


# Revisiting the soil carbon saturation concept to inform a risk index in European agricultural soils

Received: 22 April 2024

Accepted: 19 February 2025

Published online: 18 March 2025

 Check for updates

T. S. Breure<sup>1</sup>✉, D. De Rosa<sup>2</sup>, P. Panagos<sup>1</sup>, M. F. Cotrufo<sup>3</sup>, A. Jones<sup>1</sup> & E. Lugato<sup>1</sup>

The form in which soil organic carbon (SOC) is stored determines its capacity and stability, commonly described by separating bulk SOC into its particulate- (POC) and mineral-associated (MAOC) constituents. MAOC is more persistent, but the association with mineral surfaces imposes a maximum MAOC capacity for a given fine fraction content. Here, we leverage SOC fraction data and spectroscopy to investigate POC/MAOC distribution, together with SOC changes data over 2009–2018 period, across pedo-climatic zones in the European Union and the UK. We find that rather than a universal mineralogy-dependent maximum MAOC capacity, an emergent effective MAOC capacity can be identified across pedo-climatic zones. These findings led us to propose the SOC risk index, combining SOC changes and effective MAOC capacity. We find that between 43 and 83 Mha of agricultural soils are classified as high risk, mostly constrained to cool and humid regions. The index provides a synthetic information to decision makers for preserving and accruing POC and MAOC.

The pathway to climate neutrality foresees the contribution of the land to offset the residual sectorial greenhouse gas emissions by 2050, incrementing the carbon (C) removal from vegetation and soil. In the European Union (EU), operative policy instruments to increase the land C sink are atmospheric carbon dioxide (CO<sub>2</sub>) removal targets in the land use, land use change and forestry (LULUCF) regulation<sup>1</sup> as well as the recent Carbon Removal Certification regulation<sup>2</sup>, including carbon farming. Agricultural soils in the EU, in particular, are depleted in soil organic carbon (SOC) as compared to other land uses<sup>3</sup>. Furthermore, the majority of EU agricultural soils are far from saturation of the stable mineral-associated organic carbon (MAOC) fraction<sup>4,5</sup>, allowing the storage of additional C by changing to appropriate management practices<sup>6,7</sup>. However, a recent data-driven study estimated a relative SOC loss of 0.75% for the period 2009–2018 in European agricultural soils<sup>8</sup>. These SOC losses occurred despite the introduction of both mandatory and voluntary schemes in 2013, aiming at increasing agricultural sustainability<sup>9</sup>.

Assessing current bulk SOC content and its change over time ( $\Delta$ SOC), while fundamental, does not provide enough information for effective SOC sequestration interventions. In the last decades, a new conceptual framework has highlighted the advantage to separate bulk SOC in two fractions that underlie prevailing mechanisms of SOC formation and stabilization, namely the MAOC and the particulate organic carbon (POC)<sup>10</sup>. MAOC is mostly composed of plant and microbial derived compounds low in molecular weight, which can be stabilized by interaction with the soil matrix via sorption and physical protection<sup>11</sup>. Consequently, MAOC is more resilient to degradation compared to POC, and it has a lower turnover time on average<sup>4</sup> which promotes the long-term accrual of atmospheric CO<sub>2</sub> into soil. However, MAOC has a ‘theoretical mineral capacity’ due to a finite number of mineral surface binding sites, as postulated and demonstrated by a large body of studies<sup>12–14</sup>. Therefore, the degree of MAOC saturation indicates the proportion of measured MAOC over the theoretical capacity. The theoretical mineral capacity is commonly calculated

<sup>1</sup>European Commission, Joint Research Centre, Ispra, Italy. <sup>2</sup>Department of Agriculture, Forestry, Food and Environmental Sciences, University of Basilicata, Potenza, Italy. <sup>3</sup>Department of Soil and Crop Science and Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, USA.

✉ e-mail: [timo.breure@ec.europa.eu](mailto:timo.breure@ec.europa.eu)

based on soil texture and clay mineralogy to benchmark the saturation deficit of soils from global databases<sup>15,16</sup>. Here, we argue that this mineralogical capacity has a low practical importance for carbon accrual actions as, for instance, Mediterranean soils would never reach the MAOC content of acidic soils under the cold climate of northern Europe, even when sharing the same texture<sup>17,18</sup>. Commonly, the method used to calculate a unifying theoretical mineral MAOC capacity consists of pooling data together from different soil types, environmental and management conditions. Then, a linear regression is applied between MAOC and the soil's fine fraction content separately for high and low activity minerals (e.g. Georgiou et al.<sup>16</sup>). However, this approach does not acknowledge that the theoretical mineral capacity may not be achievable as MAOC storage is constrained by additional emerging ecosystem properties<sup>19</sup> that regulate SOC formation and stabilization such as pH, microbiome characteristics, type of litter and plant productivity<sup>4,20–24</sup>. The theoretical mineral capacity has also been questioned by recent studies<sup>25</sup>, suggesting oversaturation of mineral particles due to the binding of organic matter to other organic matter bonded to minerals, therewith posing the fundamental question to what degree of surface loading MAOC can be still considered as “stabilized” by mineral-association<sup>26</sup>.

Based on these premises, we use a clustered approach to calculate the ‘effective MAOC capacity’. We followed Stewart et al. (2007), in that an apparent saturation limit can be reached since pedo-climatic and management conditions impose constraints even with increased C inputs<sup>18</sup>. The effective MAOC capacity in our clustered approach thus constitutes the biophysically achievable MAOC given the cluster's pedo-climatic properties, for soils under agricultural land use (Supplementary Fig. 1). Further, to account for the oversaturation of mineral particles<sup>25,27</sup>, we formulated three different regression methods to estimate the effective MAOC capacity.

Additionally, we leveraged information from four datasets to map a risk index (Supplementary Fig. 1): i) the SOC content in locations that have been repeatedly surveyed (2009–2018) in the EU Land Use and Land Cover Survey (LUCAS)<sup>28,29</sup>, ii) the SOC changes ( $\Delta$ SOC) between the repeated surveys<sup>8</sup>, iii) associated visible- and near-infrared (VNIR) spectroscopy measurements<sup>30</sup>, and iv) a subset of measured SOC fractions<sup>3,27</sup>. The risk index builds on the exposure-vulnerability-hazard concept from the Intergovernmental Panel on Climate Change<sup>31</sup>,

(Fig. 1). The exposure component consists of the areal extent, which are all soils under agricultural land use. The hazard is represented by  $\Delta$ SOC, which is the effect of climate and management on SOC storage. Vulnerability is represented by the level of MAOC saturation within biophysically homogeneous European agricultural regions. We suggest that mapping the vulnerability and hazard components of agricultural SOC is informative for SOC management. While we applied this conceptual framework to the EU, which may be further refined with additional data, we suggest to apply it in other regions to identify areas at risk of SOC loss as well as areas with the highest potential for SOC accrual.

## Results and discussion

### Clustering of pedo-climatic zones across Europe

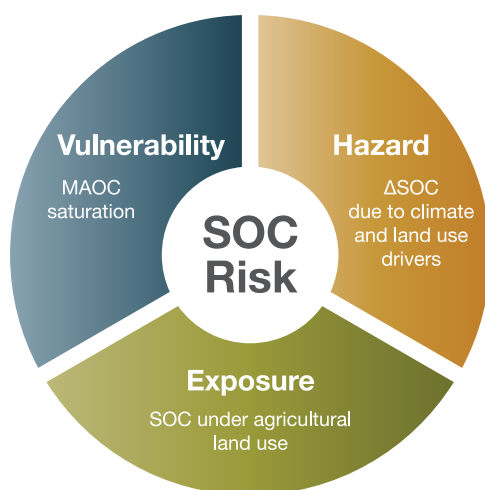
Bulk SOC storage is known to be an ecosystem property controlled by climatic conditions, management, plant productivity, soil properties such as texture and pH, and geomorphological features such as elevation or slope<sup>19,21,23</sup>. Therefore, approaches using pedo-climatic clustering can provide reliable estimates of bulk SOC storage, as recently demonstrated across Europe<sup>32</sup>.

Similarly to bulk SOC, fractions vary with environmental, geochemical and landform gradients<sup>22,24</sup>. Thus, accounting for these conditions by applying a clustering approach can enable more accurate estimation of the effective MAOC capacity<sup>24</sup>. We therefore applied a *k*-means clustering procedure based on aridity<sup>33</sup>, net primary productivity (NPP)<sup>34</sup>, measured pH in H<sub>2</sub>O<sup>28</sup> and landform<sup>35</sup> for the LUCAS soil sampling locations. Soil pH was included as a proxy of clay mineralogy and SOC turnover (microbial composition), landform to account for how the erosion and depositional setting affects preferential displacement of SOC fractions, NPP as a driver of saturation through C inputs<sup>36</sup> and aridity as a synthetic climate parameter controlling SOC storage<sup>37–39</sup>. We identified sixteen pedo-climatic clusters such as the coastal areas in mid- and southern-Europe (cluster 1) (Fig. 2), which generally receive high precipitation rates and show a large net primary productivity (Fig. 2b).

