nature geoscience

Soil carbon storage informed by particulate and mineral-associated organic matter

M. Francesca Cotrufo [ID][1]*, Maria Giovanna Ranalli[2], Michelle L. Haddix[1], Johan Six[3] and
Emanuele Lugato [ID][4]

Effective land-based solutions to climate change mitigation require actions that maximize soil carbon storage without generating surplus nitrogen. Land management for carbon sequestration is most often informed by bulk soil carbon inventories, without considering the form in which carbon is stored, its capacity, persistency and nitrogen demand. Here, we present coupling of European-wide databases with soil organic matter physical fractionation to determine continental-scale forest and grassland topsoil carbon and nitrogen stocks and their distribution between mineral-associated and particulate organic matter pools. Grasslands and arbuscular mycorrhizal forests store more soil carbon in mineral-associated organic carbon, which is more persistent but has a higher nitrogen demand and saturates. Ectomycorrhizal forests store more carbon in particulate organic matter, which is more vulnerable to disturbance but has a lower nitrogen demand and can potentially accumulate indefinitely. The share of carbon between mineral-associated and particulate organic matter and the ratio between carbon and nitrogen affect soil carbon stocks and mediate the effects of other variables on soil carbon stocks. Understanding the physical distribution of organic matter in pools of mineral-associated versus particulate organic matter can inform land management for nitrogen-efficient carbon sequestration, which should be driven by the inherent soil carbon capacity and nitrogen availability in ecosystems.

The impact of increasing atmospheric carbon dioxide (CO_2) concentration on global climate[1] calls for effective atmospheric CO_2 capture strategies. Sequestration of carbon (C) in soil organic matter (SOM) is considered one of those strategies[2]. It can also bring important co-benefits such as improving soil health and providing soil services[3]. Temperate forest and grassland soils occupy a vast land surface area[4], are often managed and can store large amounts of C[5]. Thus they can play a critical role in future soil C management.

Organic C in soil is stored in a myriad of different chemical compounds, many of which contain nitrogen (N) and/or are formed through microbial activity that demands N. Furthermore, SOM requires more N per unit C than does plant biomass. Thus, the ability of soils to store C is linked to N availability. Increasing soil C storage without increasing N fertilization or immobilizing N in soil and thus impacting plant productivity is a major challenge for soil C sequestration strategies[6,7]. Soil C/N ratio has been suggested as an indicator for the potential of soils to store C, with systems characterized by a higher soil C/N ratio being able to accrue more C per unit N[8]. Following this logic, it has been suggested that ectomycorrhizal (ECM) systems, characterized by a higher C/N ratio, have a higher potential for C storage than arbuscular mycorrhizal (AM) systems[8]. However, the effect of N availability on soil C storage remains controversial, and both increased and decreased soil C storage have been observed as a result of long-term N additions[9,10]. Recent frameworks[11–13] suggest that soil C accrual, persistence and response to N availability can be better described if SOM is broadly divided into a particulate organic matter (POM) and a mineral-associated organic matter (MAOM) pool. POM, being predominantly of plant origin, contains many structural C compounds with low N content[14] and persists in soil through inherent biochemical recalcitrance, physical protection in aggregates and/or microbial inhibition. MAOM is

largely made of microbial products richer in N and persists in soil because of chemical bonding to minerals and physical protection in small aggregates[15]. These two fractions can be analytically separated by size and/or density and show consistent differences in turnover times, with POM being more vulnerable to disturbance and cycling faster than MAOM[16,17]. In this study we used size separation and defined MAOM as the SOM fraction <53 μm and POM as the SOM fraction 53–2,000 μm (refs. [17,18]).

Because of the contrasting nature of these two SOM pools, we posit that soil C/N ratio and C sequestration potentials can be better understood through a framework based on the separation between POM and MAOM. For soil C/N ratio, for example, we can apply a mass balance approach described as:

$$C/N_{SOM} = C/N_{MAOM} \times f_{MAOM} + C/N_{POM} \times (1 - f_{MAOM})$$

where C/N_{SOM}, C/N_{POM} and C/N_{MAOM} are the C/N ratios of the total SOM, POM and MAOM, respectively, and f_{MAOM} is the MAOM proportion of the total SOM. The proportion of POM is $1 - f_{MAOM}$. Given that MAOM accrual is expected to be limited by mineral surface availability[18,19], we posit that variation in soil C/N ratio and the ability of soils to sequester C is related to f_{MAOM}, C/N_{POM} and C/N_{MAOM}. Elucidating their drivers and impact on soil C storage can provide a useful guidance for soil C sequestration strategies.

