soil carbon vulnerability, benchmarking soil carbon models\(^{17}\), and assessing the mineralogical potential for carbon sequestration globally.

**Fig. 3 Data-driven predictive relationships of mineral-associated organic carbon as a function of individual controlling factors and interactions.** A random forest model was used to disentangle the significance and emergent relationships of individual covariates, using partial dependence relationships that describe the marginal effect of each variable on the predicted mineral-associated carbon (MOC) response (see “Methods”). a Predicted MOC as a function of particulate organic C (POC). For each plot, the points show random forest model predictions; best fit lines and shading show 99% confidence intervals on each trend. b Predicted MOC as a function of mean annual temperature (MAT), conditional on three distinct subsets of POC—namely, low (L), medium (M), and high (H) POC corresponding to unsaturated to saturated regimes. The breakpoints were selected as the median and third quartile of the POC distribution, which also correspond to approximate saturation thresholds in the partial dependence plot of POC alone.
Global soil mineralogical carbon capacity. We used the same classification of mineral types from our observational synthesis to map high- and low-activity minerals (HM and LM, respectively; see “Methods”, Supplementary Fig. 14). Together with a global map of CS and our derived relationships for MOC_max in HM and LM soils, we calculated the corresponding MOC_max globally (Fig. 4 and Supplementary Fig. 17). This spatially explicit and depth-resolved global product provides insights into the capacity of the world’s soils to store MOC. Globally, we estimate that the soil mineralogical C capacity is 1443 ± 141 Pg C and 3153 ± 312 Pg C in topsoils and subsoils, respectively, totaling 4596 Pg C ± 453 Pg C to a depth of 1 m (excluding tundra, peatlands, and deserts; Supplementary Table 1). Many sites in temperate and subequatorial zones show the greatest mineralogical C deficit (MOC_max minus MOC), and thus potential for additional MOC storage, in part because of the prevalence of agricultural soils there (Fig. 4c and Supplementary Figs. 18–20). Indeed, we estimate that the mineralogical C deficit of croplands is 184 Pg C (5–95% range: 148, 225) and 509 Pg C (439, 569) in topsoils and subsoils, respectively (Supplementary Table 1). While reaching MOC_max is difficult and strongly limited by climate (Fig. 3), if these soils achieved the average %C saturation levels of natural lands (Fig. 2), this potential sequestration would total nearly 104 Pg C to a depth of 1 m (62 and 42 Pg C in topsoils and subsoils, respectively) (Supplementary Table 1). Grazing lands also show a large mineralogical C deficit covering a large expanse globally, but will require tailored strategies for soil C restoration, especially in arid and semi-arid regions that are additionally limited by climate (e.g., the Southwestern United States and Ustyurt Plateau). Our maps reveal hotspots of mineralogical C deficit, often associated with long-standing cropping and grazing lands (Fig. 4 and Supplementary Figs. 20 and 21), and highlight priority regions for soil C restoration efforts.

Current MOC stocks correspond to global average %C saturation levels of 42% (5–95% range: 24%, 61%) and 21% (14%, 30%) in topsoils and subsoils, respectively (Fig. 4c and Supplementary Figs. 18 and 19; Supplementary Table 2). However, as explained earlier, reaching C saturation is not necessarily feasible (given climate limitations, for example) or a recommended target. Rather, the potential lies initially in restoring the natural %C saturation in managed areas (Supplementary Figs. 18–22; Supplementary Table 1) and, especially, those furthest from %C saturation (Supplementary Fig. 9; Supplementary Discussion). We identify geographic locations with the most promise (namely, temperate and subequatorial zones in the northern hemisphere; Fig. 4c) and highlight the potential benefit of practices that sequester C deeper in the soil profile, where soil minerals are further from %C saturation (Supplementary Figs. 18–20). In addition to practices that leverage deeper-rooted vegetation41, recent studies have indicated that new approaches to deep ploughing55 or other practices resulting in topsoil-subsoil flipping56 may lead to considerable overall C sequestration,
although the cost and full impact of such practices on other soil functions needs to be better understood.