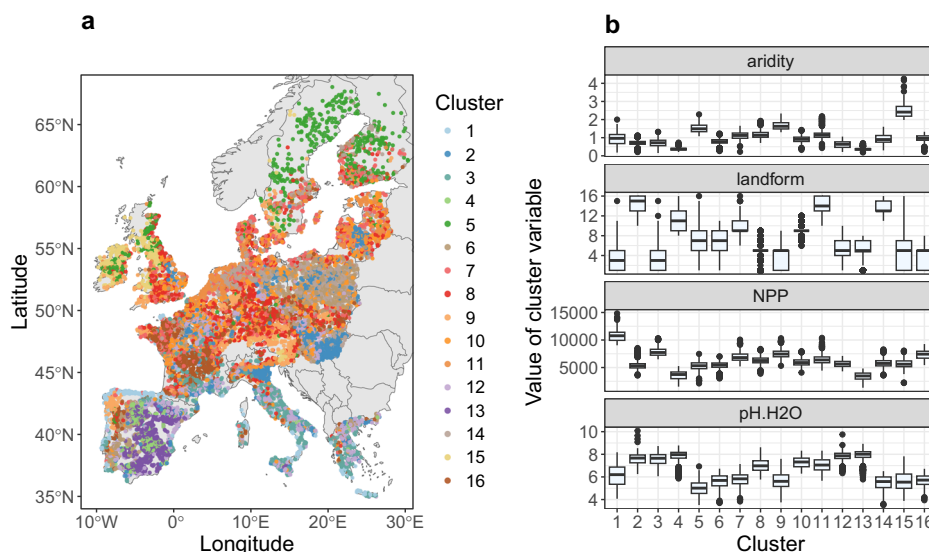
Relatively arid Mediterranean areas were attributed to separate clusters (3, 4 and 13), depending on their differences in landform, whereas their pH range was comparable. Other characteristic pedo-climatic zones were temperate lowland areas and the acid soils in north-western Europe (clusters 2, 5 and 15; Fig. 2). Pedo-climatic clusters also varied across smaller geographical scales. For example, considering the island of Sardinia (IT), the coastal areas were separated from the inland which showed further variation depending on the landform and soil pH (Fig. 2a).

### MAOC and POC predictions and total SOC changes in agricultural soils

Based on a subset of measured C fractions<sup>3</sup>, we predicted POC and MAOC for the remainder of the LUCAS 2009 survey using visible near-infrared (VNIR) soil spectra. The VNIR spectra allowed for an independent estimate of POC and MAOC from the covariates used in the pedo-climatic clustering. Predicted MAOC showed good correspondence with measured values for the validation dataset, although to a lesser extent for POC (Supplementary Fig. 2). However, predicted POC and MAOC (i.e. POC + MAOC) showed relatively good correspondence with measured bulk SOC (Supplementary Fig. 3a), considering the number of samples and geographic extent of the LUCAS survey ( $R^2 = 0.59$ , RMSE = 8.8 g kg<sup>-1</sup>, RPIQ = 1.7, Bias = 0.48, CCC = 0.76). We related the MAOC:SOC ratio to predicted carbon changes ( $\Delta$ SOC) between the 2009–2018 surveys based on De Rosa et al. When plotting  $\Delta$ SOC versus the MAOC:SOC ratio, pedo-climatic clusters showed different ranges for both variables (Fig. 3). Based on a linear least-squares model, all clusters showed a positive slope, where the interaction term of  $\Delta$ SOC x Cluster was significant for different slope estimates (Supplementary Fig. 4, Supplementary Table 1–2). The positive

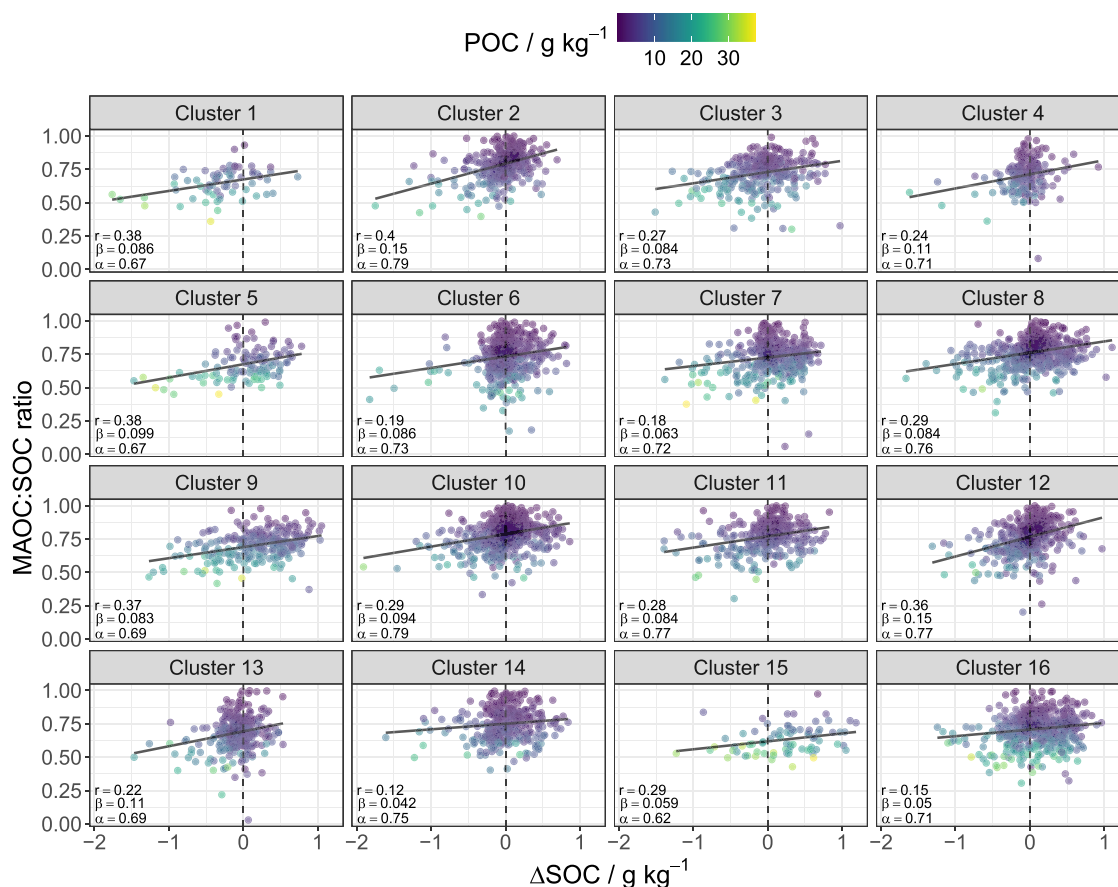


**Fig. 1 | Soil organic carbon (SOC) risk framework.** based on the exposure-hazard-vulnerability risk concept from the Intergovernmental Panel on Climate Change (IPCC)<sup>31</sup>. Soil organic carbon under agricultural land use is considered exposed. Vulnerability, the mineral-associated organic carbon (MAOC) saturation, determines the magnitude of the exposure. The hazard (soil organic carbon changes,  $\Delta$ SOC) is the integrated effect of climate change and land use that acts on exposed SOC.



**Fig. 2 | Allocated clusters based on the *k*-means method using the Hartigan and Wong (1979) algorithm.** **a** Geographical representation of the LUCAS 2009 survey. **b** Boxplots of aridity, landform, net primary productivity (NPP) and soil pH used for

the *k*-means in their original scale, by cluster association ( $n = 13,295$ ). Vector map data used from the ‘rnaturland’ R package<sup>31</sup>. Copyright (CC) (2025), (CRAN).

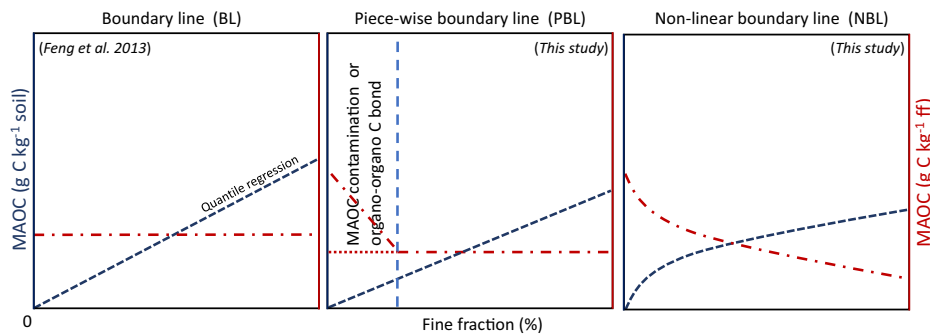


**Fig. 3 | Predicted changes in soil organic carbon ( $\Delta$ SOC) versus the predicted mineral-associated organic carbon fraction (MAOC) of total soil organic carbon [MAOC:SOC ratio (with SOC = MAOC + POC)]. colored by the soil particulate**

organic carbon (POC) concentration for each pedo-climatic cluster. Cluster numbers are as in Fig. 2.  $r$  is the Pearson's correlation coefficient,  $\alpha$  and  $\beta$  are the intercept and slope estimates of the linear least-squares model (solid black line). ( $n = 5482$ ).

slope suggests that SOC losses (negative  $\Delta$ SOC) were generally associated with both a higher contribution of POC to total SOC (i.e., low MAOC:SOC ratio) and a high POC content ( $\text{g kg}^{-1}$  of soil; Fig. 3). These results are in line with previous findings, suggesting that POC is more

vulnerable to disturbance than MAOC<sup>3,38</sup>, and that MAOC is responsible for the largest amount of SOC increase from high quality inputs in agriculture<sup>40,41</sup>. Furthermore, cold and wet regions (clusters 5 and 15) had a lower MAOC:SOC ratio and high POC content (Fig. 3<sup>39</sup>). Data



**Fig. 4 | Conceptual representation of different regression methods to estimate the effective mineral-associated organic carbon (MAOC) capacity.** Boundary line (BL) assumes that MAOC content ( $\text{g C kg}^{-1}$  in fine fraction (ff)) has a constant maximum across the fine fraction range<sup>15</sup>. Piece-wise boundary line (PBL) assumes a constant for the effective MAOC capacity (like Feng et al.) but isolates the disturbance of MAOC contamination and organo-organano C bonds. The non-linear

boundary line (NBL) assumes that MAOC content can change across the fine fraction range; effective MAOC capacity includes SOC organic layers that interact via organo-organano C bonds rather than directly to the clay surface. All regressions were done for the 90th quantile with a forced intercept to 0 as per Feng et al. Data and fitting statistics for the LUCAS survey is presented in Supplementary Figs. 5–7.

points with SOC gains (positive  $\Delta\text{SOC}$ ) generally showed higher MAOC:SOC ratios (Fig. 3; Supplementary Fig. 4), although SOC losses occurred also in soils with high MAOC:SOC ratio, due to varying type of perturbations driving SOC changes at the LUCAS sites.

