










# Global stocks and capacity of mineral-associated soil organic carbon

Katerina Georgiou <sup>1,2</sup>✉, Robert B. Jackson <sup>2,3,4</sup>, Olga Vindušková <sup>5,6</sup>, Rose Z. Abramoff<sup>7,8</sup>, Anders Ahlström <sup>9</sup>, Wenting Feng <sup>10</sup>, Jennifer W. Harden<sup>2,11</sup>, Adam F. A. Pellegrini <sup>12,13</sup>, H. Wayne Polley <sup>14</sup>, Jennifer L. Soong<sup>15,16</sup>, William J. Riley <sup>17</sup> & Margaret S. Torn <sup>17,18</sup>

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.

<sup>1</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA. <sup>2</sup>Department of Earth System Science, Stanford University, Stanford, CA 94305, USA. <sup>3</sup>Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA. <sup>4</sup>Precourt Institute for Energy, Stanford University, Stanford, CA 94305, USA. <sup>5</sup>Department of Biology, University of Antwerp, Antwerp 2000, Belgium. <sup>6</sup>Institute for Environmental Studies, Charles University, Prague 128 01, Czech Republic. <sup>7</sup>Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette F-91191, France. <sup>8</sup>Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA. <sup>9</sup>Department of Physical Geography and Ecosystem Science, Lund University, Lund SE-22100, Sweden. <sup>10</sup>Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 10081, China. <sup>11</sup>U.S. Geological Survey, Menlo Park, CA 94035, USA. <sup>12</sup>Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. <sup>13</sup>Cambridge Conservation Institute, University of Cambridge, Cambridge CB2 3EA, UK. <sup>14</sup>Agricultural Research Service, U.S. Department of Agriculture, Temple, TX 76502, USA. <sup>15</sup>Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80523, USA. <sup>16</sup>Granular, Inc, San Francisco, CA 94103, USA. <sup>17</sup>Climate and Ecosystem Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>18</sup>Energy and Resources Group, University of California, Berkeley, Berkeley, CA 94720, USA. ✉email: [georgiou1@llnl.gov](mailto:georgiou1@llnl.gov)

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<sup>1</sup>. Globally, SOC contains more carbon than the atmosphere and vegetation combined<sup>2–4</sup>. 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<sup>5–7</sup>. Improved soil management practices that promote soil carbon sequestration, especially in stable carbon pools, are needed to reverse this trajectory and mitigate climate change<sup>8,9</sup>.

Field observations suggest that more than half of soil organic carbon is chemically or physically associated with soil minerals<sup>10,11</sup> (Supplementary Fig. 1). These interactions limit microbial access to otherwise decomposable substrates<sup>12,13</sup> 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<sup>14,15</sup>. Thus, increasing MOC may be a key to lasting carbon sequestration in soils<sup>16</sup>. 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<sup>17</sup>. Yet, little attention has been afforded to the explicit representation and parameterization of mineral-organic associations in Earth system models<sup>18</sup>. Organic matter and minerals form associations via myriad mechanisms, including ligand exchange, hydrophobic interactions, and cation bridging<sup>19</sup>, and the appropriate—representative yet tractable—mathematical formulation of MOC dynamics in these models is still the subject of research<sup>20,21</sup>. Nevertheless, existing and candidate approaches require a means to constrain the maximum mineralogical capacity ( $MOC_{max}$ ). This  $MOC_{max}$  is a property of the soil mineralogy<sup>22,23</sup>. A data-driven approach to convert readily measured soil mineralogical variables to  $MOC_{max}$  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<sup>-1</sup>), 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  $MOC_{max}$  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.

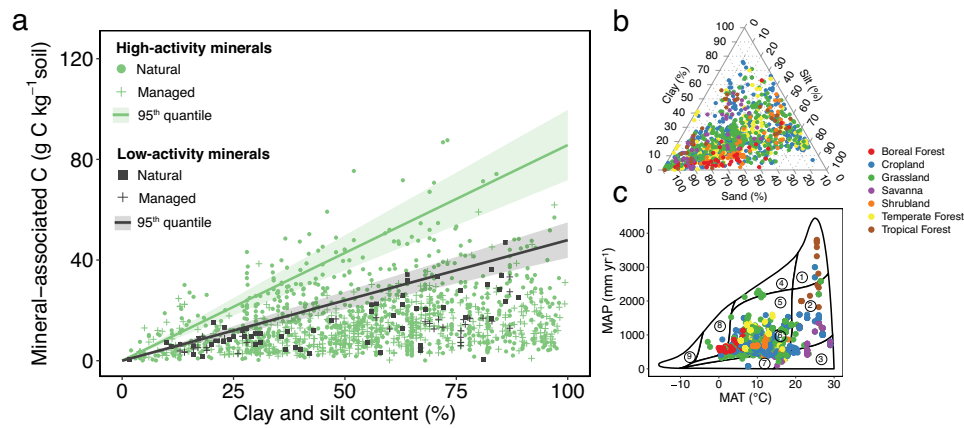
## Results and discussion

**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<sup>24–26</sup>, 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<sup>27</sup>, and the effective mineral area on which C can bind ( $m^2\ g^{-1}\ mineral$ ) which is a function of the type of mineral<sup>22,28,29</sup>. However, since  $MOC_{max}$  is driven primarily by the amount ( $g\ mineral\ kg^{-1}\ soil$ ) and type of mineral<sup>23,28</sup>, we hypothesized that if soils were partitioned based on their dominant mineral type, CS would emerge as a significant predictor of  $MOC_{max}$  globally.

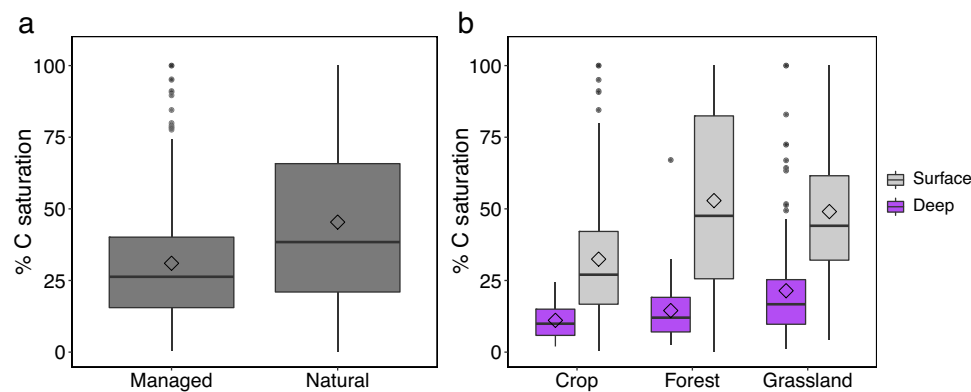
To test this hypothesis, we classified soils according to their mineral type and explored whether the upper quantiles of observed MOC form distinct boundaries for different mineral categories, i.e., whether a distinct  $MOC_{max}$  specific to each mineral type arises (Supplementary Figs. 6 and 7). Here, we classified soils as containing high- or low-activity minerals (HM or LM, respectively) based on their primary composition of high-activity (illite, smectite, vermiculite, chlorite) and low-activity (kaolinite, gibbsite) clays<sup>29–31</sup> (see “Methods” for the robustness of results to additional categories).

Indeed, our data synthesis quantifies and supports this hypothesis, demonstrating that a robust  $MOC_{max}$  estimate can emerge as a function of only CS and the type of mineral (Fig. 1; and irrespective of vegetation type, Supplementary Fig. 8). We obtained a conservative estimate of  $MOC_{max}$  as the 95th quantile of MOC values for each mineral type, resulting in  $MOC_{max}$  estimates of  $86 \pm 9$  and  $48 \pm 6\ mg\ C\ g^{-1}\ mineral$  for HM and LM, respectively (Fig. 1; slope  $\pm 90\%$  confidence interval, following a unit conversion). These data-derived estimates of  $MOC_{max}$  provide a maximum carbon potential for a given CS and mineral type, which can be used to inform the location and management practices of soil restoration and sequestration efforts (Supplementary Discussion). These values broadly agree with regional studies that have included high-activity minerals at comparable soil depths<sup>28,32,33</sup>, but are significantly greater than studies that have omitted mineral activity<sup>24,34–37</sup>. Given the important role of mineralogy (Supplementary Fig. 4), our data-driven  $MOC_{max}$  estimates are also a notable update to regional studies that have reported a maximum carbon potential across European soils (e.g.,  $45$ – $50\ g\ C\ kg^{-1}\ soil$  based on model-predicted values in ref. 11). While it may be difficult for soils to reach  $MOC_{max}$  in practice, the proximity of a given soil to its maximum capacity is an important factor for determining its effectiveness in sequestering additional carbon (see Supplementary Fig. 9 for a conceptual schematic)<sup>38,39</sup> and, consequently, underestimating  $MOC_{max}$  can have significant implications for estimates of carbon deficit and potential accrual.