In addition to informing locations for targeted restoration and carbon sequestration, our data and results are also an essential input for Earth system models (ESMs) that seek to represent soil–mineral–organic associations explicitly—directly informing the mineralogical C capacity and constraining the proportion of total SOC that is mineral-associated (Supplementary Figs. 16 and 17)\textsuperscript{17}. Our global estimate of MOC\textsubscript{max} constitutes an effective potential for sorption on minerals, unlike that derived from laboratory DOC sorption experiments alone\textsuperscript{57,58}. Laboratory assays tend to underestimate maximum sorption potential, especially if they use native soil which already has some carbon loading on mineral surfaces\textsuperscript{57,58}. Moreover, microbial necromass and residues contribute substantially to MOC measured in the field\textsuperscript{19}, and thus, our estimates have the potential to uniquely improve model parameterizations beyond those obtained from laboratory measurements.

Carbon sequestration implications and future perspectives.

Geographic regions with higher C deficit can be used to inform soil C sequestration efforts, such as the "2 per 1000" initiative\textsuperscript{60,61} and other nature-based climate solutions\textsuperscript{62,63}. Soils with higher deficit (Supplementary Fig. 20), and thus lower %C saturation (Fig. 4c, Supplementary Fig. 18), may provide greater C sequestration efficacy (Supplementary Fig. 9), through C accumulation rates that are larger and can be sustained for years to decades until a new steady-state is reached\textsuperscript{39}. We synthesized studies that measured C accrual rates following management interventions intended to promote C storage (n = 103, ranging from 2 to 65 years; Supplementary Figs. 23 and 24) and found that soils furthest from their mineralogical capacity (i.e., lowest %C saturation) achieve the highest rates of C accrual (Fig. 4d). In fact, carbon accrual rates were on average three times higher in soils from sites with <10% C saturation compared to sites with 50% C saturation (Fig. 4d). This result is critical for identifying global regions that, in addition to a significant capacity for C storage, may also provide greater C sequestration efficacy (Supplementary Fig. 9; Supplementary Discussion). Indeed, many of these potential sequestration hotspots are located in extensive cropping regions (e.g., Midwestern United States and India; Fig. 4c and Supplementary Figs. 21 and 22) that could be incentivized to sequester C, with additional co-benefits for soil productivity and food security\textsuperscript{64,65}.

Estimates of global potential soil C sequestration rates vary considerably, but recent reviews suggest the annual technical potential of individual practices (e.g., no-till agriculture and agro-forestry) applied globally is on the order of 2–5 Pg CO\textsubscript{2} yr\textsuperscript{-1} (0.5–1.4 Pg C yr\textsuperscript{-1})\textsuperscript{62,66,67}; implementation potentials given economic considerations fall at the lower end of this range\textsuperscript{68}, while nutrient limitations may further limit achievable sequestration\textsuperscript{69,70}. While the implementation of soil management interventions across the world’s >500 million active farms poses formidable socioeconomic challenges\textsuperscript{71,72}, actionable valuations of in targeted areas\textsuperscript{73,74}, with enduring co-bene...
properties (e.g., cation exchange capacity, iron, and aluminum content, among others), where available. To locate studies, we searched Google Scholar for co-
minated keywords: "high C and Mg proportions", "carbonate associated", "clay fraction", "density fractionation", and "carbon saturation". We found that studies that measured MOC by both size and density fractionation (e.g., refs. 89–91) were consistent and within the limit of MOCmax for each mineral type, and that both methods concluded similar ranges of MOC to SOC ratios92 (Supplementary Fig. 1). Indeed, our synthesis supported the use of either fractionation method for the goals of our particular study.

The observational synthesis totalled 1144 mineral soil profiles from 78 studies that reported fractionation and bulk measurements of carbon and nitrogen across depths. The dataset included: 359 (<2 μm; clay fraction) and 1451 (<63 μm; clay and silt fraction) MOC measurements, 1107 POC measurements, and 1432 SOC measurements, as well as corresponding MAT, MAP, CS, soil order, mineral type, and vegetation type (see section “Defining and evaluating ecosystem categories” for classification details). We also collected other auxiliary soil properties, including diithionite- and oxalate-extractable iron (FeD and FeOx, respectively), oxalate-extractable aluminium (AlOx), cation exchange capacity (CEC), specific surface area (SSA), and total reserve bases (TRB), though these measurements were sparse across the collected studies. A complete list of studies and their location, climate, and vegetation is provided in Supplementary Table 3. Studies in our synthesis spanned intensively managed (e.g., crop) to natural/ less- managed (e.g., grassland, forest) ecosystems across diverse climates and soil types (Fig. 1 and Supplementary Fig. 2). We obtained a representative mix of MOC measurements from managed (n = 559 profiles) and natural (n = 585 profiles) lands. We focused on the current management status herein, but note that future studies may consider information on historical management, where available. Within each climate and vegetation type, we expected to have a greater representation of temperate ecosystems, as opposed to tropical and boreal ecosystems (Supplementary Figs. 2, 3, and 14)— especially when including the European LUCAS database11,92 (Supplementary Fig. 2b, d), but our model results were robust to its exclusion. Furthermore, measurements in tundra and deserts were limited, highlighting the need for studies in these regions in the future (Supplementary Figs. 3a, b). Nevertheless, our observational synthesis spanned diverse climates and soil types: MAT ranged from −2.9 to 29 °C, MAP from 79 to 3806 mm yr−1, and CS from 1.5 to 100%, across different mineral and vegetation types (Fig. 1b, c and Supplementary Fig. 2).