#### Estimating the effective MAOC capacity by clusters

Given that MAOC is less associated with SOC losses by perturbation than POC, efficient C management should better consider the degree to which soils are from their effective MAOC capacity, which we hypothesized being an emergent property of pedo-climatic clusters. We further formulated three different methods to estimate the effective MAOC capacity based on underlying MAOC saturation theory (Fig. 4).

The concept of MAOC saturation was first proposed by Hassink (1997) that described MAOC saturation as a linear function of the soil's fine fraction content (clay + silt content)<sup>12</sup>. Successively, Feng et al. proposed the boundary line (BL) regression as a method to better restrict the inference to data from soils close to MAOC saturation<sup>15</sup> (Fig. 4). Here we propose a variation of the BL method, which we called PBL, to better isolate soils that have mineral association between organic carbon and the fine fraction. Conversely, suggesting oversaturation of mineral particles due to the binding of organic matter to other organic matter bonded to minerals<sup>25</sup>, we used a non-linear regression (NBL) to calculate the MAOC saturation (Fig. 4).

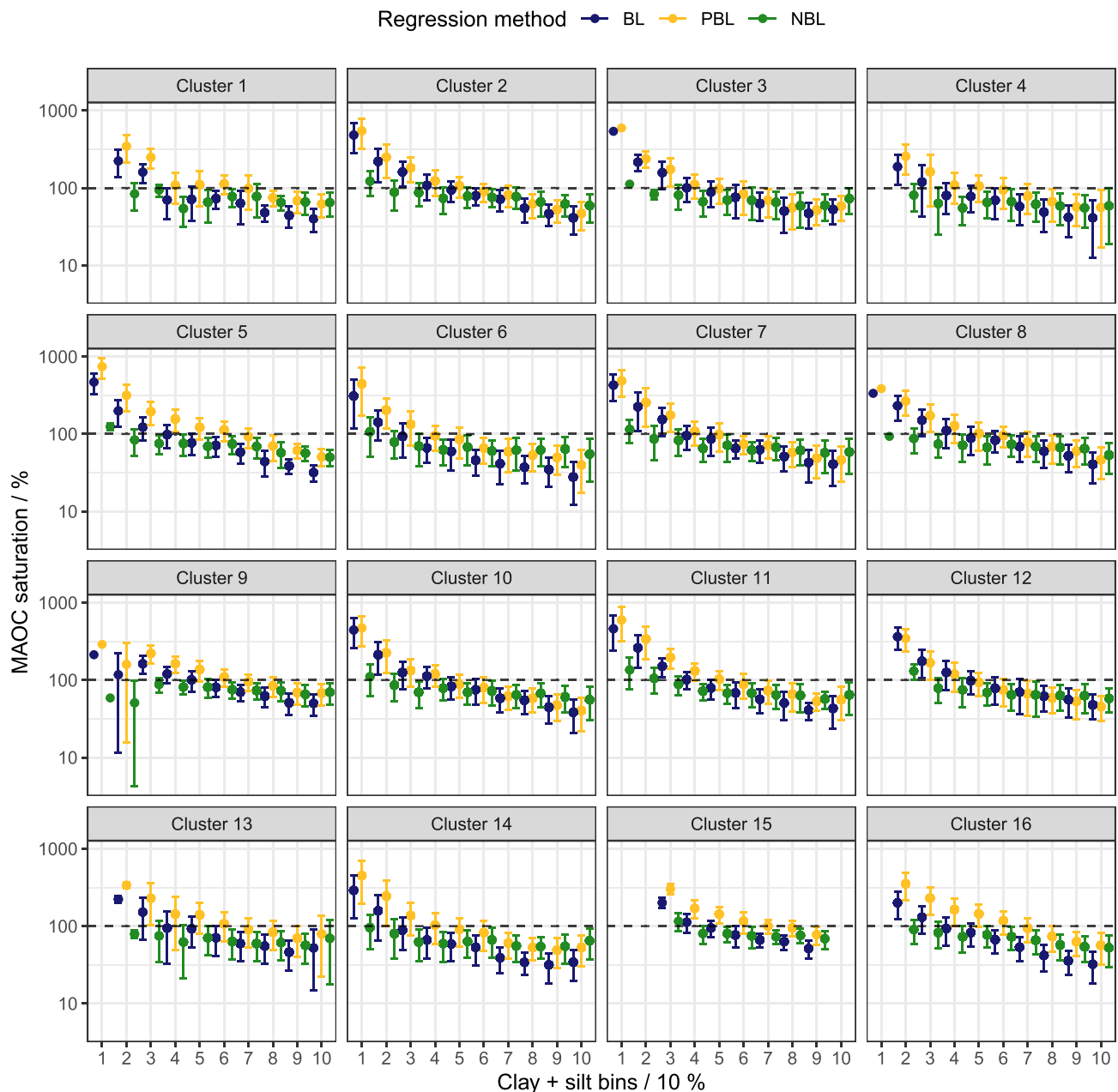
Following the different approaches illustrated, we performed quantile regressions to determine the cluster effective MAOC capacity (Supplementary Figs. 5–7). Parameter estimates varied by up to 200% between pedo-climatic clusters (Supplementary Figs. 5–7), which provides evidence that the effective MAOC capacity is an emergent property based on pedo-climatic conditions<sup>19,24</sup>. From the pedo-climatic variables used in the clustering, aridity and net primary productivity showed larger effects on the spread of parameter estimates compared to pH and landform. This suggests that aridity and NPP play a larger role in the distinction between the theoretical and effective MAOC capacity for our dataset (Supplementary Fig. 8), although the importance of pedo-climatic controls might vary across different regional scales<sup>22</sup>.

However, the functional relationship of the BL method did not seem to fit the data well for most clusters, since the 90th quantile regression line under-fits for coarse soils (i.e., low in fine fraction content) and over-fits for soils high in fine fraction content (Supplementary Fig. 5). This fitting phenomenon has been reported previously in a study for German soils<sup>5</sup> and our study provides further evidence on a continental scale. The estimated MAOC capacity for the PBL method was less variable between clusters compared to the BL method. The estimated breakpoints, however, varied widely in their magnitude between clusters (19–68%), depending on the level at which MAOC in

fine fraction reached a plateau (Supplementary Fig. 6). The non-linear NBL method allowed for a good fit of the upper boundary accounting for a slight increase in MAOC along the fine fraction range (Supplementary Fig. 7). Recently, Viscarra-Rossel et al. estimated the effective MAOC capacity across the Australian continent by soil groups. Their regression method is analogous to the underlying theory of our NBL method, in that it assumes a fine fraction-dependent MAOC concentration. To align with the existing literature, we have included estimates based on their method (Supplementary Fig. 9). These estimates can be considered an upper-limit of the effective MAOC capacity for our dataset, given that the frontier line method is specifically aimed at estimating the maxima of the data<sup>24</sup>. We preferred to use a more conservative quantile approach (see Method section), the parametric nature of which allows for comparison with previous studies.

Whereas two of our proposed regression methods to estimate MAOC saturation are novel to this study (Fig. 4), there are few studies that used the exact same regression type (90th quantile, 0 intercept) to determine MAOC saturation<sup>42</sup>. Here we compared our results to the existing literature. We first converted all parameter estimates to the same unit ( $\text{g MAOC kg}^{-1}$  fine fraction). The mean of the  $\beta$  parameter estimates across clusters for the boundary line method (BL) was  $45.1 \pm 11.3$  (SD) (Supplementary Fig. 5). The mean  $\alpha$  value for the piecewise regression method (PBL) was  $34.1 \pm 6.6$  (SD). Differences between the PBL method compared to the BL method occur because the MAOC concentration for soils low in fine fraction is assumed to be contaminated by POC or characterized by organo-organano C bonds<sup>25,26</sup> and, thus, identified before the estimated breaking-point in the piecewise regression (Supplementary Fig. 6). These results show that disregarding coarse soils with high MAOC content leads to lower estimates of the effective MAOC capacity. The mean estimate for the NBL method was  $28.5 \pm 6$  (SD) and spanned the smallest range across pedo-climatic zones (Supplementary Figs. 5–7). The upper limit parameter values for the NBL and PBL methods (41, 45, respectively) were lower than for BL (62), which is lower than previous estimates for 2:1 mineral dominated soils:  $84 \pm 4$  (SE)<sup>15</sup>, 90th quantile regression) and  $86 \pm 9$  (90% CI)<sup>16</sup>, 95th quantile regression).