**Vegetation and management controls on carbon undersaturation across depths.** Our analysis suggests that many soils



**Fig. 1** Maximum mineral-associated carbon as a function of clay and silt content for high- and low-activity mineral soils estimated from a global observational synthesis. **a** Mineral-associated organic carbon (MOC;  $\text{g C kg}^{-1}$  soil) as a function of clay and silt content (CS; %) across sites. The maximum slope (fit as the 95th quantile) for each soil type represents the intrinsic capacity of minerals to store carbon and depends on mineral composition. High-activity minerals (HM) include soils dominated by illite, smectite, vermiculite, and chlorite ( $n = 1303$ ) and low-activity minerals (LM) include soils dominated by kaolinite and gibbsite ( $n = 93$ ). Shading depicts 90% confidence intervals on the slope. Soils near the upper quantiles are closer to MOC saturation. Within HM and LM soils, filled markers and crosses denote natural and managed ecosystems, respectively. **b, c** Distribution of soil texture (**b**) and climate (**c**) across sites. Black polygons depict Whittaker’s biomes<sup>119</sup> according to mean annual temperature (MAT;  $^{\circ}\text{C}$ ) and mean annual precipitation (MAP;  $\text{mm yr}^{-1}$ ) values, following: (1) tropical rainforest; (2) tropical seasonal rainforest/savanna; (3) subtropical desert; (4) temperate rainforest; (5) temperate seasonal forest; (6) woodland/shrubland; (7) temperate grassland/desert; (8) boreal forest; and (9) tundra.



**Fig. 2** Percent mineralogical carbon saturation across ecosystems and soil depths. Percent mineralogical C saturation (%C saturation)—i.e., the proximity of a soil to its mineralogical carbon capacity ( $\text{MOC}_{\text{max}}$ )—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 (cropland, forest, and grassland) and depth (surface  $< 30$  cm and deep  $> 30$  cm). Only categories with  $> 5$  observations for both surface (cropland  $n = 425$ ; forest  $n = 242$ ; grassland  $n = 355$ ) and deep (cropland  $n = 21$ ; forest  $n = 53$ ; grassland  $n = 136$ ) soils are shown. Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), 1.5 $\times$  interquartile range (whiskers), and means (diamonds).

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 ratio of observed MOC to  $\text{MOC}_{\text{max}}$  (see “Methods”; Supplementary Fig. 10). Across all sites, soils averaged  $40 \pm 2\%$  C saturation ( $\pm 95\%$  confidence interval of the mean). Land-use change and/or poor management practices that degrade soils can decrease MOC, a result evident in the shift of C saturation distributions from less-managed to more intensively managed ecosystems (Fig. 2a). The former (namely, grasslands and forests), which we refer to here as natural ecosystems for brevity, had on average higher levels of C saturation than did agricultural systems ( $p < 0.0001$ ; Fig. 2a). Indeed, whereas soils in natural ecosystems averaged  $46 \pm 3\%$  C saturation, agricultural systems averaged only

$31 \pm 2\%$  C saturation. This contrast suggests that restoring degraded or intensively managed lands could push their soils towards higher C saturation levels (i.e., their highest attainable MOC given the site-specific environmental factors that limit stocks below  $\text{MOC}_{\text{max}}$ ) and allow them to serve as a substantial C sink.

We also observed that deeper soils were further from C saturation than were surface soils ( $p < 0.0001$ ; Fig. 2b). On average, surface soils (0–30 cm) were  $43 \pm 2\%$  C saturated, as compared to  $19 \pm 3\%$  C saturation for deeper soils (30–120 cm). This difference is likely due, in part, to lower rates of C inputs with depth<sup>40,41</sup>. The location and type of C inputs are also important, as a higher proportion of plant C is retained in SOC from belowground root inputs than from aboveground litter inputs<sup>42–44</sup>. We may then expect a difference in C saturation levels between grasslands and forests, which have different root depth distributions<sup>45</sup>. Indeed, although there was no significant difference in the level of saturation between grasslands and forests

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<sup>41,46</sup>.

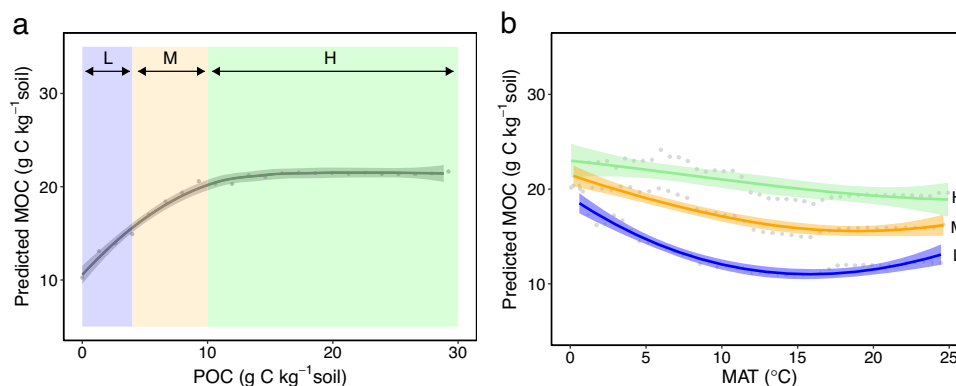
**Climate controls on mineral-associated organic carbon.** Climate also affects the achieved MOC in soil and its departure from saturation. However, MOC does not vary linearly with extrinsic factors such as temperature, but rather its response is governed by ‘pedogenic thresholds’ and nonlinearities<sup>47</sup>. 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<sup>22,48</sup>. 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<sup>48,49</sup>. 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

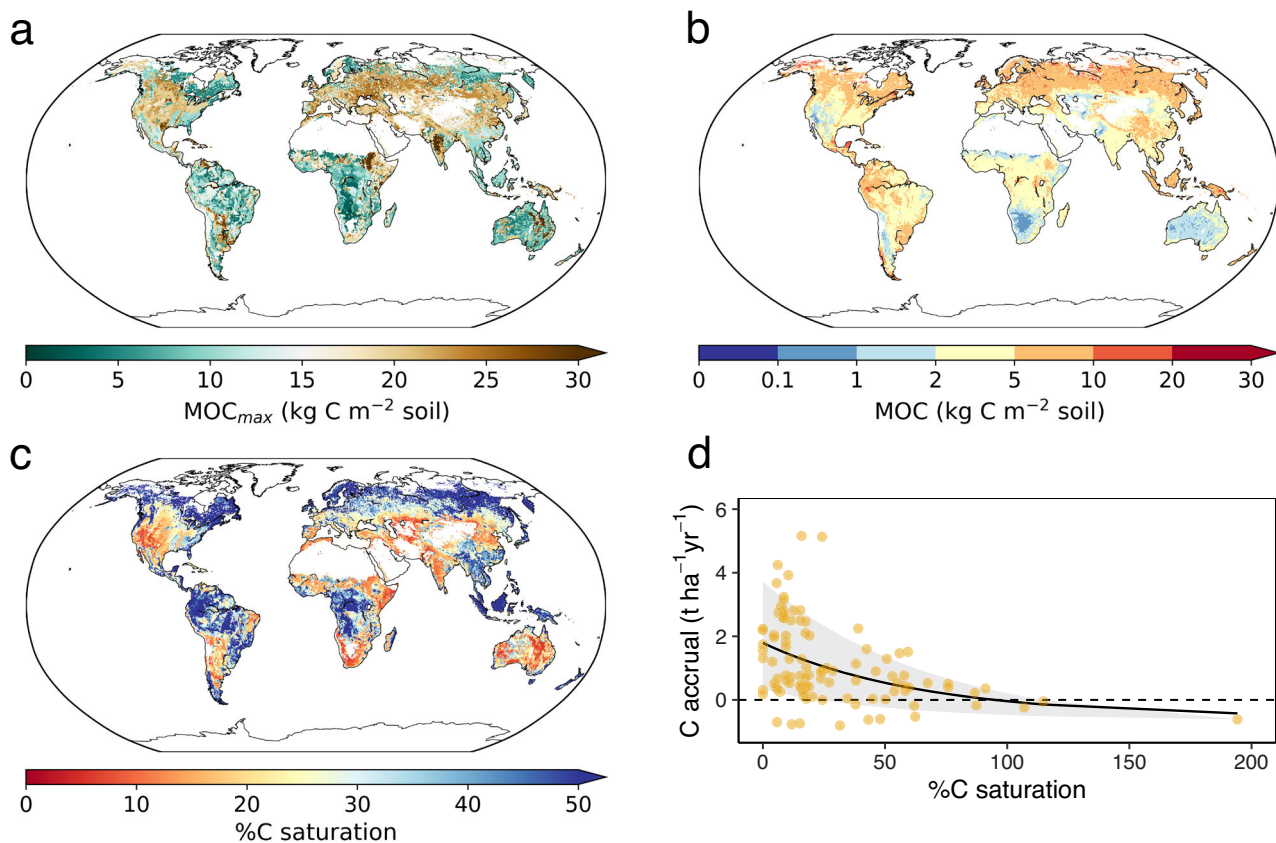
findings that aboveground productivity and litterfall are often not good proxies for carbon inputs to mineral soils<sup>42</sup>.