We used the Land Use/Land Cover Area Frame Survey (LUCAS) database[20] to determine topsoil C and N storage in European forests and grasslands on 9,415 georeferenced points. We limited our analyses to soils with an organic C content lower than 12% to avoid organic soils. The LUCAS survey is restricted to the topsoil (0–20 cm), which represents 40–50% of the total soil C along the profile of grasslands and forest soils worldwide[5]. For a subset of soils ($n = 186$) representative of the full dataset (Supplementary Fig. 1),

[1]Department of Soil and Crop Science and Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, USA. [2]Department of Political Science—Statistics Section, University of Perugia, Perugia, Italy. [3]Department of Environmental Systems Science, Swiss Federal Institute of Technology, Zurich, Switzerland. [4]European Commission, Joint Research Centre, Ispra, Italy. *e-mail: Francesca.cotrufo@colostate.edu
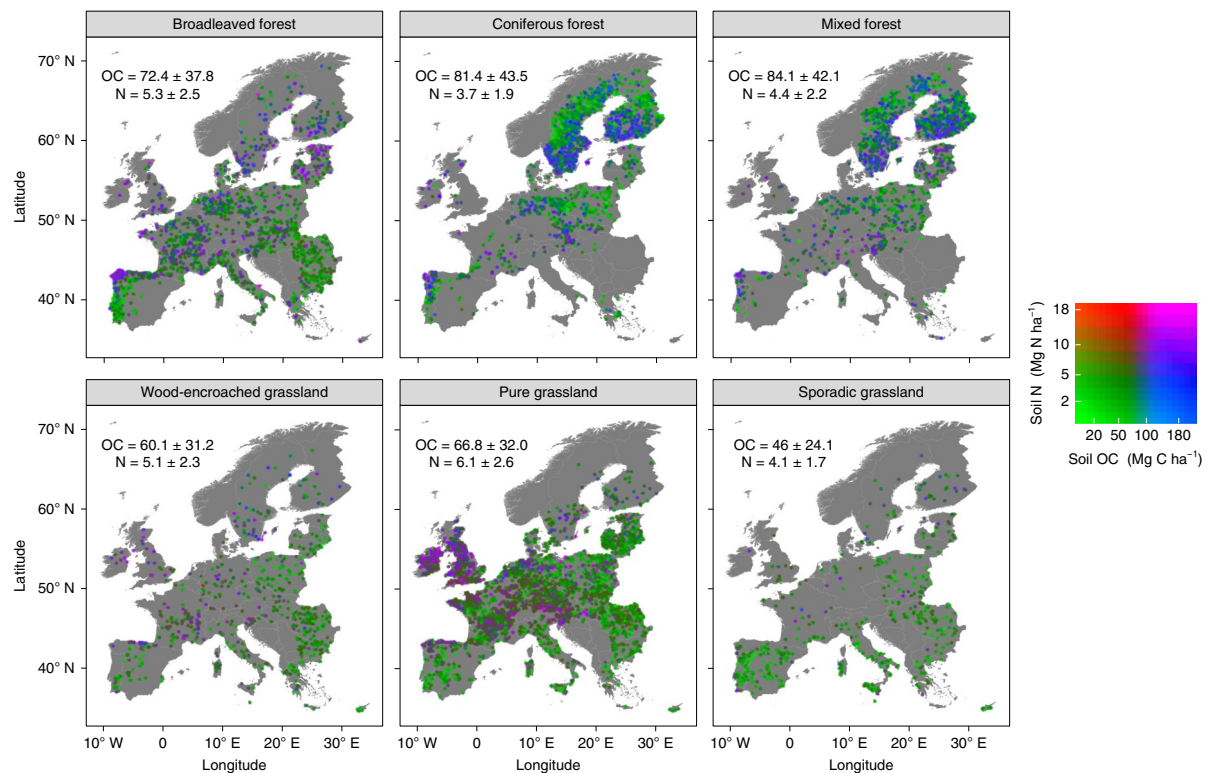
Fig. 1 | Geographical distribution of soil OC and N stocks in the top mineral soil (0–20 cm) of European forests and grasslands. The six panels represent the LUCAS points in broadleaved ($n = 1,617$), coniferous ($n = 1,713$) and mixed ($n = 1,399$) forests and wood-encroached ($n = 635$), pure ($n = 3,533$) and sporadic ($n = 518$) grasslands. Mean OC and N stocks (Mg ha^{-1}) and their standard deviations for each land cover are reported in the top left corner of each plot.

we separated SOM into POM and MAOM by size (53 μm) and measured their C and N content. We used these measured data to predict the MAOM and POM fraction share of the SOM and their C and N content for the full dataset using the random forest (RF) approach (Supplementary Figs. 2–7). We assessed the role of land cover and plant–mycorrhizal associations on the relative contribution of MAOM and POM to SOM and their C/N ratios. We used coniferous, mixed and broadleaved forests and wood-encroached, sporadic and pure grasslands as categories for land cover. For broadleaved and mixed forests, we identified tree species composition and their mycorrhizal associations by coupling the LUCAS database with a high-resolution pan-European tree occurrence dataset[21]. Finally, we used path analyses, based on the 186 samples for which we had independent measured values, to investigate drivers of soil C stocks as affected by f_{MAOM}, C/N$_{POM}$ and C/N$_{MAOM}$ (Methods).