Given the LUCAS sampling design, our analysis is likely to be representative of the most abundant soil types across Europe<sup>28</sup>. However, the dataset that we used to calibrate VNIR spectra against C fractions to predict the 6,548 samples is relatively small and therefore can impose a limitation. For example, recent studies<sup>42</sup> pointed to soils with higher MAOC content that may be formed under particular conditions (e.g., very high clay, hydromorphic conditions) and can exceed  $50 \text{ g MAOC kg}^{-1}$  soil; the maximum MAOC content in the C fraction dataset used in this study. For example, MAOC accumulation



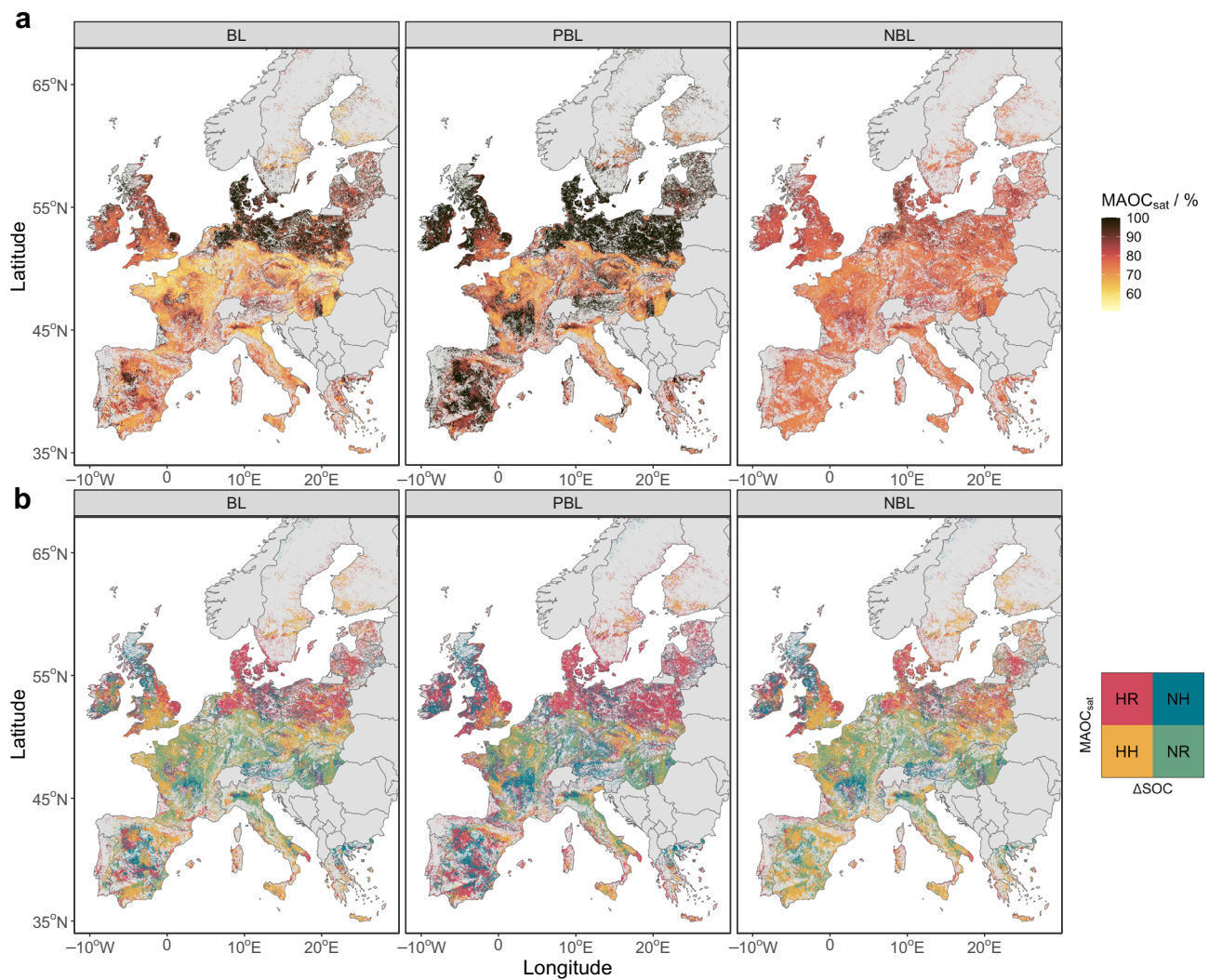
**Fig. 5 | Degree of mineral-associated organic carbon (MAOC) saturation (MAOC / effective MAOC capacity  $\times$  100%) as a function of fine fraction (clay + silt, %). Values below 100% indicate a saturation deficit relative to the cluster-dependent**

effective MAOC capacity. Fine fraction content has been binned by intervals of 10%, points and lines represent the mean degree of MAOC saturation  $\pm$  standard deviation for each bin. The y-axis is on a  $\log_{10}$  scale. ( $n = 6548$ ).

can occur due to oxygen limitation rather than mineral stabilization<sup>42</sup>, such as in Stagnosols. Another component that might interfere with MAOC accounting is the presence of geogenic C, i.e., the organic C present in the bedrock that was deposited during sedimentation<sup>42,43</sup>. These findings further support that a clustered approach is more meaningful for the inference of effective MAOC capacity from data spanning a broad range of soil types. That is, disaggregation prevents a limited number of points, potentially belonging to one particular soil type, from having high leverage in the regression. Future research could investigate how oxygen limited conditions and geogenic C affect regional estimates of MAOC saturation. Nevertheless, we have also investigated the effect of including the existing legacy soil C fraction data<sup>516</sup> on the parameter estimates (Supplementary Figs. 10–13). Based

on these results, we anticipate that the exclusion of soil under-represented in the LUCAS dataset might underestimate the maximum MAOC capacity for fine-textured soils in some pedo-climatic clusters. Nonetheless, the mean estimate across our EU pedo-climatic clusters was more similar to those found for 2:1 mineral dominated soils (global coverage) under cropland:  $45 \pm 5(\text{SE})^{15}$  (Supplementary Figs. 5–7). Also the estimates for non-clustered data align very closely with that of Feng et al. for 2:1 mineral dominated soils under cropland<sup>15</sup> (Supplementary Fig. 14). The NBL method notably led to lower estimates (29 g MAOC kg<sup>-1</sup> fine fraction).

Similar differences between regression methods were found after calculating the degree (as percentage) of MAOC saturation (MAOC / effective MAOC capacity  $\times$  100%) and computing the mean and its



**Fig. 6 | Geographical representation of mineral-associated organic carbon (MAOC) degree of saturation ( $MAOC_{sat}$ ) and the soil organic carbon (SOC) risk index. **a** degree of MAOC saturation mapped to grid cells by cluster-specific MAOC saturation relationship for fine fraction bins. Values below 100% indicate a saturation deficit relative to the cluster-dependent effective MAOC capacity. Panels correspond to different methods to estimate the effective MAOC capacity (see Fig. 4). **b** follows the same panel structure and shows the SOC risk index:**

whether MAOC saturation is above or below its median and whether  $\Delta SOC$  is below 0 or not. 'HR': high risk, negative  $\Delta SOC$ , MAOC saturation above median. 'HH': high hazard, negative  $\Delta SOC$  and MAOC saturation below the median. 'NR': no risk, positive  $\Delta SOC$  and MAOC saturation below the median. 'NH': no hazard, positive  $\Delta SOC$  and MAOC saturation above the median. Vector map data used from the 'rnatuarearth' R package<sup>31</sup>. Copyright (CC) (2025), (CRAN).

standard deviation for fine fraction intervals by cluster (Fig. 5). Values below 100% indicate a saturation deficit relative to the cluster-dependent effective MAOC capacity. The figures where the degree of MAOC saturation has not been binned by fine fraction intervals can be found in Supplementary Fig. 15.

When the BL or PBL was applied, MAOC content exceeded the theoretical saturation (> 100%) in sandy soils likely because the quantile regression is underfitting the data low in fine fraction content. In total, 1202 samples exceeded 100% for BL and 1408 samples for PBL. For fine soils, the BL generally estimated lower MAOC saturation compared to the other methods. The NBL method did not frequently exceed 100% saturation (708 samples) and remained more constant across the range of fine fraction content compared to the other two methods. These characteristics can be attributed to a better fit across the range of fine fraction content, reducing variability in MAOC saturation at the extremes. However, we restricted MAOC saturation estimates to a maximum of 100% for subsequent analysis, under the assumption that any values above 100% indicate saturation.

### The SOC risk index

Since we demonstrated that effective MAOC saturation is cluster-dependent and interplaying with SOC vulnerability, here we propose a synthetic 'risk index', which may guide most effective action to protect or accrue SOC. We did this by borrowing the hazard-exposure-vulnerability risk framework used by the Intergovernmental Panel on Climate Change<sup>31</sup> (Fig. 1). Soil organic carbon under agricultural land use is considered exposed to anthropogenic and environmental drivers, and thus determines the areal extent of soils under exposure in the EU. The degree of MAOC saturation was taken as a measure of vulnerability, given that soils saturated in MAOC are more likely to either have a high SOC content and/or have proportionally higher POC. The SOC changes ( $\Delta SOC$ ) were considered as the level of hazard, that is, SOC changes driven by climatic conditions and land management<sup>8</sup>. By assessing the degree of hazard and vulnerability, we constituted four index classes (high risk, high hazard, no risk, no hazard) that allow for a spatial assessment of SOC status. High hazard (HH) and high risk (HR) are both subject to SOC losses but have low and high levels of vulnerability, respectively (MAOC saturation below or above the

median of 68.9%). No risk (NR) and no hazard (NH) have SOC gains with low and high vulnerability, respectively (Fig. 1, Fig. 6b). To assess the SOC risk index across Europe, we imposed the effective MAOC capacity as a function of the fine fraction for each pedo-climatic cluster (Method section, Supplementary Figs. 16–17). Upscaling the clusters thus allowed us to: i) map the degree of MAOC saturation across Europe and relate these estimates to  $\Delta$ SOC and ii) reclassify both MAOC saturation and  $\Delta$ SOC to assess the SOC risk index in agricultural soils.