We observed a decline in MOC with MAT (Fig. 3b and Supplementary Fig. 12c) which is consistent with increased relative desorption with increasing temperatures<sup>50,51</sup>, 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<sup>52</sup>. 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  $MOC_{max} \neq f(T)$  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<sup>53</sup> 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 ( $0.66 \pm 0.13$ ; mean  $\pm$  s.d.) compared to in subsoils ( $0.70 \pm 0.17$ ) ( $0.69 \pm 0.15$  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<sup>54</sup> (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<sup>17</sup>, 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.



**Fig. 4** Global soil mineralogical carbon capacity, mineral-associated carbon stocks, and potential for carbon accrual. **a** Mineralogical carbon capacity ( $MOC_{max}$ ;  $\text{kg C m}^{-2}$  soil), **b** mineral-associated organic carbon (MOC;  $\text{kg C m}^{-2}$  soil), and **c** percent mineralogical carbon saturation (%C saturation) in topsoil (0–30 cm) at  $0.5^\circ$  resolution, excluding tundra, peatlands, and deserts.  $MOC_{max}$  was estimated using our data-derived relationships with the amount and type of mineral. MOC was predicted using a random forest model trained on our observational dataset. %C saturation was calculated as  $MOC/MOC_{max}$ . Low %C saturation corresponds to a greater C deficit, and therefore highlights regions where targeted soil management could lead to higher C sequestration. **d** Carbon accrual in topsoil ( $\text{t ha}^{-1}\text{yr}^{-1}$ ; mean depth = 29 cm) as a function of %C saturation across sites, based on our global synthesis of studies that measured C accrual following management interventions ( $n = 103$ ). Nonlinear fit depicts asymptotic regression model with shaded areas representing 10th and 90th quantiles.

**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  $1443 \pm 141$  Pg C and  $3153 \pm 312$  Pg C in topsoils and subsoils, respectively, totaling  $4596$  Pg C  $\pm 453$  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 vegetation<sup>41</sup>, recent studies have indicated that new approaches to deep ploughing<sup>55</sup> or other practices resulting in topsoil-subsoil flipping<sup>56</sup> 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 mineral-organic C capacity and constraining the proportion of total SOC that is mineral-associated (Supplementary Figs. 16 and 17)<sup>17</sup>. Our global estimate of MOC<sub>max</sub> constitutes an effective potential for sorption on minerals, unlike that derived from laboratory DOC sorption experiments alone<sup>57,58</sup>. Laboratory assays tend to underestimate maximum sorption potential, especially if they use native soil which already has some carbon loading on mineral surfaces<sup>57,58</sup>. Moreover, microbial necromass and residues contribute substantially to MOC measured in the field<sup>59</sup>, 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 “4 per 1000” initiative<sup>60,61</sup> and other nature-based climate solutions<sup>62,63</sup>. 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<sup>39</sup>. 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 mineral-organic 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<sup>64,65</sup>.

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 agroforestry) applied globally is on the order of 2–5 Pg CO<sub>2</sub> yr<sup>-1</sup> (0.5–1.4 Pg C yr<sup>-1</sup>)<sup>62,66,67</sup>; implementation potentials given economic considerations fall at the lower end of this range<sup>68</sup>, while nutrient limitations may further limit achievable sequestration<sup>69,70</sup>. While the implementation of soil management interventions across the world’s >500 million active farms poses formidable socioeconomic challenges<sup>71,72</sup>, actionable valuation of soil C accumulation may incentivize restorative soil management in targeted areas<sup>73,74</sup>, with enduring co-benefits for soil productivity<sup>64,65</sup>. Despite the many challenges of such interventions, if the average C accrual rates from our synthesis (Fig. 4d) were achieved globally over all croplands (Supplementary Fig. 22), we estimate that soil C sequestration efforts could store over 1 Pg C yr<sup>-1</sup> given current stocks, on par with the bottom-up estimates of soil C sequestration potential for agricultural lands<sup>62,66,67</sup>. We differ, however, in the duration that such carbon uptake rates could be maintained. Effective C sequestration would push soils towards higher %C saturation (Fig. 4d), thereby lowering their C sequestration efficiency over time.

While many soil C sequestration initiatives are centered on improved management of current croplands<sup>60,61</sup> (which often show more modest C accrual rates and %C saturation; Supplementary Figs. 20–22 and 23b), the potential for soil C sequestration is relevant across nature-based climate solutions more broadly<sup>62,63</sup>. For example, there are growing efforts in quantifying the potential for C sequestration on abandoned pastures and croplands<sup>75–79</sup>, as well as restoring degraded rangelands<sup>80–84</sup>. Though agricultural lands are still expanding in some parts of the world, nearly 90 million hectares of agricultural lands were abandoned between 1985–2005 in North America, Europe, Oceania, and South Asia<sup>85</sup> (and up to 472 million hectares have been abandoned globally over the last few centuries<sup>86</sup>). Many of these lands are in regions with low %C saturation, where our synthesis suggests that revegetation and restoration interventions have the potential to achieve high rates of C accrual (Fig. 4d; Supplementary Figs. 23–25, Supplementary Table 4). Finally, we note that given the considerable departure of most current and abandoned agricultural lands from their respective mineral-organic capacities (Figs. 1, 2, Supplementary Figs. 20–22) (following losses in soil C content<sup>5</sup>), incentives that encourage widespread soil C sequestration efforts are crucial for scaling and achieving mitigation targets.

Our study provides a spatially explicit global estimate of the mineral-organic capacity of the world’s soils to store carbon and an improved understanding of the factors—natural and managed—that drive soils below their capacity. While MOC constitutes the majority (69%) of total SOC in the top meter of non-permafrost mineral soils (totaling 899 Pg C), we find that this represents less than half of the mineral-organic carbon-storage capacity. The understanding of how climate and vegetation affect MOC stocks, and the pronounced undersaturation in managed lands, in particular, suggests that there is significant potential for restoring or enhancing MOC storage. Our estimate of MOC<sub>max</sub> can be refined in future studies as scientists collect more data on additional mineral types, but already provides a data framework to inform ESMs, paleo reconstructions of soils, and policies for soil carbon sequestration. Long-term changes in C inputs may have lasting effects on MOC that are particularly pertinent for land management decisions that seek to sequester carbon<sup>87</sup>. Other soil modification options—e.g., adding more clay to existing sandy soils<sup>88</sup>—to sequester more MOC may also be plausible, but require substantial research to investigate the gains, co-benefits, and potential consequences. Our findings shed light on potential responses and vulnerabilities of MOC to novel natural or managed conditions and emphasize that it is critical to consider the whole system when comparing policies and practices.