Indeed, our synthesis spanned the majority of the climate-edaphic covariate space encompassed by the world’s soils in WoSIS (World Soil Information Service)39 (Supplementary Fig. 3a, b). We further illustrated the general representativeness of our synthesis within a multi-dimensional covariate space using a principal component analysis (PCA) and comparing to randomly sampled WoSIS profiles (Supplementary Fig. 3c). We note that even WoSIS soil profiles showed a greater representation in temperate ecosystems compared to tropical and boreal ecosystems84 (Supplementary Fig. 3b), further highlighting the need for additional measurements in these regions.

Defining and evaluating ecosystem categories. Here we classified soils into two broad categories of mineral type and activity; namely, low-activity minerals (LM) and high-activity minerals (HM). We used soil order from each individual study and, when provided, details on clay composition to categorize the soil of each site. Phyllosilicate clay minerals—composed of tetrahedral silicate sheets and octahedral hydroxide sheets—are composed of either 1:1 (one tetrahedral and one octahedral sheet) or 1:2 (one octahedral sheet sandwiched between two tetrahedral sheets) clays. 1:1 clays are prevalent in kaolinitic soils and generally have a lower SSA14,28,29,95; therefore, such soils generally exhibit a lower capacity to stabilize MOCmax for each depth was not theoretically or empirically warranted. While it is more difficult to reach MOCmax in subsoils due to lower carbon inputs, MOCmax is an intrinsic property of the soil mineralogy and constitutes a theoretical upper limit that does not depend on depth; indeed, the depth independent of depth; indeed, this depth-independence is important for modeling C cycling and predicting future C sequestration, because ecosystems differ in their MOC capacity, the value of MOCmax is independent of vegetation type86,100. Regressions for both mineral types were fit with a forced intercept through the origin, since, by definition, no clay- and silt-associated organic carbon (i.e., MOC) can exist without clay and silt minerals present. While some size fractionation studies report MOC in sandy soils with very low amounts of clay and silt minerals92, further exploration is needed to understand the relative contribution of fine POC and DOC in such soils16.

Statistical modeling and predictive relationships. Due to potential nonlinear relationships between MOC and environmental variables, we used an ensemble modeling approach—Random Forest (RF)—to rank MOC predictors and their effects on MOC. This method has been shown to reduce overfitting to a training dataset compared to other machine learning methods, and does not suffer from multi-collinearity (i.e., the linear dependence among predictor variables), as do multiple regression analyses102. We used the RandomForest package in R for RF analyses103,104. We used an ensemble of 300 independently trained RF models (using a 75–25% train-test split; 400 decision trees each) with bootstrapped sampling to robustly assess model performance105—mean absolute error (MAE) and mean-squared error (MSE) to quantify the model error and R2 to estimate the proportion of variance in MOC explained by the model (Supple-

centary Fig. 11). Our aim was to: (i) elucidate key variables by assessing their relative contribution to the MOC of test-set predictions when removed or permuted, (ii) calculate the total variance explained by selected predictors, and (iii) predict the response of MOC to a given variable while controlling for all other variables. RF is often used to rank the importance of each variable in a regression through permutations of the data39,101. Nevertheless, after preliminary investigation to select the most important variables, while also considering variables that were measured in sufficient studies to retain the greatest number of complete observations, we identified CS, MAT, MAP, and POC as key continuous predictors and vegetation (land cover) and mineral type as key categorical predictors (Supplementary Fig. 1f; R2 = 0.60 ± 0.06). Specifically, variables were ranked from most to least important, where the most important variables showed the highest increase in MSE of test set predictions when permuted in the model.