The PBL method calculated a higher MAOC saturation in particular for the Baltic Sea area, northern UK and the Iberian Peninsula, although it followed the same geographical pattern as the BL method (Fig. 6a). This reflects that the PBL method estimates a lower effective MAOC capacity for coarse textured soils. That is, the abovementioned regions are characterized by low clay content and/or high sand. The NBL method was distinct compared to the two other approaches given its relative narrow range of MAOC saturation across Europe. For example, a smaller area is calculated to be MAOC saturated (mostly restricted to Denmark and north-east Germany) (Fig. 6a), likely due to a better fit for coarse soils leading them to have lower MAOC saturation (Fig. 5). The differences in MAOC saturation across Europe illustrate how the methodological decision to calculate the MAOC capacity can have implications for SOC management. The NBL method seemed to provide a better fit to our data, given the assumption that MAOC content can change across the fine fraction range, suggesting that future work

should be directed towards the NBL method. Whereas the type (mineral or organo-organ) and strength of C bonds can only be evaluated at nanoscale level with time- and cost-intensive analysis, we suggest that the NBL method can lead to a less-biased quantification of the SOC risk.

The SOC risk index showed a more refined distinction in C accrual potentials compared to considering only the MAOC saturation (Fig. 6b). Across Europe, there are a variety of locations that are under high hazard, in the sense that they are characterized by a negative  $\Delta$ SOC but are below the median MAOC saturation ('HH'), spanning between 30–70 Mha, depending on the regression method. This situation occurs in particular across Scandinavia, central England, western France and some parts of the Mediterranean. The opposite combination, above the median MAOC saturation and positive  $\Delta$ SOC ('NH'), occurs in the northern UK, the Massif Central (FR), as well as in Austria and southern Germany. The areal extent under no hazard, 'NH', ranges from 25–48 Mha. Locations that are at high risk, above median MAOC saturation and negative  $\Delta$ SOC ('HR') cover an area of 43–83 Mha and occur mostly in countries bordering the Baltic Sea as well as northern Germany and east England. Lastly, the areas where SOC is least sensitive to losses and are at deficit in MAOC (below median MAOC saturation) are in the no risk class ('NR'). These areas could be potential locations for efficient C accrual through carbon farming. The main areas stretch from the west-coast Europe towards the east, across the countries of northern France, Belgium, and southern Germany to Hungary. Other notable locations include southwestern France and the Po valley (IT) (Fig. 6b). The total area for the 'NR' category covers 26–50 Mha. We note, however, that we have listed generic geographic patterns here and there is large variability within different pedo-climatic zones, also depending on the regression method (BL, PBL, NBL) used. The 'high hazard' and 'high risk' index classes were associated with larger uncertainty, given the larger range in their areal extent based on an uncertainty propagation analysis (Table 1).

Overall, however, there was 59.6% agreement between all three regression methods on the SOC risk index, while there was some method-dependent spatial disagreement, in central and south-west England, the Iberian peninsula and south Germany (Supplementary Fig. 18). The ratio between agreement/disagreement varied strongly between SOC risk index classes (Table 2), in particular for classes below the median MAOC saturation (NR and HH). These differences can likely be attributed to the different assumption of each regression method to estimate the effective MAOC capacity for coarse soils. Based on this 'convergence of evidence' approach, we conclude that the SOC risk index provides robust information for areas where to prioritize measures to revert degrading processes or protect the existent SOC pool (HR and NH classes), while there is more uncertainty on areas suitable and with some potential for SOC accrual (NR and HH classes).

**Table 1 | Summary table of the corresponding area (Mha) for each SOC risk index class**

SOC risk index	Method	Mean	Q5	Q95
High hazard (HH)	BL	59.2	54.4	61.7
	PBL	30.0	25.9	33.7
	NBL	69.8	67.0	75.4
No risk (NR)	BL	46.0	42.6	48.0
	PBL	26.2	23.6	29.1
	NBL	49.5	47.0	52.3
High risk (HR)	BL	54.0	51.5	58.9
	PBL	83.2	79.5	87.3
	NBL	43.4	37.8	46.2
No hazard (NH)	BL	28.7	26.7	32.0
	PBL	48.5	45.5	51.0
	NBL	25.1	22.4	27.6

For details on the regression methods, see Fig. 4 in the main manuscript. The mean values in Mha correspond to the values discussed in the main text, whose spatial patterns across Europe are provided in Fig. 6b. Q5 and Q95 refer to the 5<sup>th</sup> and 95<sup>th</sup> quantiles that were calculated based on an uncertainty propagation analysis (Methods section, Supplementary Figs. 22–23).

**Table 2 | Agreement between methods to estimate effective mineral-associated organic carbon (MAOC) capacity and their estimated area for each soil organic carbon (SOC) risk index class in million hectare (Mha)**

Agreement	SOC risk index	MAOC	$\Delta$ SOC	Area (Mha)	%
Agreement	High hazard (HH)	< median	< 0	27.5	46.4
	No risk (NR)	< median	> 0	24.6	53.5
	High risk (HR)	> median	< 0	39.0	72.2
	No hazard (NH)	> median	> 0	20.8	72.6
Disagreement	High hazard (HH)	< median	< 0	31.7	53.6
	No risk (NR)	< median	> 0	21.4	46.5
	High risk (HR)	> median	< 0	15.0	27.8
	No hazard (NH)	> median	> 0	7.9	27.4

Agreement means that all three methods allocate a raster cell to the same SOC risk index class. The relative percentage of agreement between the regression methods is calculated as area of agreement / total area for each SOC risk index class. Where the SOC risk index classes based on the non-linear boundary line (NBL) method have been taken as a reference for Table 2. The columns MAOC and  $\Delta$ SOC indicate which quadrant of the risk index that row belongs to (Fig. 6b). For details on the regression methods, see Fig. 4 in the main manuscript.

The magnitude of saturation also determines the rate of soil C accrual<sup>16,44</sup>, which affects the extent to which carbon farming is likely to be a cost-effective practice<sup>45</sup>. Conversely, the risk index is built on the concept that soils close to MAOC saturation lead to more rapid losses, as shown in a global synthesis<sup>16</sup>, due to higher POC content and weaker MAOC sorption/bonds to the mineral matrix<sup>11,13,44</sup>. Our data showed that high levels of MAOC saturation also led to higher C losses on a continental scale, consistently across all three boundary line regression methods (Supplementary Fig. 19, Supplementary Table 3). We further note that suitability for soil C accrual or SOC protection measures should not only focus on the degree of saturation but also on the absolute amount (i.e. in terms of  $\text{g kg}^{-1}$  of C) of MAOC. For example, locations that are under low risk and, thus, have more potential for SOC accrual ('NR'), have a range of effective carbon storage potentials, which depends on their pedo-climatic cluster associations and the boundary line regression method (Supplementary Fig. 20). While other biophysical limitations exist for SOC accrual practices, such as the availability of nutrients<sup>46,47</sup> and soil depth<sup>17</sup>, our index identifies potential regions for C accrual and protection, acknowledging constraints in terms of soil characteristics, NPP, climate and land use. Socio-economic and technical constraints may also limit the adoption of farming activities that aim at accruing or protecting SOC as, for instance, access to farm advisory services or risk aversion with respect to alternative management practices<sup>48</sup>. Future research could focus on expanding the MAOC dataset by additional soil sampling and C fraction measurements, to cover a wider range of soils and environmental conditions. Lastly, we have shown that calculating the degree of C saturation is affected by methodological decisions and we hope that our findings lead to further research towards a unified approach.

## Methods

### Analytical data

The LUCAS dataset consists of records from a 2009 sampling campaign, based on a random sampling design stratified by land use and topography. Soil cores were taken at a depth of 0–20 cm, see Tóth et al. for further details<sup>28</sup>. The bulked soil samples were air-dried and sieved to their <2 mm fractions. Soil analytical data for clay, silt, organic carbon and pH in  $\text{H}_2\text{O}$  was determined by standard methods following ISO protocols (Supplementary Table 4). Soil spectra in the visible- and near-infrared range (VNIR, 380–2500 nanometer (nm) range) were measured with a XDS Rapid Content Analyzer (FOSS NIRSystems, Inc., Denmark) at 0.5 nm spectral resolution. The protocols of the instrument manufacturer and the soil spectroscopy group<sup>49</sup> where followed for the spectroscopic measurements. For each sample, the mean spectrum was taken of two replicates. We restricted the LUCAS 2009 dataset to locations that were both under agricultural land use, as recorded by the surveyors corresponding to cropland and grassland under 1000 m a.s.l.<sup>8</sup>, and had associated VNIR spectra ( $n = 13,295$ ).

The analytical soil C fraction data was originally measured for a selection of soil samples from the LUCAS 2009 survey<sup>3</sup>, the procedure of which we briefly summarize here. Firstly, the aggregates were dispersed. Five grams of oven dried, <2 mm sieved soil was shaken for 18 hours in dilute (0.5%) sodium hexametaphosphate with beads. After aggregate dispersion, samples were fractionated by size through rinsing the soil samples onto a 53  $\mu\text{m}$  sieve (see<sup>3,27</sup> for further details). Where the < 53  $\mu\text{m}$  fraction was considered MAOC and the > 53  $\mu\text{m}$  was considered POC<sup>50</sup>. We then also restricted the soil C fraction dataset to locations that were both under agricultural land use and had associated VNIR spectra ( $n = 240$ ).