### Methods

**Observational synthesis of mineral-associated carbon.** In the present study, we aimed to quantify the maximum capacity of soils to stabilize mineral-associated organic carbon as a function of soil mineralogy, and to resolve the underlying variability due to climate, vegetation, and management. We hypothesized that the maximum mineral-organic capacity (MOC<sub>max</sub>) of soils is a function of the amount (% clay + silt; CS) and type (high- or low-activity; HM or LM, respectively) of mineral, while the achieved mineral-associated carbon (MOC) is additionally a function of climate, vegetation, and management:

$$\text{MOC}_{\text{max}} = f_1(\text{mineral amount, mineral type}) \quad (1)$$

$$\text{MOC} = f_2(\text{MOC}_{\text{max}}, \text{climate, vegetation, management, } \dots) \quad (2)$$

Therefore, we performed an observational synthesis from studies that included MOC (as measured by size and density fractionation, retaining information about the size and density cutoffs), independent particulate organic carbon (POC) and soil organic carbon (SOC) measurements, as well as auxiliary data that included mean annual temperature (MAT), mean annual precipitation (MAP), clay and silt (CS), mineral type and soil order, vegetation type and carbon inputs, total, particulate, and mineral-associated nitrogen, and a variety of mineral-organic soil

properties (e.g., cation exchange capacity, iron, and aluminum content, among others), where available. To locate studies, we searched Google Scholar for combinations of the keywords “soil carbon”, “mineral-associated”, “clay”, “silt”, “size fractionation”, “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  $MOC_{max}$  for each mineral type, and that both methods concluded similar ranges of MOC to SOC ratios<sup>10</sup> (Supplementary Fig. 1). These findings support 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. Specifically, the dataset included: 359 (<2  $\mu\text{m}$ ; clay fraction only) and 1451 (<63  $\mu\text{m}$ ; 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 dithionite- and oxalate-extractable iron ( $Fe_d$  and  $Fe_o$ , respectively), oxalate-extractable aluminum ( $Al_{ox}$ ), 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 natural lands, there was 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 database<sup>11,92</sup> (Supplementary Fig. 2b, d), but our model and results were robust to its exclusion. Furthermore, measurements in tundra and deserts were limited, highlighting the need for studies in these regions (Fig. 1c, Supplementary Figs. 2 and 3). Nevertheless, our observational synthesis spanned diverse climates and soil types: MAT ranged from  $-2.9$  to  $29^\circ\text{C}$ , MAP from  $79$  to  $3806$   $\text{mm yr}^{-1}$ , 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)<sup>93</sup> (Supplementary Fig. 3a, b). We further illustrated the global 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 ecosystems<sup>94</sup> (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 2:1 (one octahedral sheet sandwiched between two tetrahedral sheets) clays. 1:1 clays are prevalent in kaolinitic soils and generally have a lower SSA<sup>14,28,29,95</sup>; therefore, such soils generally exhibit a lower capacity to stabilize carbon and can be characterized as LM. In contrast, 2:1 clays are present in smectitic (e.g., montmorillonitic) and illitic soils and can be characterized as HM, on account of their high SSA<sup>10,14,28,29</sup>. Finally, soils with amorphous, poorly crystalline minerals and mineraloids (e.g., allophane) also depicted high stabilization (described in greater detail below) and were grouped with HM.

Soil orders (USDA taxonomy) represented in this study include Alfisols, Andisols, Aridisols, Entisols, Gelisols, Inceptisols, Mollisols, Oxisols, Spodosols, Ultisols, and Vertisols; we excluded organic soils in wetlands and peat, and, thus, Histosols were not considered. When detailed information on mineral composition (e.g., the dominant presence of 2:1 vs. 1:1 clays) was not explicitly stated in the original studies, the following categorization was used based on the primary mineral composition by soil order<sup>31</sup>: Oxisols and Ultisols at all depths were LM; and Alfisols, Aridisols, Gelisols, Inceptisols, Mollisols, Spodosols, and Vertisols were HM. Andisols and Entisols were HM in the topsoil (0–30 cm) and LM in the subsoil (30–100 cm). We note that, in future studies, allophanic soils with large amounts of amorphous minerals (i.e., Andisols) could be treated as a third category due to their propensity to form stable Al- and Fe-complexes and accumulate SOC<sup>14,32</sup>; however, we did not observe a significant difference in their MOC as compared to HM soils in our synthesis. Furthermore, there are complexities, including the amount and type of metal oxides and hydroxides, that can make the distinction between low- and high-activity minerals difficult in certain soils<sup>29,96</sup> (and can also complicate the comparison of soil clays and “reference” or synthetic clays<sup>97</sup>), the classification of LM and HM soils has been widely used and supported<sup>23,28,29</sup> and emerges from our data globally (Fig. 1 and Supplementary Figs. 7 and 8). Indeed, the MOC patterns were consistent within our classifications

and, thus, these categories were deemed sufficient for the scope of this global-scale study.

We retained the classification of vegetation type, and all details on management practices or lack thereof, reported in the original studies and assigned one of eight land cover types (Fig. 1b–c). Land cover types included boreal forest, temperate forest, tropical forest, grassland, savanna, shrubland, and cropland (Supplementary Fig. 14). To further compare managed and natural lands, we grouped sites into broad categories of forests, grasslands (including savannas), and croplands. We observed stark differences between natural or less-managed ecosystems (namely, forest and grassland) and more intensively managed (crop) ecosystems (Fig. 2). We calculated the %C saturation of soils (as depicted in Supplementary Fig. 10) to compare ecosystems and management practices.

**Boundary line analyses.** To estimate the mineralogical carbon capacity ( $MOC_{max}$ ) as a function of soil mineralogy, we conducted a boundary line analysis across the full range of clay and silt (CS) for each mineral category (LM and HM, see section “Defining and evaluating ecosystem categories”; Supplementary Fig. 26). Specifically, we explored the relationship of MOC (y-axis;  $\text{g C kg}^{-1}$  soil) as a function of CS (x-axis; %) and fit the 95th quantile for natural ecosystems of each mineral type to obtain maximum boundary-line slopes for HM and LM (Fig. 1; quantile sensitivity analysis in Supplementary Fig. 7). We chose the 95th quantile as a conservative approximation of this boundary line, given existing experimental uncertainty of individual observations (i.e., total % recovery of C from size fractionations ranged from  $\sim 80$  to  $120\%$  among studies). We used MOC observations from clay and silt (<63  $\mu\text{m}$ ) size classes to derive  $MOC_{max}$  (Fig. 1) and note that our results were robust to the choice of particle size (e.g., <20  $\mu\text{m}$  or <63  $\mu\text{m}$ ; Supplementary Fig. 5). The 63  $\mu\text{m}$  silt size threshold chosen is also that used in global datasets and mapped products of soil clay and silt<sup>4,98</sup>. We also note that our estimates of  $MOC_{max}$  for HM and LM were robust across topsoils and subsoils. Both topsoils and subsoils approached  $MOC_{max}$  (Fig. 2b), and thus, a separate  $MOC_{max}$  for each depth was not theoretically or empirically warranted. While it is more difficult to reach  $MOC_{max}$  in subsoils due to lower carbon inputs,  $MOC_{max}$  is an intrinsic property of the soil mineralogy and constitutes a theoretical upper limit that is independent of depth; indeed, this depth-independence is important for capturing MOC stocks and ages in process-based soil carbon models<sup>99,100</sup>. Our estimates of  $MOC_{max}$  for HM and LM were also robust across vegetation types (Supplementary Fig. 8), which corroborates studies showing that microbial processing and reactive minerals act as effective filters that lessen chemical differences in organic matter inputs<sup>59,101</sup>. Theory and models further support this finding and suggest that, while litter quality may affect the approach of a given soil to its mineralogical capacity, the value of  $MOC_{max}$  is independent of vegetation type<sup>28,102</sup>. 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 minerals<sup>92</sup>, further exploration is needed to understand the relative contribution of fine POC and DOC in such soils<sup>16</sup>.

**Statistical modeling and predictive relationships.** Due to potential nonlinear relationships between MOC and environmental variables, we used an ensemble machine learning method—namely, Random Forest (RF)—to identify key 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 analyses<sup>103</sup>. We used the RandomForest package in R for RF analyses<sup>103,104</sup>. 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 performance<sup>105</sup>—mean absolute error (MAE) and mean-squared error (MSE) to quantify the model error and  $R^2$  to estimate the proportion of variance in MOC explained by the model (Supplementary Fig. 11). Our aim was to: (i) elucidate key variables by assessing their relative contribution to the MSE 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 data<sup>103</sup>, as implemented herein. 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. 11;  $R^2 = 0.60 \pm 0.06$ ). 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)<sup>106</sup>. That is, for a variable  $x_i$  and complement set of all other variables  $x_C$ , the general model function is  $f(x_i, x_C)$  and depends on all input variables. The partial dependence function of  $x_i$ , denoted  $\hat{f}_{x_i}(x_i)$ , then describes the marginal effect of  $x_i$  on the prediction, and is

estimated as:

$$\hat{f}_{x_s}(x_s) = \frac{1}{n} \sum_{i=1}^n f(x_s, x_c^{(i)}) \quad (3)$$

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 that corresponded to different saturation regimes—unsaturated, medium, and saturated (Fig. 3a). The breakpoints for these three regimes were selected as the median and third quartile of the POC distribution (namely, 4 and 10 g C kg<sup>-1</sup> soil, respectively), which also correspond to approximate saturation thresholds in the partial dependence plot of POC (Fig. 3a).