We then investigated the underlying relationship of MOC to key variables (Fig. 3 and Supplementary Fig. 12). RF allowed us to explore the effect of each individual variable on MOC, by integrating the predicted response over the contribution of all other variables (i.e., partial dependence plot)106. That is, for a variable x, the partial dependence of MOC on x is approximated by f(x, x) − f(x̄, x̄), where x̄ denotes the marginal effect of x̄ on the prediction, and is
estimated as:
\[
f(x, y) = \frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i)
\]

This was calculated for each predictor variable over the full range of the other predictor variables (Supplementary Fig. 12), and also over relevant subsets of the other predictor variables to explore conditional partial dependencies and visualize two-way variable interactions (e.g., Fig. 3). For the case of MAT and POC, we explored the temperature sensitivity of MOC (i.e., the partial dependence relationship with MAT) conditional on low (L), medium (M), and high (H) values of POC (details in Supplementary Fig. 26). We note that a global map of POC values differed from the results of a generalized linear model (GLM) and two-way interactions on the response variable of interest (here MOC) and uncovers emergent (linear, monotonic, or nonlinear) relationships, providing unique insights for MOC response to future climate and management interventions (e.g., increasing POC or CS). For comparison, we also repeated the above analysis with a generalized linear model (GLM) and find qualitatively similar, though linearized, trends between MOC and the individual environmental predictors.

Global mineral-associated carbon predictions and uncertainty. Following the exploration of underlying climatic and edaphic properties, we trained a RF model on our observational synthesis to predict MOC globally. Specifically, we trained the RF model to predict MOC from MAT, MAP, CS, SOC, vegetation type, and mineralogy (Supplementary Fig. 26). We note that a global map of MOC values did not exist to be used as an explanatory variable, as was used in the attribution analyses (Supplementary Fig. 11); rather, global POC values can be derived as a product of this study. The Scikit-learn Library in the Python environment was used for global predictions and Matplotlib Basemap for global mapping. We again used an ensemble of 300 independently trained RF models (using a 75% product of this study. We note that the predictive RF model \((R^2 = 0.79 \pm 0.05)\); Supplementary Table 1), both the HWSD and SoilGrids\(^4,98\), and also as the average of the two gridded data sets1015 g), we used BD to calculate MOCmax stocks (in kg C m\(^{-2}\) soil) and then summed over all grid cells containing mineral soils (Supplementary Tables 1, 2).

The mineralogical capacity MOCmax (Supplementary Fig. 17) was then compared to predicted MOC stocks (Supplementary Fig. 15) to calculate the mineralogical C deficit (by subtracting MOC from MOCmax) and %C saturation (by dividing MOC/MOCmax) globally (Fig. 4; Supplementary Figs. 18–20). We note that this limit is based on mineralogy alone, and many lands may be further limited by climate. Thus, in addition to the C deficit relative to the mineralogical capacity (MOCmax), we also estimated the C deficit relative to an environmental limit calculated using the natural land average %C saturation (i.e., average MOC/ MOCmax) observed in our MOC synthesis at each depth (Fig. 2; i.e., 51 and 19% for the average %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.

Our spatially explicit, depth-resolved global estimates highlight regions where larger C deficits (lower %C saturation) can be targeted for soil restoration and C sequestration efforts, especially on managed lands (Fig. 4c; Supplementary Figs. 21 and 22). While environmental conditions may make carbon accumulation more challenging in certain regions than elsewhere, the global majority of current soil carbon stocks (and consequently MOC) stocks (Supplementary Table 2). However, our estimates using the averaged soil carbon data product agree with the distributions of MOC stocks and %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.

Observational synthesis and analysis of carbon accrual. Here we aimed to quantify the ability of topsoils ranging in mineralogical carbon saturation to further accrete carbon. We hypothesized that soils further from their mineralogical capacity (i.e., lower MOC saturation) would accrue carbon faster than soils closer to their capacity. To this end, we performed an observational synthesis from manipulation and chronosequence studies that included initial and final carbon stocks or concentrations to calculate these stocks, bulk density, experimental duration, and carbon content, together with a variety of other information that would facilitate MOC stocks, in lands already dominated by managed lands (Fig. 4c; Supplementary Figs. 21 and 22). While environmental conditions may make carbon accumulation more challenging in certain regions than elsewhere, the global majority of current soil carbon stocks (and consequently MOC) stocks (Supplementary Table 2). However, our estimates using the averaged soil carbon data product agree with the distributions of MOC stocks and %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.