The processing of the spectra was done by sub-setting every 10<sup>th</sup> wavelength, trimming the spectra to the range of 400–2450 nm and computing the 1<sup>st</sup> derivative. Subsequently, the  $\text{H}_2\text{O}$  bands were removed from the spectra by excluding the 1350–1460 nm and 1790–1960 nm wavelength regions<sup>51</sup>.

### Calibration regression

Based on the results from an exploratory analysis (Supplementary Fig. 2), we decided to use a local partial least squares regression method. For the calibration regression method, we adapted the method described in Summerauer et al. (2021) to our purpose<sup>52</sup>. We used the moving-window correlation as a metric to select  $k$ -nearest neighbors based on spectral similarity. In order to choose the window size, we computed the RMSE between nearest neighbors for different window sizes (11–151 in steps of 10). We selected the window size with the lowest RMSE (Supplementary Fig. 21). After the nearest neighbors had been selected, a local model was fitted based on the weighted average partial least squares regression algorithm as per Shenk et al. (1997)<sup>53</sup>. For each number of components used in the PLS, from 1 to  $j$ , a weight is calculated based on the spectral residuals of the observation to be predicted. These weights are then used to average multiple PLS models computed for different number of components:

$$w_j = \frac{1}{\delta_{1j}g_j} \quad (1)$$

Where  $\delta_{1j}$  is the RMSE of the spectral residuals for a predicted sample based on  $j$  PLS components,  $g_j$  is the RMSE of the regression coefficients which corresponds to the  $j^{\text{th}}$  PLS component (more details in ref. 53). We considered a range of 5 to 15 PLS components.

The number of  $k$ -nearest neighbors was optimized by using nearest neighbor cross-validation<sup>54</sup> (Ramirez-Lopez et al. 2013). This method is essentially equivalent to a leave-one-out approach where for  $k$  nearest neighbors, each neighbor is excluded iteratively and predicted by a weighted PLS regression using the  $k-1$  nearest neighbors. The predictions are then cross-validated against their analytical values. We considered a value of  $k$  between 20 and 100 where the final  $k$  value was selected based on the minimum RMSE in the nearest neighbor cross-validation. We then restricted the LUCAS 2009 dataset to the SOC range of the soil C fraction calibration set (3.6–85.1  $\text{g kg}^{-1}$  SOC,  $n = 12,019$ ) and predicted MAOC and POC using the method described above.

### Determination of the calibration applicability domain

The aim of the VNIR predictions was to extend the soil C fractionation data across the entire LUCAS 2009 soil dataset. Thus, we needed to determine the applicability domain of our calibration regression based on our calibration set ( $n = 240$ ). An established method to do this for PLS predictions is through use of the  $F$ -ratio<sup>55–57</sup>. The main idea is to assess how well the PLS scores can reproduce the spectra of the validation set compared to those in the calibration set. This is achieved by dividing the residual variance of the spectra of the validation set by those of the calibration set:

$$F = \frac{(\mathbf{u} - \hat{\mathbf{u}})^T (\mathbf{u} - \hat{\mathbf{u}}) n_c}{s_c^2} \quad (2)$$

Where  $\mathbf{u}$  is the spectrum of the observation to be predicted,  $\hat{\mathbf{u}}$  is the spectrum of the observation to be predicted produced from the PLS scores,  $n_c$  is the number of observations in the calibration set and  $s_c^2$  the residual spectral variance of the calibration set. We computed the residual spectral variance of the calibration set with the projected PLS scores, whereas residual spectral variance of the validation set is computed by use of the predicted PLS scores. We then computed the probability of the  $F$ -ratio as per Dangal et al. (2019) and assigned a prediction as being out of the calibration applicability domain for probabilities exceeding 0.99. We then merged the POC and MAOC predictions into a single dataset, disregarding outliers for both predictions ( $n = 6548$ ) and setting negative predictions to 0. In order to validate our predictions, we assessed our predictions against the measured SOC content. We compared the sum of POC and MAOC (Supplementary Fig. 3a) and SOC predictions directly from the VNIR

spectra (Supplementary Fig. 3b). We evaluated predictions based on the following metrics: root mean squared error (RMSE), correlation coefficient ( $R^2$ ), bias, Lin's concordance correlation coefficient (CCC)<sup>58</sup> and the ratio of the standard prediction error over the inter-quartile range (RPIQ)<sup>59</sup>.

We also visually examined the model applicability domain through the use of principal components analysis (PCA) on the VNIR spectra. We plotted the joint distribution of the first two PCA components (explaining 75.8% of the variance) to assess whether the samples within the model applicability domain lay within the range of the calibration set (Supplementary Fig. 3c-d). We note that defining the model applicability domain reduced the number of predictions by almost half. This can be partially attributed to the limited representation of the calibration dataset with respect to the LUCAS 2009 survey (Supplementary Fig. 3c). Ideally samples should span the range of spectral variability, which did not seem to be the case according to this diagnostic plot. Additionally, the spectral information might be limited in terms of the absorption features related to carbon fractions in the VNIR range. That is, previous studies have found that mid-infrared soil spectra can lead to better predictions of soil C fractions<sup>60,61</sup>, although this likely depends on the fractionation method<sup>62</sup> and soil characteristics in the study<sup>63</sup>.

### Clustering

We then clustered the LUCAS dataset ( $n=13,295$ ) into pedo-climatic zones based on a combination of climate, pedological and landscape factors<sup>21–23</sup>: i.) measured pH in  $H_2O$ <sup>28</sup>, ii.) landform classes computed from digital elevation data<sup>35</sup>, iii.) MODIS/Terra cumulative net primary productivity (NPP,  $kg\ C / m^2 / year$ )<sup>34</sup> and iv.) the aridity index (precipitation / potential evapotranspiration) based on the TerraClim dataset<sup>33</sup>. Both NPP and the aridity index were extracted using Google Earth Engine and the mean was calculated over the period 01-01-2001–01-01-2021 at 1000 m resolution. We applied the  $k$ -means method using the Hartigan and Wong (1979) algorithm<sup>64</sup>. All variables were scaled to unit variance. We ran the following iteration 100 times for different seeds (random number generators): within each iteration the  $k$ -means was ran 100 times for different random allocations of initial centers. We considered a maximum of  $k=20$  to minimize over-dispersion of the LUCAS dataset, and selected the number of clusters that minimized the within-cluster sum of squares based on the elbow method, i.e. the minimum of the 2<sup>nd</sup> derivative. From the 100 iterations, we then selected  $k$  that was most frequent (see Suppl. Material for more details). These cluster associations were then used for the dataset that disregarded outliers of POC and MAOC predictions ( $n=6548$ ).

Additionally, we computed a random forest regression between the cluster associations and the variables used in the  $k$ -means method (scaled pH in  $H_2O$ , NPP, landform and aridity). The regression allowed us to upscale the cluster associations and thus was used to predict clusters with 1000 m grid resolution and the same extent as the Europe-wide raster dataset provided in De Rosa et al., reporting SOC changes in the period 2009–2018 ( $\Delta SOC$ )<sup>8</sup>. The raster extent corresponds to areas that were under agricultural land use (cropland or grassland), as per the Corine Land Cover dataset (<https://land.copernicus.eu/pan-european/corine-land-cover>). The pH in  $H_2O$  and the fine fraction rasters were taken from the study by Ballabio et al.<sup>65,66</sup>. The raster with predicted cluster associations was used in the last step of our methodology to calculate the SOC risk index. All rasters that were not at 1000 m grid resolution (pH in  $H_2O$ , fine fraction,  $\Delta SOC$  and landform) were resampled using bilinear interpolation.

### Quantifying the degree of MAOC saturation and the SOC changes

We explored three different regression methods to estimate the effective MAOC saturation capacity as a function of the fine fraction, based on different hypotheses of MAOC saturation dynamics (Fig. 4). The first

was the boundary line regression (BL)<sup>15</sup>. The second, PBL, was an alternative method to filter for samples that are likely to contain POC in the MAOC fraction. This might occur during size separation, both due to POC fragmentation and dispersion of POC into dissolved organic carbon (DOC)<sup>26,42</sup>. We first expressed the predicted MAOC as a function of the fine fraction (g MAOC in  $kg^{-1}$  fine fraction). We then determined the break-point of a piecewise linear 90<sup>th</sup> quantile regression while constraining the slope of the second linear equation to 0, such that the effective MAOC capacity was constant across the remaining fine fraction range (as first proposed in Hassink, 1997<sup>12</sup>). Any values prior to the breakpoint were then considered as MAOC saturated. The breakpoint was determined through use of the segmented() function in R<sup>67</sup>. The third method, NBL, was a non-linear boundary line regression, also on the 90<sup>th</sup> quantile and restricting the intercept to 0. For the non-linear quantile regression we considered a logarithmic model:

$$y = \alpha + \beta \log(x) \quad (3)$$

where  $y$  = the effective MAOC capacity,  $\alpha = 0$ ,  $\beta$  is the coefficient to be estimated and  $x$  is the fine fraction (clay + silt / %). For all three regression methods, the degree of MAOC saturation was calculated as (MAOC / effective MAOC capacity)  $\times$  100. Values below 100% indicate a saturation deficit relative to the cluster-dependent effective MAOC capacity. Values above 100% were considered saturated and set to 100% for subsequent analysis. This procedure was repeated for each pedo-climatic cluster.

To investigate the change in SOC ( $\Delta SOC$ ) as a function of the ratio of MAOC to SOC (Fig. 3), we used  $\Delta SOC$  values obtained from the non-linear regression model developed by De Rosa et al.<sup>8</sup>. The trained model was used to analyze the changes in SOC between the LUCAS 2009 and 2018 surveys across the EU + UK. Since the regression model used in De Rosa et al. study depends on land use information collected over time, our analysis was constrained to sites that had repeated recordings of land use across surveys. As a result our dataset of predicted MAOC and POC was reduced to 5482 points for our investigation of  $\Delta SOC$  as a function of MAOC:SOC (Fig. 3).