Our approach permits us to disentangle the role of each controlling variable 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 mineral type (Supplementary Fig. 26). We note that a global map of POC 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 75–25% train-test split; 400 decision trees each) with bootstrapped sampling to rigorously assess model performance<sup>105</sup>. For these RF models that included SOC, a comparison of predicted and observed MOC on the independent test datasets across the ensemble yielded an  $R^2 = 0.79 \pm 0.05$  (mean  $\pm$  s.d.; Supplementary Fig. 13). We note that, while the predictability of SOC is often lower across large global datasets<sup>4,107</sup>, the predictability of MOC is generally high given readily measured SOC, climate, and edaphic variables (e.g., see refs. <sup>11,92</sup> on a regional scale). Indeed, the  $R^2$  was consistently high across the ensemble cross-validation (Supplementary Fig. 13b). We also performed a spatially buffered leave-one-out cross-validation<sup>108</sup> to ensure that spatial auto-correlation did not significantly compromise the RF model. Namely, for each leave-one-out test data point, all data within a given buffer radius (0–150 km) were excluded from the training set; i.e., only data points outside the buffer zone were used to train each RF model.  $R^2$  values did not vary with buffer radius (regression slope not statistically different from zero;  $p = 0.81$ ) and were in agreement with the ensemble cross-validation approach (Supplementary Fig. 13c). Furthermore, we also used a hold-out (or block) cross-validation<sup>109</sup> on the basis of soil order to assess extrapolation performance and confirm the robustness of our RF estimates across soil types—i.e., all soil profiles from a given soil order were completely withheld from the training dataset and then used to independently test the resulting RF model. Model performance was consistently high across the out-of-sample test datasets ( $R^2 = 0.75 \pm 0.11$ ; mean  $\pm$  s.d. across soil orders); only Andisols (which span ~1% of the ice-free land surface) were more difficult to predict when completely withheld from the training dataset ( $R^2 = 0.40$ ), and we encourage more measurements in these soils (see section “Defining and evaluating ecosystem categories”). Hold-out cross-validation on the basis of individual studies also confirmed the robustness of our RF estimates to additional data (e.g., our results were insensitive to adding the independent LUCAS dataset<sup>92</sup>; Supplementary Fig. 2). We rigorously demonstrate the RF model performance and provide uncertainty ranges for the 90% prediction intervals, derived from the 5th and 95th quantiles (Supplementary Fig. 13a).

The climatic and edaphic driver variables used for the global predictions were all re-gridded to  $0.5^\circ \times 0.5^\circ$  resolution (Supplementary Fig. 14). Mean annual temperature (MAT) was estimated from the CRU data set (version 3.10)<sup>110</sup>, and mean annual precipitation (MAP) from the GPCC dataset<sup>111</sup>, both as 30-year annual averages. Clay and silt (CS) content was obtained from the Harmonized World Soil Database (HWSD)<sup>98</sup>. Soil organic carbon (SOC) was obtained from both the HWSD and SoilGrids<sup>4,98</sup>, and also as the average of the two gridded data products<sup>112</sup> (Supplementary Fig. 19 and Supplementary Table 2). Land cover (vegetation type) was obtained from the MODIS MCD12C1 product<sup>113</sup>, and the 16 categories were combined into 10 for consistency with those reported in the observations (Fig. 1b, c and Supplementary Fig. 14 for details). The primary mineral type was calculated from global estimates of mineral composition<sup>31</sup> (Supplementary Fig. 14; see section “Estimating the global soil mineralogical carbon capacity”). We focused our analysis on mineral soils, and thus excluded grid cells containing a majority of organic soils (i.e., >50% Histosols and Gelisols<sup>114,115</sup>, Tundra and deserts were also excluded, due to data limitations in these regions

(Supplementary Figs. 2 and 14). We used two depth intervals for SOC, CS, and primary mineral type, namely topsoil (0–30 cm) and subsoil (30–100 cm).

The predictive RF model ( $R^2 = 0.79 \pm 0.05$ ; Supplementary Fig. 13), trained on the observational data, was then used to predict MOC globally for each depth interval to 1 m (Supplementary Fig. 15; Supplementary Tables 1 and 2 summarize results for all SOC data products). We include MOC uncertainty ranges for the 90% prediction intervals (5th and 95th quantiles) globally (Supplementary Fig. 15; Supplementary Table 2). Interestingly, we note that the distributions of MOC and SOC (and %C saturation) from our globally distributed observational dataset most closely agreed with SOC and predicted MOC (and %C saturation) from the averaged data product (Supplementary Figs. 18 and 19; Supplementary Table 2), whereas the distribution of SOC (and %C saturation) from SoilGrids was higher and HWSD was substantially lower (Supplementary Fig. 19; Supplementary Table 2). We focus here on reporting results using the averaged SOC data product, as a conservative estimate of MOC stocks given the uncertainty and range between soil carbon data products, and note that more accurate estimates and consensus on current soil carbon stocks are critical for improved predictions<sup>116</sup>.

Leveraging our global MOC predictions, we calculated the proportion of SOC that is mineral-associated globally (Supplementary Fig. 16). Our global estimates of MOC/SOC ratios agree with those observed in our observational synthesis and confirm the dominant contribution of MOC to soil carbon at the global scale (Supplementary Figs. 1 and 16). Furthermore, our predictions suggest that high-latitudes have a lower fraction of MOC/SOC globally, and thus a higher fraction of particulate organic carbon (POC), but data limitations of mineral-associated carbon measurements at high-latitudes warrant further exploration of these patterns.

**Estimating the global soil mineralogical carbon capacity.** To estimate the spatially explicit mineralogical carbon capacity ( $MOC_{max}$ ) globally, we used global data products of clay and silt content (CS), soil order and mineral composition, and bulk density (BD), together with our derived relationship for  $MOC_{max}$  (Eq. 1; Fig. 1). Global maps of % clay, % silt, and BD were obtained from the HWSD<sup>98</sup>. A map of soil mineral composition<sup>31</sup> was used to generate a corresponding map of primary mineral type (Supplementary Fig. 14; see section “Defining and evaluating ecosystem categories”). Our focus here was on mineral soils, and thus, we excluded organic soils (grid cells with >50% Histosols and Gelisols) from the estimation of global MOC storage potential. Tundra and deserts were also excluded due to data limitations (Supplementary Figs. 2 and 14). We calculated CS content and applied our  $MOC_{max}$  relationship (Fig. 1; slope  $\pm$  90% confidence intervals) for each mineral type (HM and LM) globally to obtain  $MOC_{max}$  (g C kg<sup>-1</sup> soil) at  $0.5^\circ \times 0.5^\circ$  resolution. For a global total (in Pg C;  $1 \text{ Pg} = 10^{15} \text{ g}$ ), we used BD to calculate  $MOC_{max}$  stocks (in kg C m<sup>-2</sup> soil) and then summed over all grid cells containing mineral soils (Supplementary Tables 1, 2).

The mineralogical capacity  $MOC_{max}$  (Supplementary Fig. 17) was then compared to predicted MOC stocks (Supplementary Fig. 15) to calculate the mineralogical C deficit (by subtracting MOC from  $MOC_{max}$ ) and %C saturation (by dividing MOC/ $MOC_{max}$ ) 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 ( $MOC_{max}$ ), we also estimated the C deficit relative to an environmental limit calculated using the natural land average %C saturation (i.e., average MOC/ $MOC_{max}$ ) observed in our MOC synthesis at each depth (Fig. 2; i.e., 51 and 19% for topsoil and subsoil, respectively). These results are summarized in Supplementary Table 1. We note that our estimates of C deficit and %C saturation using SoilGrids are the most conservative for potential accrual, given their higher current SOC (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 difficult in some regions than in others (Supplementary Fig. 12), this estimate constrains  $MOC_{max}$  for a given soil mineralogy and provides a starting point for areas to focus on management (Supplementary Discussion). Furthermore, while deep soils have the largest C deficit, surface soils in lands already dominated by management may be the most cost-effective areas on which to focus.

**Observational synthesis and analysis of carbon accrual.** Here we aimed to quantify the ability of topsoils ranging in mineralogical carbon saturation to further accrue 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 edaphic properties; namely, clay and silt content, texture class or soil series to estimate these contents and soil classification or other information that would allow us to assign the soil to a high- or low-activity mineral type.