Our spatially explicit, depth-resolved global estimates highlight regions where larger C deficits (lower %C saturation) can be targeted for soil restoration and C sequestration efforts, especially on managed lands (Fig. 4c; Supplementary Figs. 21 and 22). While environmental conditions may make carbon accumulation more challenging in certain regions than elsewhere, the global majority of current soil carbon stocks (and consequently MOC) stocks (Supplementary Table 2). However, our estimates using the averaged soil carbon data product agree with the distributions of MOC stocks and %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.

Observational synthesis and analysis of carbon accrual. Here we aimed to quantify the ability of topsoils ranging in mineralogical carbon saturation to further accrete carbon. We hypothesized that soils further from their mineralogical capacity (i.e., lower MOC saturation) would accrue carbon faster than soils closer to their capacity. To this end, we performed an observational synthesis from manipulation and chronosequence studies that included initial and final carbon stocks or concentrations to calculate these stocks, bulk density, experimental duration, and carbon content, together with a variety of other information that would facilitate MOC stocks, in lands already dominated by managed lands (Fig. 4c; Supplementary Figs. 21 and 22). While environmental conditions may make carbon accumulation more challenging in certain regions than elsewhere, the global majority of current soil carbon stocks (and consequently MOC) stocks (Supplementary Table 2). However, our estimates using the averaged soil carbon data product agree with the distributions of MOC stocks and %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.

Observational synthesis and analysis of carbon accrual. Here we aimed to quantify the ability of topsoils ranging in mineralogical carbon saturation to further accrete carbon. We hypothesized that soils further from their mineralogical capacity (i.e., lower MOC saturation) would accrue carbon faster than soils closer to their capacity. To this end, we performed an observational synthesis from manipulation and chronosequence studies that included initial and final carbon stocks or concentrations to calculate these stocks, bulk density, experimental duration, and carbon content, together with a variety of other information that would facilitate MOC stocks, in lands already dominated by managed lands (Fig. 4c; Supplementary Figs. 21 and 22). While environmental conditions may make carbon accumulation more challenging in certain regions than elsewhere, the global majority of current soil carbon stocks (and consequently MOC) stocks (Supplementary Table 2). However, our estimates using the averaged soil carbon data product agree with the distributions of MOC stocks and %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.

Observational synthesis and analysis of carbon accrual. Here we aimed to quantify the ability of topsoils ranging in mineralogical carbon saturation to further accrete carbon. We hypothesized that soils further from their mineralogical capacity (i.e., lower MOC saturation) would accrue carbon faster than soils closer to their capacity. To this end, we performed an observational synthesis from manipulation and chronosequence studies that included initial and final carbon stocks or concentrations to calculate these stocks, bulk density, experimental duration, and carbon content, together with a variety of other information that would facilitate MOC stocks, in lands already dominated by managed lands (Fig. 4c; Supplementary Figs. 21 and 22). While environmental conditions may make carbon accumulation more challenging in certain regions than elsewhere, the global majority of current soil carbon stocks (and consequently MOC) stocks (Supplementary Table 2). However, our estimates using the averaged soil carbon data product agree with the distributions of MOC stocks and %C saturation in our observational synthesis (Supplementary Fig. 19; Supplementary Table 2), and provide a robust estimate that incorporates the range of current soil carbon stocks.
For each entry (a pair of initial and final carbon stocks), we calculated the maximum mineralogical capacity (MOC\textsubscript{max}) using the data-derived relationships (Fig. 4b) for clay + silt + sand (CS) and fine fraction (HM or LM, respectively) of mineral in each soil layer and used respective bulk densities to calculate the layer MOC\textsubscript{max} stocks (t C ha\textsuperscript{-1}) which were then summed across the considered depth to obtain the profile MOC\textsubscript{max} stock (t C ha\textsuperscript{-1}). In cases where initial and final soil profiles differed in clay and silt content, their average was used. However, if the difference resulted in these soil profiles falling into different texture classes, they were excluded from further analysis. Furthermore, to calculate initial and final stocks, we considered initial and final profiles that were taken at equal depths, which ranged between 10 and 40 cm across studies. Where necessary, initial and final profiles were standardized to equal depths using the depth of the shallower profile as a cutoff ranging between 30 and 39 cm (n = 5 profiles).

We calculated the mineralogical % C saturation as the initial SOC stock divided by the initial MOC\textsubscript{max} stock. Because MOC was not measured in all accrual studies, we used the measured SOC and CS values to calculate the corresponding MOC using our random forest model trained on the MOC synthesis data with analogous SOC and CS predictors (R\textsuperscript{2} = 0.70 ± 0.06; using Scikit-Learn in Python, see section “Global mineral-associated carbon predictions and uncertainty”).