### The SOC risk index

Lastly, we mapped at 1000 m resolution the degree of MAOC saturation and predicted  $\Delta SOC$  to assess the vulnerability and hazard of exposed SOC in agricultural lands. We considered the level of risk for SOC as a function of both vulnerability and the hazard. This concept follows the framework introduced by the Intergovernmental Panel on Climate Change<sup>31</sup> which we have adapted to our purpose (Fig. 1). The degree of MAOC saturation was taken as a measure of vulnerability, given that soils saturated in MAOC are more likely to either have a high SOC content and/or have proportionally higher POC. The  $\Delta SOC$  was considered as the level of hazard, that is, SOC changes driven by climatic conditions and land use change<sup>8</sup>. We transferred the relationship between the degree of MAOC saturation and the fine fraction content based on the cluster associations. We did this by calculating the mean degree of MAOC saturation across fine fraction bins of 10%. We then mapped the mean MAOC saturation by fine fraction bin (Fig. 5) to a raster of the pedo-climatic clusters (Supplementary Fig. 17a) and the fine fraction raster<sup>65</sup> that was reclassified to align with the fine fraction bins. In a few cases, the range of fine fraction bins of the data (Fig. 5) did not cover those present in the raster. In that case, we considered the mean degree of MAOC saturation to be missing and those locations were thus excluded from the subsequent analysis. Finally, the SOC risk index was calculated across regression methods (BL, PBL, NBL) by determining for each raster cell whether: i.) it was above- or below the median degree of MAOC saturation across Europe, ii.) the  $\Delta SOC$  was below 0 or equal to 0 and above. We thus ended up with the following four classes: 'HR': at risk, above the median MAOC saturation and negative  $\Delta SOC$ . 'HH': high hazard, negative  $\Delta SOC$  and below the

median MAOC saturation. 'NR': low risk, positive  $\Delta$ SOC and below the median MAOC saturation. 'NH': positive  $\Delta$ SOC and above the median MAOC saturation. Although the median degree of MAOC saturation across the three regression methods (68%) is arbitrary, there is no scientific consensus yet on a generic threshold of MAOC saturation where C accrual diminishes and a C losses become more likely. We note that the linear model fitted on the  $\Delta$ SOC and the MAOC saturation (Supplementary Fig. 19, Supplementary Table 3) supports the decision to take 68% as a threshold. Given the fitted parameters, the MAOC saturation before  $\Delta$ SOC goes negative is 59%, 68% and 71%, for the BL, PBL and NBL method, respectively.

### Uncertainty propagation

We have performed an uncertainty propagation analysis based on the associated error with the MAOC predictions from the soil VNIR spectra. To assess the effect of marginal uncertainties in our MAOC predictions, we have approximated the expected error based on the predicted POC + MAOC vs. measured SOC (Supplementary Fig. 3). Given the negatively skewed distribution of SOC, we calculated the mean absolute log error (MALE). The MALE is robust to outliers (high SOC values). MALE reduces the effect of large differences between the predicted and measured values and provides a better measure of the relative difference. That is, the exponential of the MALE (EMALE) represents the relative multiplicative error (once we subtracted 1). We assumed the error to be normally distributed around the mean prediction and that POC and MAOC contribute equally, so we divided the EMALE by two. We then performed 500 simulations where we resampled the mean MAOC prediction with a standard deviation represented by (EMALE-1)  $\times$  MAOC (Supplementary Fig. 22). We then calculated the MAOC saturation and SOC index for each of these 500 realizations of MAOC. We calculated the 5<sup>th</sup> and 95<sup>th</sup> quantile for the MAOC saturation (Supplementary Fig. 23) and for the areas of each SOC index class (Table 1). See the Supplementary Material for further details.

Data handling, analysis and visualization was done using the following R packages: data handling with **tidyverse**<sup>68</sup>, **prospectr**<sup>69</sup>, **broom**<sup>70</sup>, regressions with **quantreg**<sup>71</sup>, **pls**<sup>72</sup>, **caret**<sup>73</sup>, **resemble**<sup>74</sup>, **mgcv**<sup>75</sup>, **mgcViz**<sup>76</sup>, **segmented**<sup>67</sup>, **emmeans**<sup>77</sup> handling of spatial objects using the **raster**<sup>78</sup> and **sf**<sup>79</sup> packages, clustering with the **cluster**<sup>80</sup> package. Graphics were created with base R functions, **rnaturalearth**<sup>81</sup>, **patchwork**<sup>82</sup> and the package **ggplot2**<sup>83</sup>.

### Data availability

The LUCAS 2009 soil survey and SOC fractionation data used in this study are available in the European Soil Data Centre (ESDAC) of the European Commission – Joint Research Centre under: <http://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data> and <https://esdac.jrc.ec.europa.eu/content/soil-organic-matter-som-fractions>. The main outputs of this study are available at: <https://esdac.jrc.ec.europa.eu/content/soil-carbon-risk-index>.

### Code availability

The most relevant R scripts to this manuscript are available at: <https://esdac.jrc.ec.europa.eu/content/soil-carbon-risk-index>.

### References

- Council, E. P. On the inclusion of greenhouse gas emissions and removals from land use, land use change and forestry in the 2030 climate and energy framework, and amending Regulation (EU) No 525/2013 and Decision No 529/2013/EU. <http://data.europa.eu/eli/reg/2018/841/2023-05-11> (2018).
- Council, E. P. Regulation (EU) 2024/3012 of the European Parliament and of the Council of 27 November 2024 establishing a Union certification framework for permanent carbon removals, carbon farming and carbon storage in products. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R3012> (2024).
- Lugato, E., Lavalley, 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).
- Cotrufo, M. F. & Lavalley, J. M. *Soil Organic Matter Formation, Persistence, and Functioning: A Synthesis of Current Understanding to Inform Its Conservation and Regeneration. Advances in Agronomy* (Elsevier Inc., 2021).
- Begill, N., Don, A. & Poeplau, C. No detectable upper limit of mineral-associated organic carbon in temperate agricultural soils. *Glob. Chang. Biol.* **29**, 4662–4669 (2023).
- Lugato, E., Bampa, F., Panagos, P., Montanarella, L. & Jones, A. Potential carbon sequestration of European arable soils estimated by modelling a comprehensive set of management practices. *Glob. Chang. Biol.* **20**, 3557–3567 (2014).
- Angst, G. et al. Unlocking complex soil systems as carbon sinks: multi-pool management as the key. *Nat. Commun.* **14**, 2967 (2023).
- De Rosa, D. et al. Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade? *Glob. Chang. Biol.* **30**, e16992 (2024).
- Council, E. *Decision No 1386/2013/EU of the European Parliament and of the Council. Off. J. Eur. Union* 171–200 (2013).
- Lavalley, 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).
- Kleber, M. et al. Mineral-organic associations: formation, properties, and relevance in soil environments. *Adv. Agron.* **130**, 1–140 (2015).
- Hassink, J. Effects of soil texture and grassland management on soil organic C and N and rates of C and N mineralization. *Soil Biol. Biochem.* **26**, 1221–1231 (1994).
- Gulde, S., Chung, H., Amelung, W., Chang, C. & Six, J. Soil carbon saturation controls labile and stable carbon pool dynamics. *Soil Sci. Soc. Am. J.* **72**, 605–612 (2008).
- Castellano, M. J., Kaye, J. P., Lin, H. & Schmidt, J. P. Linking carbon saturation concepts to nitrogen saturation and retention. *Ecosystems* **15**, 175–187 (2012).
- Feng, W., Plante, A. F. & Six, J. Improving estimates of maximal organic carbon stabilization by fine soil particles. *Biogeochemistry* **112**, 81–93 (2013).
- Georgiou, K. et al. Global stocks and capacity of mineral-associated soil organic carbon. *Nat. Commun.* **13**, 3797 (2022).
- Ingram, J. S. I. & Fernandes, E. C. M. Managing carbon sequestration in soils: concepts and terminology. *Agric. Ecosyst. Environ.* **87**, 111–117 (2001).
- Stewart, C. E., Paustian, K., Conant, R. T., Plante, A. F. & Six, J. Soil carbon saturation: concept, evidence and evaluation. *Biogeochemistry* **86**, 19–31 (2007).
- Schmidt, M. W. I. et al. Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
- Stewart, C. E., Plante, A. F., Paustian, K., Conant, R. T. & Six, J. Soil carbon saturation: linking concept and measurable carbon pools. *Soil Sci. Soc. Am. J.* **72**, 379–392 (2008).
- Doetterl, S. et al. Soil carbon storage controlled by interactions between geochemistry and climate. *Nat. Geosci.* **8**, 780–783 (2015).
- Viscarra Rossel, R. A. et al. Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls. *Nat. Geosci.* **12**, 547–552 (2019).
- Wiesmeier, M. et al. Soil organic carbon storage as a key function of soils—a review of drivers and indicators at various scales. *Geoderma* **333**, 149–162 (2019).
- Viscarra Rossel, R. A. et al. How much organic carbon could the soil store? the carbon sequestration potential of Australian soil. *Glob. Chang. Biol.* **30**, e17053 (2024).
- Schweizer, S. A., Mueller, C. W., Höschel, C., Ivanov, P. & Kögel-Knabner, I. The role of clay content and mineral surface area for soil