For each entry (a pair of initial and final carbon stocks), we calculated the maximum mineralogical capacity ( $MOC_{max}$ ) using the data-derived relationships (Fig. 1) with an amount (% clay + silt; CS) and type (high- or low-activity; HM or LM, respectively) of mineral in each soil layer and used respective bulk densities to calculate the layer  $MOC_{max}$  stocks ( $t\ C\ ha^{-1}$ ) which were then summed across the considered depth to obtain the profile  $MOC_{max}$  stock ( $t\ C\ ha^{-1}$ ). 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 SOC stocks were standardized to equal depths using the maximum 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 MOC stock divided by the initial  $MOC_{max}$  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 ( $R^2 = 0.70 \pm 0.06$ ; using Scikit-Learn in Python, see section “Global mineral-associated carbon predictions and uncertainty”).

Carbon accrual ( $t\ C\ ha^{-1}\ y^{-1}$ ) 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 sites of 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), doubled litter or wood input from the DIRT (Detrital Input and Removal Experiment) network, and natural revegetation and mine reclamation (formerly degraded lands or newly exposed land surfaces as a result of mining, landsliding, alluviation, or marine terrace uplift, where soil C accrual can be expected as a consequence of natural or human-accelerated revegetation and soil formation). We note that most studies did not explicitly quantify C inputs over time, especially those where changes in C inputs stemmed from downstream impacts on vegetation, and we strongly urge the reporting of C input estimates in future studies. After the initial compilation, profiles were excluded in which the studied soil depth was lower than 15 cm ( $n = 2$ ) and in which organic amendments were applied annually, in an initial one-time application higher than  $20\ t\ ha^{-1}$  without a control plot or at unknown intensity ( $n = 11$ ).

Our accrual synthesis included 103 observations from 34 studies, with depths ranging between 15 and 40 cm (mean = 29 cm). The synthesis included soils with both high-activity minerals ( $n = 84$ ), especially Entisols ( $n = 41$ ) and Alfisols ( $n = 20$ ), and low-activity minerals ( $n = 19$ ), belonging to Oxisols ( $n = 10$ ) and Ultisols ( $n = 9$ ). There was a greater representation of temperate climates, highlighting the need for additional studies in poorly represented regions including tropical and boreal climates in future work. However, globally, our study suggests that the greatest potential for accrual (i.e., lowest %C saturation) often occurs in these temperate regions (Fig. 4b) and was thus a focus of our work.

While most soils in the C accrual synthesis fell below 100% C saturation, we note that  $MOC_{max}$  was derived as a conservative estimate from the 95th quantile of our MOC synthesis (Fig. 1; see section “Boundary line analyses”) and therefore, by definition, soils can fall near or above this maximum. Indeed, in our MOC synthesis, ~200% C saturation values were observed for soils with low values of MOC and  $MOC_{max}$ . Of the 103 accrual measurements, one sandy soil (i.e., low MOC and  $MOC_{max}$ ) displayed seemingly high %C saturation, yet still within expected ranges (Fig. 4d). Furthermore, this soil was categorized as an Andisol, which could be treated as a third category in future studies (see section “Defining and evaluating ecosystem categories”), due to its propensity to form stable Al- and Fe-complexes and accumulate carbon<sup>14</sup>. Nevertheless, removing this individual measurement, or even all Andisols ( $n = 6$ ) in the C accrual synthesis, did not change the fitted model.

Furthermore, we note that the reclaimed mine soils in our dataset constituted two major categories of abandonment and regeneration—those with and without topsoil application (Supplementary Fig. 23b). Mine reclamation that employs the application of topsoil (usually a 30 cm layer, originally excavated before mining and stored in stockpiles;  $n = 31$ ) is, in terms of management procedures, comparable to the conversion of cropland (disturbed by tillage and/or erosion) to grassland or forest. Reclamation can involve some mechanical preparation of soil, application of fertilizer or other organic amendments, and planting vegetation. This reapplied topsoil is sampled as a baseline and subsequent sampling in the following years is used to monitor C accrual. In reclamation without topsoil application ( $n = 13$ ), vegetation either spontaneously establishes itself or is planted directly into the parent material. Such conditions represent early stages of soil formation and are parallel to the restoration of soils heavily impacted by erosion that has led to the complete loss of topsoil with only the infertile parent material left behind. While mine soils are well-described systems suitable for the study of soil management,

included in previous syntheses of soil carbon dynamics after land-use change<sup>117,118</sup>, the fitted model and observed negative trend in C accrual with %C saturation were robust to their exclusion (Supplementary Fig. 23c).

Statistical analyses were conducted using the ‘nlms’ package in R to fit a nonlinear asymptotic relationship between carbon accrual rates and the mineralogical percent saturation of the respective soil profiles. Specifically, we used a three-parameter self-starting asymptotic regression model (SSAsymp) on the 103 rate measurements ( $R^2 = 0.18$ ; 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 ( $R^2 = 0.18$ ) than by initial SOC ( $R^2 = 0.04$ –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>).

## Code availability

All code relating to this study is available from the corresponding author upon request, and is archived at Zenodo (<https://doi.org/10.5281/zenodo.6539765>).

Received: 20 January 2022; Accepted: 9 June 2022;

Published online: 01 July 2022

## References

1. Tiessen, H., 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**, e0169748 (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 CO<sub>2</sub> 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**, 307–309 (2017).
9. Rumpel, C. et al. Put more carbon in soils to meet paris climate pledges. *Nature* **564**, 32–34 (2018).
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).
11. Cotrufo, M. F., Ranalli, M. G., Haddix, M. L., Six, J. & Lugato, E. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nat. Geosci.* **12**, 989–994 (2019).
12. Hemingway, J. D. et al. Mineral protection regulates long-term global preservation of natural organic carbon. *Nature* **570**, 228–231 (2019).
13. Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Deneff, K. & Paul, E. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant

- litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Glob. Chang. Biol.* **19**, 988–995 (2013).
14. Kleber, M. et al. Mineral-organic associations: formation, properties, and relevance in soil environments. *Adv. Agron.* **130**, 1–140 (2015).
  15. Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M. & Hendricks, D. M. Mineral control of soil organic carbon storage and turnover. *Nature* **360**, 3601–3603 (1997).
  16. Lavallee, J. M., Soong, J. L. & Cotrufo, M. F. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. *Glob. Chang. Biol.* **26**, 261–273 (2020).
  17. Wieder, W. R., Sulman, B. N., Hartman, M. D., Koven, C. D. & Bradford, M. A. Arctic soil governs whether climate change drives global losses or gains in soil carbon. *Geophys. Res. Lett.* **46**, 14486–14495 (2019).
  18. Wieder, W. R., Allison, S. D., Davidson, E. A., Georgiou, K. & Hararuk, O. Explicitly representing soil microbial processes in Earth system models. *Glob. Biogeochem. Cycles* **29**, 1782–1800 (2015).
  19. Kleber, M., Sollins, P. & Sutton, R. A conceptual model of organo-mineral interactions in soils: self-assembly of organic molecular fragments into zonal structures on mineral surfaces. *Biogeochemistry* **85**, 9–24 (2007).
  20. Sulman, B. N. et al. Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry* **141**, 109–123 (2018).
  21. Woolf, D. & Lehmann, J. Microbial models with minimal mineral protection can explain long-term soil organic carbon persistence. *Sci. Rep.* **9**, 1–8 (2019).
  22. Kaiser, K. & Guggenberger, G. Mineral surfaces and soil organic matter. *Eur. J. Soil Sci.* **54**, 219–236 (2003).
  23. Six, J., Conant, R. T., Paul, E. A. & Paustian, K. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant Soil* **241**, 155–176 (2002).
  24. Hassink, J. The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant Soil* **191**, 77–87 (1997).
  25. Doetterl, S. et al. Soil carbon storage controlled by interactions between geochemistry and climate. *Nat. Geosci.* **8**, 780–783 (2015).
  26. Rasmussen, C. et al. Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* **137**, 297–306 (2018).
  27. Feng, W., Plante, A. F., Aufdenkampe, A. K. & Six, J. Soil organic matter stability in organo-mineral complexes as a function of increasing C loading. *Soil Biol. Biochem.* **69**, 398–405 (2014).
  28. Feng, W., Plante, A. F. & Six, J. Improving estimates of maximal organic carbon stabilization by fine soil particles. *Biogeochemistry* **112**, 81–93 (2013).
  29. Singh, M. et al. Stabilization of soil organic carbon as influenced by clay mineralogy. *Adv. Agronomy* **148**, 33–84 (2018).
  30. Kleber, M. et al. Dynamic interactions at the mineral–organic matter interface. *Nat. Rev. Earth Environ.* **2**, 402–421 (2021).
  31. Ito, A. & Wagai, R. Global distribution of clay-size minerals on land surface for biogeochemical and climatological studies. *Sci. Data* **4**, 1–11 (2017).
  32. Beare, M. H. et al. Estimating the organic carbon stabilisation capacity and saturation deficit of soils: a New Zealand case study. *Biogeochemistry* **120**, 71–87 (2014).
  33. McNally, S. R. et al. Soil carbon sequestration potential of permanent pasture and continuous cropping soils in New Zealand. *Glob. Chang. Biol.* **23**, 4544–4555 (2017).
  34. Chen, S. et al. Fine resolution map of top- and subsoil carbon sequestration potential in France. *Sci. Total Environ.* **630**, 389–400 (2018).
  35. Angers, D. A., Arrouays, D., Saby, N. P. A. & Walter, C. Estimating and mapping the carbon saturation deficit of French agricultural topsoils. *Soil Use Manag.* **27**, 448–452 (2011).
  36. Wiesmeier, M. et al. Carbon sequestration potential of soils in southeast Germany derived from stable soil organic carbon saturation. *Glob. Chang. Biol.* **20**, 653–665 (2014).
  37. Wiesmeier, M. et al. Carbon storage capacity of semi-arid grassland soils and sequestration potentials in northern China. *Glob. Chang. Biol.* **21**, 3836–3845 (2015).
  38. Stewart, C. E., Paustian, K., Conant, R. T., Plante, A. F. & Six, J. Soil carbon saturation: Implications for measurable carbon pool dynamics in long-term incubations. *Soil Biol. Biochem.* **41**, 357–366 (2009).
  39. West, T. O. & Six, J. Considering the influence of sequestration duration and carbon saturation on estimates of soil carbon capacity. *Clim. Change* **80**, 25–41 (2007).
  40. Balesdent, J. et al. Atmosphere–soil carbon transfer as a function of soil depth. *Nature* **559**, 599–602 (2018).
  41. Lorenz, K. & Lal, R. The depth distribution of soil organic carbon in relation to land use and management and the potential of carbon sequestration in subsoil horizons. *Adv. Agron.* **88**, 35–66 (2005).
  42. Jackson, R. B. et al. The ecology of soil carbon: pools, vulnerabilities, and biotic and abiotic controls. *Annu. Rev. Ecol. Syst.* **48**, 419–445 (2017).
  43. Sokol, N. W. & Bradford, M. A. Microbial formation of stable soil carbon is more efficient from belowground than aboveground input. *Nat. Geosci.* **12**, 46–53 (2019).
  44. Villarino, S. H., Pinto, P., Jackson, R. B. & Piñeiro, G. Plant rhizodeposition: a key factor for soil organic matter formation in stable fractions. *Sci. Adv.* **7**, 1–14 (2021).
  45. Jackson, R. B. et al. A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**, 389–411 (1996).
  46. Balesdent, J. et al. Renouveau du carbone profond des sols cultivés: une estimation par compilation de données isotopiques. *Biotechnol. Agron. Soc. Environ.* **21**, 1–10 (2017).
  47. Kramer, M. G. & Chadwick, O. A. Climate-driven thresholds in reactive mineral retention of soil carbon at the global scale. *Nat. Clim. Chang.* **8**, 1104–1108 (2018).
  48. Mayer, L. M. & Xing, B. Organic Matter-Surface Area Relationships in Acid Soils. *Soil Sci. Soc. Am. J.* **65**, 250–258 (2001).
  49. Keil, R. G. & Mayer, L. M. *Mineral Matrices and Organic Matter. Treatise on Geochemistry.* 2nd edn. (Elsevier Ltd., 2013).
  50. Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).
  51. Kleber, M. et al. Old and stable soil organic matter is not necessarily chemically recalcitrant: Implications for modeling concepts and temperature sensitivity. *Glob. Chang. Biol.* **17**, 1097–1107 (2011).
  52. Rodrigues, C. C. & Moraes, D. J. Control of the emission of ammonia through adsorption in a fixed bed of activated carbon. *Adsorpt. Sci. Technol.* **20**, 1013–1022 (2002).
  53. Chenu, C. et al. Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil Tillage Res.* **188**, 41–52 (2019).
  54. Stuble, K. L. et al. Long-term impacts of warming drive decomposition and accelerate the turnover of labile, not recalcitrant, carbon. *Ecosphere* **10**, e02715 (2019).
  55. Alcántara, V., Don, A., Well, R. & Nieder, R. Deep ploughing increases agricultural soil organic matter stocks. *Glob. Chang. Biol.* **22**, 2939–2956 (2016).
  56. Schiedung, M., Tregurtha, C. S., Beare, M. H., Thomas, S. M. & Don, A. Deep soil flipping increases carbon stocks of New Zealand grasslands. *Glob. Chang. Biol.* **25**, 2296–2309 (2019).
  57. Mayes, M. A., Heal, K. R., Brandt, C. C., Phillips, J. R. & Jardine, P. M. Relation between Soil Order and Sorption of Dissolved Organic Carbon in Temperate Subsoils. *Soil Sci. Soc. Am. J.* **76**, 1027–1037 (2012).
  58. Abramoff, R. Z. et al. How much carbon can be added to soil by sorption? *Biogeochem. Lett. under Rev.* <https://doi.org/10.1007/s10533-021-00759-x> (2021).
  59. Kallenbach, C. M., Frey, S. D. & Grandy, A. S. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nat. Commun.* **7**, 1–10 (2016).
  60. Minasny, B. et al. Soil carbon 4 per mille. *Geoderma* **292**, 59–86 (2017).
  61. Amelung, W. et al. Towards a global-scale soil climate mitigation strategy. *Nat. Commun.* **11**, 1–10 (2020).
  62. Bossio, D. A. et al. The role of soil carbon in natural climate solutions. *Nat. Sustain* **3**, 391–398 (2020).
  63. Nolan, C. J., Field, C. B. & Mach, K. J. Constraints and enablers for increasing carbon storage in the terrestrial biosphere. *Nat. Rev. Earth Environ.* **2**, 436–446 (2021).
  64. Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **304**, 1623–1627 (2004).
  65. Lehmann, J., Bossio, D. A., Kögel-Knabner, I. & Rillig, M. C. The concept and future prospects of soil health. *Nat. Rev. Earth Environ.* **1**, 544–553 (2020).
  66. Fuss, S. et al. Negative emissions - Part 2: Costs, potentials and side effects. *Environ. Res. Lett.* **13**, 063002 (2018).
  67. Lal, R. Digging deeper: a holistic perspective of factors affecting soil organic carbon sequestration in agroecosystems. *Glob. Chang. Biol.* **24**, 3285–3301 (2018).
  68. Smith, P. Soil carbon sequestration and biochar as negative emission technologies. *Glob. Chang. Biol.* **22**, 1315–1324 (2016).
  69. Spohn, M. Increasing the organic carbon stocks in mineral soils sequesters large amounts of phosphorus. *Glob. Chang. Biol.* **26**, 4169–4177 (2020).
  70. Davies, C. A., Robertson, A. D. & McNamara, N. P. The importance of nitrogen for net carbon sequestration when considering natural climate solutions. *Glob. Chang. Biol.* 218–219. <https://doi.org/10.1111/gcb.15381> (2020).
  71. Amundson, R. & Biardeau, L. Soil carbon sequestration is an elusive climate mitigation tool. *Proc. Natl Acad. Sci. USA* **115**, 11652–11656 (2018).
  72. Schlesinger, W. H. & Amundson, R. Managing for soil carbon sequestration: let's get realistic. *Glob. Chang. Biol.* **25**, 386–389 (2019).
  73. Six, J. Spare our restored soil. *Nature* **498**, 180–181 (2013).
  74. Searchinger, T. D., Wiersma, S., Beringer, T. & Dumas, P. Assessing the efficiency of changes in land use for mitigating climate change. *Nature* **564**, 249–253 (2018).
  75. Bell, S. M., Terrer, C., Barriocanal, C., Jackson, R. B. & Rosell-Melé, A. Soil organic carbon accumulation rates on Mediterranean abandoned agricultural lands. *Sci. Total Environ.* **759**, 143535 (2021).