Soil organic C and N storage

Top mineral soil (0–20 cm) organic C and N storage varied geographically and by land-use cover (Fig. 1). Overall, mixed and coniferous forests had the highest average organic C stocks compared with broadleaved forests and grasslands. Pure grasslands had the highest average N stocks. However, these differences among land covers may also be due to the generally higher proportion of soil organic carbon (OC) in the topsoil (0–20 cm) of forests (50% of their total OC stock) compared with grasslands (42%)[5].

Soil C/N ratio was relatively well constrained in European forest and grassland top mineral soils. The overall soil C/N average was 15.0 ± 6.5, which is within the distribution of C/N averages (9.9 to 25.8) found for world soils[22]. Coniferous and mixed forest soils had the highest and more variable soil C/N (22.5 ± 7.1 and 20.0 ± 6.2, respectively), while broadleaved forests (13.8 ± 4.0) and pure grasslands (11.0 ± 2.1) had lower and more constrained C/N than

coniferous and mixed forests (Fig. 2), as shown by their smaller standard deviations. Soil C/N ratio decreased with increasing f_{MAOM} across all land-use classes and, as expected, increased with increasing C/N$_{POM}$ and C/N$_{MAOM}$ (Fig. 2). Overall, the C/N of MAOM (12.6 ± 4.7) was lower and less variable than that of POM (22.1 ± 14.9; Fig. 2 and Supplementary Fig. 7). While plants vary substantially in their C/N ratio, both among and within different plant species and plant organs and in response to environmental stressors[23], soil microbes have a much more constrained C/N ratio, with fungi C/N ratio typically ranging between 4.5 and 15 and bacteria C/N ratio ranging between 3 and 5 (ref. [24]). It is thus expected that POM, consisting largely of partly decomposed plant material, has a more variable C/N ratio than MAOM (Fig. 2), which is mostly of microbial origin. Because woody inputs are characterized by high C/N and persistence in POM[25] and are degraded by fungi, forests have generally higher C/N in soil, POM and MAOM relative to grasslands (Fig. 2). In forests, the C/N$_{MAOM}$ exceeds the microbial range, indicating a higher contribution of plant-derived OM to the MAOM fraction in these systems, probably through ex vivo microbial transformations of plant inputs[26] or occlusions of plant-derived particulate structures in small (<53 μm) aggregates[27].