- organic carbon storage in an arable toposequence. *Biogeochemistry* **156**, 401–420 (2021).
26. Cotrufo, M. F., Lavalley, J. M., Six, J. & Lugato, E. The robust concept of mineral-associated organic matter saturation: A letter to Begill et al., 2023. *Glob. Change Biol.* **29**, 5986–5987 (2023).
  27. 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).
  28. Toth, G., Jones, A., Montanarella, L., Alewell, C. LUCAS topsoil survey-methodology, data and results. (2013).
  29. Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. & Fernández-Ugalde, O. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *Eur. J. Soil Sci.* **69**, 140–153 (2018).
  30. Stevens, A., Nocita, M., Tóth, G., Montanarella, L. & van Wesemael, B. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS One* **8**, <https://doi.org/10.1371/journal.pone.0066409> (2013).
  31. Ara Begum, R. et al. *Point of Departure and Key Concepts. In: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor].* (2022).
  32. Pacini, L. et al. A new approach to estimate soil organic carbon content targets in European croplands topsoils. *Sci. Total Environ.* **900**, 165811 (2023).
  33. Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Sci. Data* **5**, 170191(2018).
  34. Running, S. & Zhao, M. MODIS/Terra net primary production gap-filled yearly L4 global 500m SIN grid V061. 2021. *NASA EOSDIS Land Process. DAAC* <https://doi.org/10.5067/MODIS/MOD17A3HGF.061> (2021).
  35. Iwahashi, J. & Pike, R. J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* **86**, 409–440 (2007).
  36. Poeplau, C., Dechow, R., Begill, N. & Don, A. Towards an ecosystem capacity to stabilise organic carbon in soils. *Glob. Change Biol.* **30**, e17453 (2024).
  37. 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**, 6713 (2021).
  38. Rocci, K. S., Lavalley, J. M., Stewart, C. E. & Cotrufo, M. F. Soil organic carbon response to global environmental change depends on its distribution between mineral-associated and particulate organic matter: a meta-analysis. *Sci. Total Environ.* **793**, 148569 (2021).
  39. García-Palacios, P. et al. Dominance of particulate organic carbon in top mineral soils in cold regions. *Nat. Geosci.* **17**, 145–150 (2024).
  40. Prairie, A. M., King, A. E. & Francesca Cotrufo, M. Restoring particulate and mineral-associated organic carbon through regenerative agriculture. *Proc. Natl Acad. Sci. Usa.* **120**, e2217481120 (2023).
  41. King, A. E. et al. A soil matrix capacity index to predict mineral-associated but not particulate organic carbon across a range of climate and soil pH. *Biogeochemistry* **165**, 1–14 (2023).
  42. Six, J., Doetterl, S., Laub, M., Müller, C. R. & Van De Broek, M. The six rights of how and when to test for soil C saturation. *SOIL* **10**, 275–279 (2024).
  43. Kalks, F. et al. Geogenic organic carbon in terrestrial sediments and its contribution to total soil carbon. *SOIL* **7**, 347–362 (2021).
  44. 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).
  45. Smith, P. Managing the global land resource. *Proc. R. Soc. B Biol. Sci.* **285**, 20172798 (2018).
  46. Spohn, M. Increasing the organic carbon stocks in mineral soils sequesters large amounts of phosphorus. *Glob. Chang. Biol.* **26**, 4169–4177 (2020).
  47. Davies, C. A., Robertson, A. D. & McNamara, N. P. The importance of nitrogen for net carbon sequestration when considering natural climate solutions. *Glob. Change Biol.* **27**, 218–219 (2021).
  48. Demenois, J. et al. Barriers and strategies to boost soil carbon sequestration in agriculture. *Front. Sustain. Food Syst.* **4**, <https://doi.org/10.3389/fsufs.2020.00037> (2020).
  49. Viscarra Rossel, R. A. et al. A global spectral library to characterize the world's soil. *Earth-Sci. Rev.* **155**, 198–230 (2016).
  50. Cambardella, C. A. & Elliott, E. T. Particulate soil organic-matter changes across a grassland cultivation sequence. *Soil Sci. Soc. Am. J.* **56**, 777–783 (1992).
  51. Bowers, S. A. & Hanks, R. J. Reflection of radiant energy from soils. *Soil Sci.* **100**, 130–138 (1965).
  52. Summerauer, L. et al. The central African soil spectral library: a new soil infrared repository and a geographical prediction analysis. *SOIL* **7**, 693–715 (2021).
  53. Shenk, J. S., Westerhaus, M. O. & Berzaghi, P. Investigation of a LOCAL calibration procedure for near infrared instruments. *J. Infrared Spectrosc.* **5**, 223–232 (1997).
  54. Ramirez-Lopez, L. et al. Distance and similarity-search metrics for use with soil vis-NIR spectra. *Geoderma* **199**, 43–53 (2013).
  55. Martens, H. & Næs, T. *Multivariate Calibration.* in *Chemometrics* 147–156 (Springer Netherlands, 1984).
  56. Leifeld, J. Application of diffuse reflectance FT-IR spectroscopy and partial least-squares regression to predict NMR properties of soil organic matter. *Eur. J. Soil Sci.* **57**, 846–857 (2006).
  57. Dangal, S., Sanderman, J., Wills, S. & Ramirez-Lopez, L. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. *Soil Syst.* **3**, 11 (2019).
  58. Lin, L. I.-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255 (1989).
  59. Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J. M. & McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC - Trends Anal. Chem.* **29**, 1073–1081 (2010).
  60. Reeves, J. B., Follett, R. F., McCarty, G. W. & Kimble, J. M. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? *Commun. Soil Sci. Plant Anal.* **37**, 2307–2325 (2006).
  61. Knox, N. M. et al. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma* **239–240**, 229–239 (2015).
  62. Greenberg, I., Seidel, M., Vohland, M. & Ludwig, B. Performance of field-scale lab vs in situ visible/near- and mid-infrared spectroscopy for estimation of soil properties. *Eur. J. Soil Sci.* **73**, (2022).
  63. Ramifehiarivo, N. et al. Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon fractions in Madagascar agricultural soils. *Geoderma Reg.* **33**, (2023).
  64. Hartigan, J. A. C., Wong, M. Algorithm AS 136: A k-means clustering algorithm. *JSTOR* (1979).
  65. Ballabio, C., Panagos, P. & Montanarella, L. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma* **11**, e0152098 (2016).
  66. Ballabio, C. et al. Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma* **355**, 113912 (2019).
  67. Muggeo, V. segmented: an R package to fit regression models with broken-line relationships. *R. N.* **8**, 20–25 (2008).

68. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
69. Stevens A. & Ramirez-Lopez L. An introduction to the prospectr package. *R package version 0.2.7*. <https://cran.r-project.org/web/packages/prospectr/vignettes/prospectr.html> (2024).
70. Robinson, D., Hayes, A. & Couch, S. Broom: convert statistical objects into tidy tibbles, 2021. *R package version 0.7* (2022).
71. Koenker, R. quantreg: Quantile regression. *R package version 5.97* <https://cran.r-project.org/package=quantreg> (2023).
72. Liland, K. H., Mevik, B. H. & Wehrens, R. pls: Partial Least Squares and Principal Component regression. *R package version 2.7-1* <https://cran.r-project.org/package=pls> (2022).
73. Kühn, M. Package 'caret' - Classification and Regression Training. CRAN Repository at <https://ui.adsabs.harvard.edu/abs/2015ascl.soft05003K/abstract> (2019).
74. Ramirez-lopez, L. et al. resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics. *R package version 2.2.3* <https://cran.r-project.org/web/packages/resemble/index.html> (2024).
75. Wood, S. N. Package 'mgcv' - R package. <http://cran.uib.no/web/packages/mgcv/mgcv.pdf> (2023) <https://doi.org/10.1201/9781315370279>.
76. Fasiolo, M., Nedellec, R., Goude, Y. & Wood, S. N. Scalable Visualization Methods for Modern Generalized Additive Models. *J. Comput. Graph. Stat.* **29**, 78–86 (2020).
77. Lenth, R. emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.10.1* <https://cran.r-project.org/package=emmeans> (2024).
78. Hijmans, R. J., van Etten, J. & Cheng, J. Raster: Geographic Data Analysis and Modeling. *R package version 3.6-20* <https://cran.r-project.org/package=raster> (2023).
79. Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R. Spatial Data Science: With Applications in R* (CRC Press, 2023).
80. Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. cluster: Cluster Analysis Basics and Extensions. *R package version 2.1.2* <https://cran.r-project.org/web/packages/cluster/index.html> (2021).
81. Massicotte, P. & South, A. rnaturalearth: World Map Data from Natural Earth. *R package version 0.3.2* <https://cran.r-project.org/package=rnaturalearth> (2023).
82. Pedersen, T. patchwork: The Composer of Plots. *R package version 1.2.0* <https://cran.r-project.org/package=patchwork> (2024).
83. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, New York, 2016).
- Climate Action (DG-CLIMA). We thank Beatrice Landoni and Christopher Havenga for their help with figure formatting.

## Author contributions

T.S.B. led the study's design, data analyses, interpretations and writing; D.D.R. contributed to the study's design, data analysis, interpretations and writing; P.P. contributed to interpretations and writing; M.F.C. contributed to interpretations and writing; A.J. conceived the funding; E.L. led the study's design and contributed to data analysis, interpretation and writing.

## 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/s41467-025-57355-y>.

**Correspondence** and requests for materials should be addressed to T. S. Breure.

**Peer review information** *Nature Communications* thanks Yang Lin, who co-reviewed with Ryan Champiny, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The European Union 2025, corrected publication 2026

## Acknowledgements

The LUCAS Soil sample collection is supported by EUROSTAT and the following Directorate-Generals of the European Commission: Environment (DG-ENV), Agriculture and Rural Development (DG-AGRI) and