76. Bell, S., Barriocanal, C., Terrer, C. & Rosell-Melé, A. Management opportunities for soil carbon sequestration following agricultural land abandonment. *Environ. Sci. Policy* **108**, 104–111 (2020).
77. Potter, K. N., Torbert, H. A., Johnson, H. B. & Tischler, C. R. Carbon storage after long-term grass establishment on degraded soils. *Soil Sci.* **164**, 718–725 (1999).
78. Burke, I. C., Lauenroth, W. K. & Coffin, D. P. Soil organic matter recovery in semiarid grasslands: implications for the conservation reserve program. *Ecol. Appl.* **5**, 793–801 (1995).
79. Cramer, V. A., Hobbs, R. J. & Standish, R. J. What's new about old fields? Land abandonment and ecosystem assembly. *Trends Ecol. Evol.* **23**, 104–112 (2008).
80. Schuman, G. E., Janzen, H. H. & Herrick, J. E. Soil carbon dynamics and potential carbon sequestration by rangelands. *Environ. Pollut.* **116**, 391–396 (2002).
81. Ryals, R., Kaiser, M., Torn, M. S., Berhe, A. A. & Silver, W. L. Impacts of organic matter amendments on carbon and nitrogen dynamics in grassland soils. *Soil Biol. Biochem.* **68**, 52–61 (2014).
82. Fynn, A. et al. Soil carbon sequestration in United States rangelands. *Integr. Crop Manag.* **11**, 57–104 (2010).
83. Sanderman, J. et al. Carbon sequestration under subtropical perennial pastures II: Carbon dynamics. *Soil Res.* **51**, 771–780 (2013).
84. Bardgett, R. D. et al. Combating global grassland degradation. *Nat. Rev. Earth Environ.* **2**, 720–735 (2021).
85. Foley, J. A. et al. Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
86. Campbell, J. E., Lobell, D. B., Genova, R. C. & Field, C. B. The global potential of bioenergy on abandoned agriculture lands. *Environ. Sci. Technol.* **42**, 5791–5794 (2008).
87. Conant, R. T., Cerri, C. E. P., Osborne, B. B. & Paustian, K. Grassland management impacts on soil carbon stocks: a new synthesis. *Ecol. Appl.* **27**, 662–668 (2017).
88. Schapel, A. & Marschner, P. & Churchman, J. Clay amount and distribution influence organic carbon content in sand with subsoil clay addition. *Soil Tillage Res.* **184**, 253–260 (2018).
89. Dalal, R. C. & Mayer, R. J. Long-term trends in fertility of soils under continuous cultivation and cereal cropping in southern Queensland. IV. Loss of organic carbon from different density functions. *Aust. J. Soil Res.* **24**, 301–309 (1986).
90. Dalal, R. C. & Mayer, R. J. Long-term trends in fertility of soils under continuous cultivation and cereal cropping in southern Queensland. III: Distribution and kinetics of soil organic carbon in particle-size fractions. *Aust. J. Soil Res.* **24**, 293–300 (1986).
91. Lawrence, C. R. et al. An open-source database for the synthesis of soil radiocarbon data: International Soil Radiocarbon Database (ISRaD) version 1.0. *Earth Syst. Sci. Data* **12**, 61–76 (2020).
92. Lugato, E., Lavalée, J. M., Haddix, M. L., Panagos, P. & Cotrufo, M. F. Different climate sensitivity of particulate and mineral-associated soil organic matter. *Nat. Geosci.* **14**, 295–300 (2021).
93. Batjes, N. H. et al. WoSIS: providing standardised soil profile data for the world. *Earth Syst. Sci. Data* **9**, 1–14 (2017).
94. Hartley, I. P., Hill, T. C., Chadburn, S. E. & Hugelius, G. Temperature effects on carbon storage are controlled by soil stabilisation capacities. *Nat. Commun.* **12**, 1–7 (2021).
95. Heckman, K. & Rasmussen, C. Role of mineralogy and climate in the soil carbon cycle. *Clim. Chang. Impacts Soil Process. Ecosyst. Prop. Chapter 4*, 93–110. <https://doi.org/10.1016/b978-0-444-63865-6.00004-1> (2018).
96. de Oliveira, J. S. et al. Soil properties governing phosphorus adsorption in soils of Southern Brazil. *Geoderma Reg.* **22**, e00318 (2020).
97. Churchman, G. J. Is the geological concept of clay minerals appropriate for soil science? A literature-based and philosophical analysis. *Phys. Chem. Earth* **35**, 927–940 (2010).
98. Wieder, W. R., Boehmert, J., Bonan, G. B. & Langseth, M. Regridded Harmonized World Soil Database v1.2. *Data set. Oak Ridge Natl. Lab. Distrib. Act. Arch. Center, Oak Ridge, Tennessee, USA.* <https://doi.org/10.3334/ORNLDAAC/1247> (2014).
99. Ahrens, B. et al. Combination of energy limitation and sorption capacity explains 14 C depth gradients. *Soil Biol. Biochem.* **148**, 107912 (2020).
100. Ahrens, B., Braakhekke, M. C., Guggenberger, G., Schrupf, M. & Reichstein, M. Contribution of sorption, DOC transport and microbial interactions to the <sup>14</sup>C age of a soil organic carbon profile: insights from a calibrated process model. *Soil Biol. Biochem.* **88**, 390–402 (2015).
101. Mikutta, R. et al. Microbial and abiotic controls on mineral-associated organic matter in soil profiles along an ecosystem gradient. *Sci. Rep.* **9**, 1–9 (2019).
102. Castellano, M. J., Mueller, K. E., Olk, D. C., Sawyer, J. E. & Six, J. Integrating plant litter quality, soil organic matter stabilization, and the carbon saturation concept. *Glob. Chang. Biol.* **21**, 3200–3209 (2015).
103. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
104. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R. N.* **2**(3), 18–22 (2002).
105. Kirchner, J. W. et al. Streamflow response to forest management. *Nature* **578**, E12–E15 (2020).
106. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
107. Hengl, T. et al. SoilGrids1km - Global soil information based on automated mapping. *PLoS ONE* **9**, e114788 (2014).
108. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 1–11 (2020).
109. Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).
110. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
111. Schneider, U. et al. GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. [https://doi.org/10.5676/DWD\\_GPCC/FD\\_M\\_V6\\_050](https://doi.org/10.5676/DWD_GPCC/FD_M_V6_050) (2011).
112. Shi, Z. et al. The age distribution of global soil carbon inferred from radiocarbon measurements. *Nat. Geosci.* **13**, 555–559 (2020).
113. Friedl, M. A. et al. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **114**, 168–182 (2010).
114. Hugelius, G. et al. The northern circumpolar soil carbon database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **5**, 3–13 (2013).
115. Koven, C. D., Hugelius, G., Lawrence, D. M. & Wieder, W. R. Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nat. Clim. Chang.* **7**, 817–822 (2017).
116. Tifafi, M., Guenet, B. & Hatté, C. Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on SoilGrids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data From USA, England, Wales, and France. *Glob. Biogeochem. Cycles* **32**, 42–56 (2018).
117. Post, W. M. & Kwon, K. C. Soil carbon sequestration and land-use change: Processes and potential. *Glob. Chang. Biol.* **6**, 317–327 (2000).
118. Vesterdal, L., Clarke, N., Sigurdsson, B. D. & Gundersen, P. Do tree species influence soil carbon stocks in temperate and boreal forests? *Ecol. Manag.* **309**, 4–18 (2013).
119. Whittaker, R. *Communities and Ecosystems*. (MacMillan, New York, 1975).

## Acknowledgements

K.G. was supported as a Lawrence Fellow at Lawrence Livermore National Laboratory (LLNL) and by a U.S. Department of Agriculture NIFA Postdoctoral Fellowship (Grant No. 2018-67012-27982). Work at LLNL was conducted under the auspices of DOE Contract DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project No. 21-ERD-045. O.V. and R.Z.A. were supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 793485 (O.V.) and No. 834169 (R.Z.A.). R.Z.A. was also supported by the US DOE, Office of Science, Office of Biological and Environmental Research at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the US DOE under contract DE-AC05-00OR22725. M.S.T. and W.J.R. were supported by the US DOE, Office of Science, Office of Biological and Environmental Research under contract number DE-AC02-05CH11231. We thank Jan Frouz, Kate Lajtha, Derek Pierson, István Fekete, and Zsolt Kotroczó for helpful comments regarding the carbon accrual synthesis.

## Author contributions

K.G. conceived the study. K.G., R.B.J., and M.S.T. designed the study, with contributions from O.V., R.Z.A., J.W.H., H.W.P., and W.J.R. on the study development. K.G., O.V., A.A., A.F.A.P., W.F., and J.L.S. collected and contributed data. K.G. and O.V. performed statistical and model analyses. K.G. wrote the manuscript, with contributions from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-31540-9>.

**Correspondence** and requests for materials should be addressed to Katerina Georgiou.

**Peer review information** *Nature Communications* thanks Emanuele Lugato and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022