The type of mycorrhizal associations may also be an important driver of soil C/N and f_{MAOM}, due to the different degradability of their own tissues and their differential ability to degrade organic matter and release mineral N. ECM fungi produce lytic enzymes[28], while arbuscular mycorrhizal fungi depend on saprotrophic organic matter decay and production of mineral N[29]. At a global scale, soils of ecosystems dominated by ecto- and erycoid mycorrhiza contained more C per unit N than soils in ecosystems dominated by arbuscular mycorrhiza[8]. However, in the framework of the plant economic spectrum, plant traits directly control the decomposition of plant residues, thus impacting C and N cycling in soil[30].
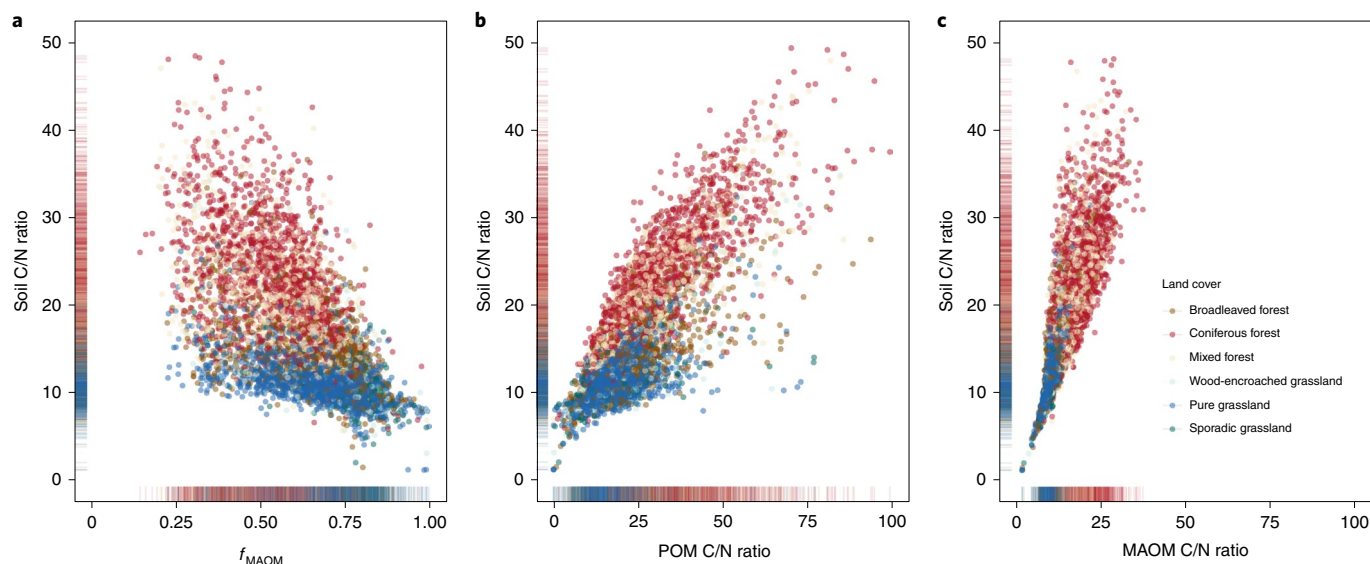
Fig. 2 | Soil C/N ratios in the top 20 cm of mineral soils of European forests and grasslands. a–c, Values are reported against the proportion of C in MAOM relative to total soil organic matter (f_{MAOM}) (**a**), the POM C/N ratio (**b**) and the MAOM C/N ratio (**c**). MAOM is defined as the soil organic matter fraction <53 µm, and POM is 53–2,000 µm, after aggregate dispersion. The coloured bands along the axes represent value distributions by land cover.
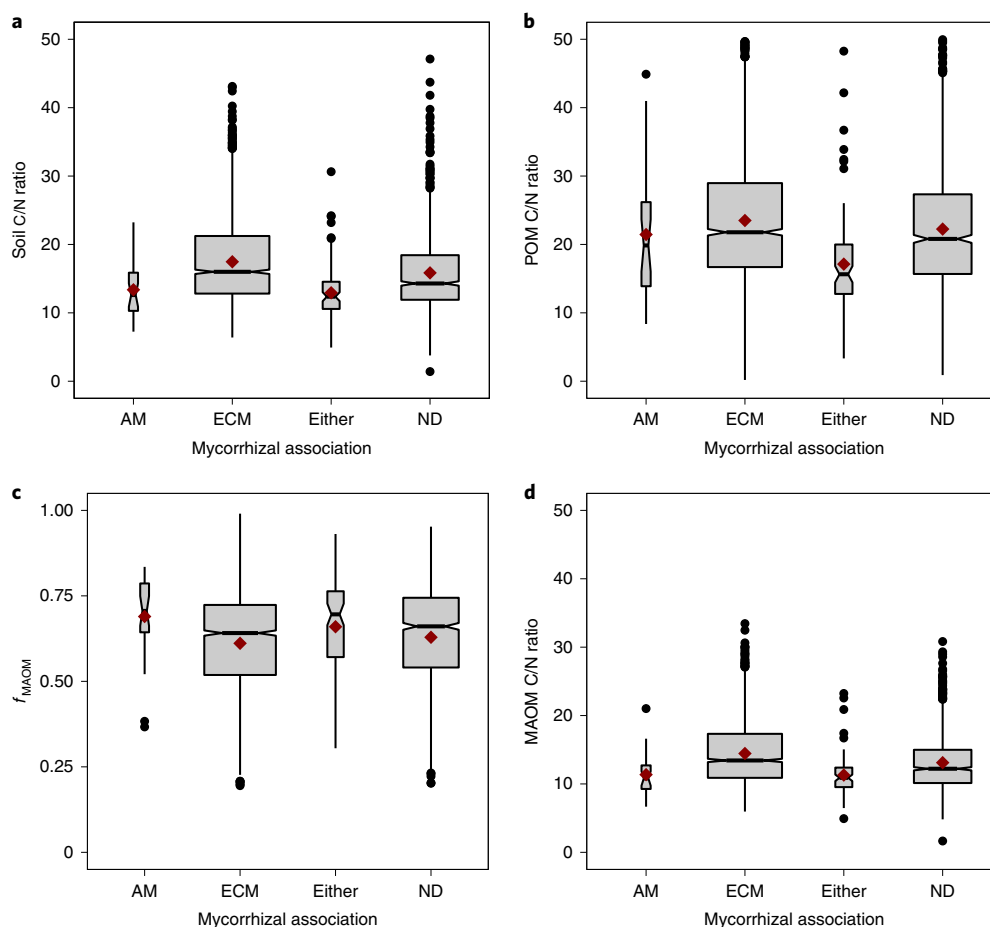