Carbon accrual (t C ha\textsuperscript{-1} y\textsuperscript{-1}) was calculated as the difference between final and initial C stock divided by the time period between the repeated samplings (n = 32), the duration of the experiment in the case of a paired plot experimental design (n = 30), the age difference between two studies - a chronosequence (n = 24), the time span at which a regression across multiple samplings was used to obtain an average rate for a chronosequence (n = 10), or the time of soil formation after emergence of a SOC-free parent material (n = 7).

Accrual studies spanned crop, pasture, grassland, and forest ecosystems across different climates and soil types (Supplementary Table 4; Supplementary Fig. 23). We focused on studies in which carbon inputs were manipulated or land-use was altered in a way that carbon accrual could be expected. These included improved agriculture (cover cropping, no till, crop-livestock rotation, deep ploughing, grazing reduction), land use change (conversion of cropland to grassland or forest), double cropped or woody vegetation (from the DIRT (Detrital Input and Removal Experiment) network), and natural revegetation and mine reclamation (formerly self-starting asymptotic regression model (SSAsymp) on the 103 rate measurements (R\textsuperscript{2} = 0.018; Fig. 4d). The same approach was also used to fit the 10th and 90th quantiles (Fig. 4d). There was substantial variability in observed accrual rates for a given % C saturation, due to differences in climate and land use across sites (Supplementary Fig. 23). However, most of this variability appeared in soils furthest from C saturation, where environmental and management factors also play a role in observed accrual rates. In soils closest to saturation, this variability was far lower as carbon accrual rates tended towards zero, illustrating the importance of the mineralogical limit. Our focus here was on constraining an average empirical rate of carbon accrual as a function of % C saturation and investigating whether soils furthest from C saturation achieve higher carbon accrual rates. We also note that the variability in C accrual rates was better explained by % C saturation (R\textsuperscript{2} = 0.18) than by initial SOC (R\textsuperscript{2} = 0.04-0.10; Supplementary Fig. 24), further demonstrating the relevance of the mineralogical limit.

We used the nonlinear asymptotic regression to provide a spatially explicit global estimate of potential C accrual rates in topsoils dominated by management (Supplementary Fig. 22). While such accrual rates would be difficult to achieve everywhere, our spatially explicit estimates of potential C accrual rates highlight priority regions for future soil restoration and C sequestration initiatives and provide an independent estimate of global soil C sequestration potential.

Data availability

The globally gridded maps of mineral-associated carbon and mineralogical carbon capacity derived in this study are freely available and archived at Zenodo (https://doi.org/10.5281/zenodo.6539765). The observational syntheses (mineral-associated carbon and accrual measurements) in support of these findings are detailed in Supplementary Tables 3 and 4, and archived at Zenodo (https://doi.org/10.5281/zenodo.5987415).

Received: 20 January 2022; Accepted: 9 June 2022; Published online: 01 July 2022

References

1. Tieszen, L. L., Cuevas, E. & Chacon, P. The role of soil organic matter in sustaining soil fertility. Nature 371, 783–785 (1994).
2. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–436 (2000).
3. Todd-Brown, K. E. O. et al. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences 10, 1717–1736 (2013).
4. Hengl, T. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12, e0196748 (2017).
5. Sanderman, J., Hengl, T. & Fiske, G. J. Soil carbon debt of 12,000 years of human land use. Proc. Natl Acad. Sci. USA 115, 1–6 (2018).
6. Houghton, R. A. et al. Carbon emissions from land use and land-cover change. Biogeosciences 9, 5125–5142 (2012).
7. Gasser, T. et al. Historical CO2 emissions from land use and land cover change and their uncertainty. Biogeosciences 17, 4075–4101 (2020).
8. Chabbi, A. et al. Aligning agriculture and climate policy. Nat. Clim. Chang. 7, 303–307 (2017).
9. Rumpel, C. et al. Put more carbon in soils to meet paris climate pledges. Nat. Microbiol. 3, 32–34 (2019).
10. Kögel-Knabner, I. et al. Organo-mineral associations in temperate soils: integrating biology, mineralogy, and organic matter chemistry. J. Plant Nutr. Soil Sci. 171, 61–82 (2008).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-31540-9 | ARTICLE

9