Fig. 3 | Box plots of SOM pools and C/N by mycorrhizal association. a–d, Box plots of soil C/N (**a**), POM C/N (**b**), f_{MAOM} (**c**) and MAOM C/N (**d**) in broadleaved and mixed forests, by mycorrhizal association. See Fig. 2 legend for POM and MAOM definitions. AM ($n=27$) is arbuscular mycorrhiza; ECM ($n=1,776$) is ectomycorrhiza; Either ($n=93$) represents AM or ECM; and ND ($n=1,097$) defines the points without univocal association. Box plots report the median, first and third quartiles, lower whisker ($\leq-1.5\times$ interquartile range) and upper whisker ($\geq+1.5\times$ interquartile range). Red dots represent mean values. Estimated mean differences and confidence intervals are reported in Supplementary Table 2.
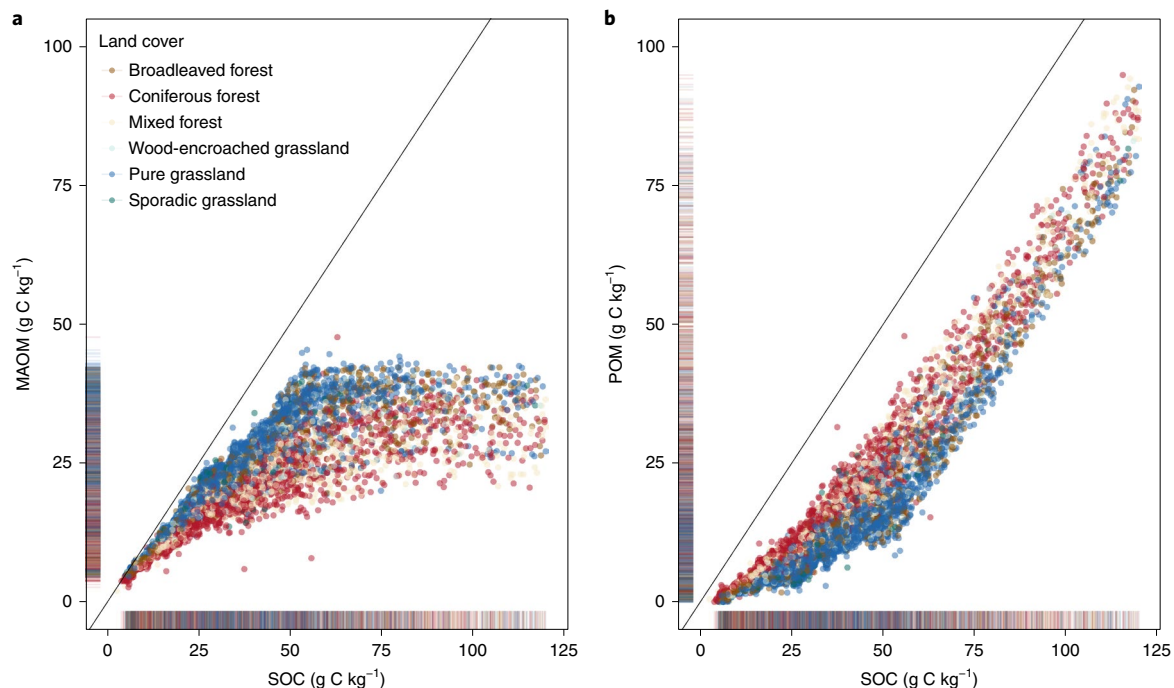
Fig. 4 | Soil organic C in MAOM and POM. a,b, Distribution of soil organic C in MAOM (**a**) and POM (**b**) in relation to total soil organic C content for European grassland and forest soils. Lines represent the 1:1 relationship.

Plant traits and mycorrhizal associations may be interconnected, and it is still an open question whether there is a causal relationship between either of them and soil C/N ratio or C storage[31]. Therefore, mycorrhizal effects on SOM and its C/N ratio are better assessed within the same land-cover class; broadleaved forests offer this opportunity since their trees may associate with either ecto- or arbuscular mycorrhiza[32]. When we compared soil C/N, C/N_{POM}, C/N_{MAOM}, f_{MAOM} and C stocks in MAOM and POM across broadleaved and mixed forests for which we had mycorrhizal association, we observed on average 24% lower soil C/N in soils with AM compared with ECM (Fig. 3 and Supplementary Table 2). This was associated with a decrease in C/N_{MAOM} and an increase in f_{MAOM} (Fig. 3) and overall higher C stocks in MAOM (Supplementary Table 2). These findings confirm and generalize recent observations of microbial residue accumulation, resulting in higher soil N in MAOM in AM-dominated temperate forests compared with temperate forests dominated by ECM[33], which store more C in POM (Supplementary Table 2). Overall, these differences between ECM and AM broadleaved forest soils resulted in higher average soil organic C stocks in ECM compared with AM systems (Supplementary Fig. 8), in line with global trends[8]. However, the difference in soil C stocks between ECM and AM was not statistically significant (Supplementary Table 3). This study assessed only mineral soils (soil organic C < 12%). It is likely that ECM forests would accumulate more POM in their organic soil layer, resulting in overall more soil C in these forests[8].

Implications for soil C sequestration

Average soil C stocks in European grasslands and forest soils varied between 46 and 84 Mg C ha^{-1}, with the highest stocks in coniferous or mixed forest (Fig. 1). This information, however, does not allow us to project the additional C storage capacity of these systems, which can be better informed by the relative distribution of C between MAOM and POM. In fact, across all sites, C storage in MAOM and POM fractions showed different dynamics with increasing total soil C content (Fig. 4). Storage of C in MAOM dominates in soils with

relatively low C content. However, it saturates[19], at which point additional C storage is only realized through POM accrual. Across the entire soil organic C range, grasslands systems consistently have proportionally more MAOM and less POM than forests, in particular compared with coniferous forests (Fig. 4 and Supplementary Table 1). Independent of land cover, the majority of the sites (80%) are below the flex point in the relationship of MAOM to soil organic C (at 50 g soil organic C kg^{-1} soil), indicating a very large capacity for C accrual. We conducted a path analysis to understand whether typical environmental control of soil C storage[3] affects it directly or through f_{MAOM}, C/N_{POM} and C/N_{MAOM} variables. Because of C saturation, soil C stocks are highly and inversely related to f_{MAOM} (Fig. 5). A highly significant positive relation exists between C/N_{MAOM} and soil C stocks, probably because MAOM is the dominant SOM pool (that is, $f_{MAOM} > 50\%$; Fig. 2) in most soils, while, surprisingly, C/N_{POM} appeared to be less important and negatively related to soil C stocks (Fig. 5). This finding demonstrates that at the continental scale, the N efficiency of C sequestration, (the amount of N required to sequester a unit of C) depends on the share between MAOM and POM, and their C/N ratios, as previously shown for broadleaved forests in the United States[33]. In another study conducted in a few European grasslands and forest soils[34], POM was identified as a strong predictor for both organic C and N, indicating a significant role of this largely plant derived and relatively vulnerable fraction for C storage in these ecosystems. In addition to land cover, soil edaphic factors (for example, soil texture and pH) control C storage through affecting f_{MAOM} and C/N_{MAOM} (Fig. 5 and Supplementary Table 4). Of interest are the negative correlations between C/N_{MAOM} and pH and between C/N_{MAOM} and silt and clay, possibly suggesting lower bacteria-derived organic matter contribution to MAOM in coarse-texture acidic soils[35].

Implementation of soil C sequestration strategies requires clear science-based guidelines that consider site-specific soil and ecosystem properties, including the relative distribution of SOM into MAOM and POM. Here, we show that C sequestration in grasslands is highly persistent but requires a lot of N and saturates because of the dominance of MAOM-C in these systems. Thus, management
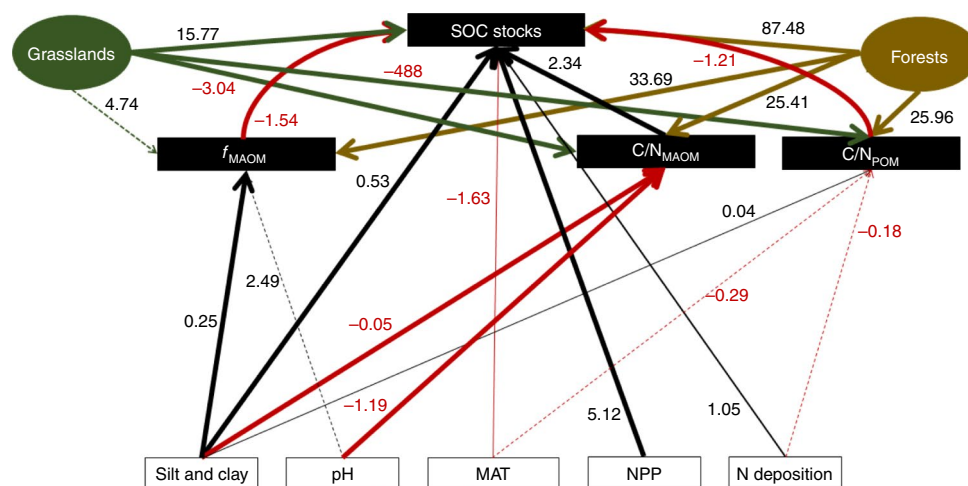
Fig. 5 | A schematic representation of the path analyses used to identify the controls on SOC stocks. Controls are mediated by the proportion of the MAOM to total organic matter (f_{MAOM}) and the C/N ratio of POM (C/N$_{POM}$) and MAOM (C/N$_{MAOM}$). Forests and grasslands are covariates. The factors reported had at least one significant correlation. Black and red arrows and/or coefficients denote positive and negative relationships, respectively. Arrow width is proportional to significance (*P* values) of regression. Regression coefficient, standard errors, *P* value and 95% confidence intervals are reported in Supplementary Table 4. NPP, net primary production.

for C accrual in grasslands should target soils below their saturation level, pointing to the need for geographical estimates of soil C saturation deficits[36]. Forests, however, have a greater plasticity in how soil C may be accrued because they can store more C in the less persistent and more vulnerable POM fraction. Afforestation for soil C sequestration should be designed on the basis of the soil properties (for example, silt and clay content, pH), C deficit and N availability and should use AM or EMC-associated tree species to maximize C accrual.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41561-019-0484-6.

References

1. IPCC *Climate Change 2013: The Physical Science Basis* (eds Stocker T. F. et al.) (Cambridge Univ. Press, 2013).
2. Chabbi, A. et al. Aligning agriculture and climate policy. *Nat. Clim. Change* **7**, 307–309 (2017).
3. 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).
4. Gong, P. et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **34**, 2607–2654 (2013).
5. Jobbagy, 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).
6. Janzen, H. The soil carbon dilemma: shall we hoard it or use it? *Soil Biol. Biochem.* **38**, 419–424 (2006).
7. Schlesinger, W. H. & Amundson, R. Managing for soil carbon sequestration: let's get realistic. *Glob. Change Biol.* **25**, 386–389 (2018).
8. Averill, C., Turner, B. L. & Finzi, A. C. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature* **505**, 543–545 (2014).
9. Zak, D. R., Freedman, Z. B., Upchurch, R. A., Steffens, M. & Kögel-Knabner, I. Anthropogenic N deposition increases soil organic matter accumulation without altering its biochemical composition. *Glob. Change Biol.* **23**, 933–944 (2017).
10. Mack, M. C., Schuur, E. A. G., Bret-Harte, M. S., Shaver, G. R. & Chapin, F. S. I. I. Ecosystem carbon storage in Arctic tundra reduced by long-term nutrient fertilization. *Nature* **431**, 440–443 (2004).
11. Cotrufo, M. F. et al. Soil organic matter formation from biochemical and physical pathways of litter mass loss. *Nat. Geosci.* **8**, 776–779 (2015).
12. Castellano, M., Mueller, K., Olk, D., Sawyer, J. & Six, J. Integrating plant litter quality, soil organic matter stabilization and the carbon saturation concept. *Glob. Change Biol.* **21**, 3200–3209 (2015).
13. Averill, C. & Waring, B. Nitrogen limitation of decomposition and decay: how can it occur? *Glob. Change Biol.* **24**, 1417–1427 (2018).
14. Golchin, A., Oades, J., Skjemstad, J. & Clarke, P. Study of free and occluded particulate organic matter in soils by solid state 13C Cp/MAS NMR spectroscopy and scanning electron microscopy. *Soil Res* **32**, 285–309 (1994).
15. Kogel-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).
16. Poeplau, C. et al. Isolating soil organic carbon fractions with varying turnover rates—a comprehensive comparison of fractionation schemes. *Soil Biol. Biochem.* **125**, 10–26 (2018).
17. 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).
18. 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).
19. 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).
20. Toth, G., Jones, A. & Montanarella, L. *LUCAS Topsoil Survey Methodology, Data and Results* (European Union, 2013).
21. Mauri, A., Strona, G. & San-Miguel-Ayanz, J. EU-Forest, a high-resolution tree occurrence dataset for Europe. *Sci. Data* **4**, 160123 (2017).
22. Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **65**, 10–21 (2014).
23. Sterner, R. W. & Elser, J. J. *Ecological Stochiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton Univ. Press, 2002).
24. Paul, E. A. *Soil Microbiology, Ecology and Biochemistry* (Academic, 2014).
25. Weedon, J. T. et al. Global meta-analysis of wood decomposition rates: a role for trait variation among tree species? *Ecol. Lett.* **12**, 45–56 (2009).
26. Liang, C., Schimel, J. P. & Jastrow, J. D. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol* **2**, 17105 (2017).
27. Mueller, C. W. et al. Submicron scale imaging of soil organic matter dynamics using NanoSIMS—from single particles to intact aggregates. *Org. Geochem.* **42**, 1476–1488 (2012).
28. Fernandez, C. W., Langley, J. A., Chapman, S., McCormack, M. L. & Koide, R. T. The decomposition of ectomycorrhizal fungal necromass. *Soil Biol. Biochem.* **93**, 38–49 (2016).

29. Gadgil, R. L. & Gadgil, P. D. Mycorrhiza and litter decomposition. *Nature* **233**, 133 (1971).

30. Freschet, G. T., Aerts, R. & Cornelissen, J. H. C. A plant economics spectrum of litter decomposability. *Funct. Ecol.* **26**, 56–65 (2012).

31. Van Groenigen, J. W. et al. The soil N cycle: new insights and key challanges. *SOIL* **1**, 235–256 (2015).

32. Phillips, R. P., Brzostek, E. & Midgley, M. G. The mycorrhizal-associated nutrient economy: a new framework for predicting carbon–nutrient couplings in temperate forests. *New Phytol.* **199**, 41–51 (2013).

33. Craig, M. E. et al. Tree mycorrhizal type predicts within-site variability in the storage and distribution of soil organic matter. *Glob. Change Biol.* **24**, 3317–3330 (2018).

34. Denef, K., Del Galdo, I., Venturi, A. & Cotrufo, M. F. Soil carbon and nitrogen stocks and distribution across organic matter fractions in European soils under different land use. *Open J. Soil Sci.* **3**, 297–313 (2013).

35. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **15**, 579–590 (2017).

36. Angers, D., Arrouays, D., Saby, N. & Walter, C. Estimating and mapping the carbon saturation deficit of French agricultural topsoils. *Soil Use Manag.* **27**, 448–452 (2011).

Methods

Soil and land-cover data. Soil data were extracted from the LUCAS database[20]. LUCAS is a land-use/cover survey consisting of 200,000 georeferenced points selected from a regular 2 km grid covering the European Union land. A Latin hypercube stratified random sampling was applied to select ~20,000 soil sampling locations representative of European landscape features. Within each georeferenced location, a first subsample was taken by sampling the topsoil, after removal of the litter layer, to a depth of 20 cm. Four other similar samples were then taken at a distance of 2 m from the original sample, following the four cardinal directions. The five subsamples were subsequently composited, and 500 g of the composite was taken as the final sample. All samples were sieved through a 2 mm mesh to remove organic fragments and analysed for coarse fragments (%), particle size distribution (% clay, silt and sand), pH, soil OC (SOC, g kg^{-1}), carbonates (CaCO$_3$, g kg^{-1}), total N (g kg^{-1}), available phosphorus (P, mg kg^{-1}), extractable potassium (K, mg kg^{-1}) and cation exchange capacity (CEC, cmol$^{(+)}$ kg^{-1}) using International Organization for Standardization protocols.

For this study, we restricted the analyses to the forest and grassland land-cover data and considered only mineral soils, defined as soils with an organic C concentration smaller than 120 g kg^{-1} (ref. [20]), for a total of 9,415 points.

In the database, forest and grassland sites are each separated in three land-cover categories assigned by the surveyor when visiting the site according to a predefined classification. No parameters regarding stand characteristics (numbers of stems, basal area and so on) were taken. Land-cover categories are coniferous ($n = 1,713$), mixed ($n = 1,399$) and broadleaved ($n = 1,617$) forests and wood-encroached ($n = 635$), pure ($n = 3,533$) and sporadic ($n = 518$) grasslands.

Tree species and mycorrhiza data in broadleaved and mixed forests. For the broadleaved and mixed forests, we used a high-resolution pan-European tree occurrence dataset (European Forest Data Center; http://forest.jrc.ec.europa.eu/efdac/) to assign tree species cover to the LUCAS points. The tree occurrence refers to the presence at the level of tree species over centroids of the 'INSPIRE compliant' 1 km^2 European grid. Tree occurrence is intended as the verified presence of a given tree in a given area (in this case, 1 km^2). First, we extracted from this database the tree species closest to the LUCAS points and compiled a tree species list for each point for which species data were obtained. Then we assigned to each tree species a mycorrhizal associate (ECM, AM or either (if a tree could associate to both)) using published plant-specific mycorrhizal associations[37]. In case of multiple mycorrhizal associations at the LUCAS point, we classified it as 'not determined' (ND) because it was impossible to define a unique association (ND = 1,078 points).

SOM fractions and their C/N ratio. To determine the relative contribution MAOM versus POM fraction to total SOM and the C/N ratios of the two fractions, a representative (Supplementary Fig. 1) subset ($n = 186$) of the LUCAS soil samples, which included 95 grassland sites and 72 forested sites, was fractionated by size (53 μm) after full soil dispersion[16]. Briefly, soils were sieved to 2 mm, and 5 g oven-dried soil was shaken in dilute (0.5%) sodium hexametaphosphate and beads for 18 h to completely disperse the soil. The dispersed soil was then rinsed onto a 53 μm sieve and the fraction passing through (<53 μm) was collected as MAOM; the fraction remaining on the sieve was collected as POM. This fractionation approach was chosen as the most appropriate to separate SOM into two meaningful fractions (POM versus MAOM) with different characteristics and dynamics while being a convenient approach for high throughput[16]. Since the approach defines POM by size (>53 μm), very small amounts of very fine POM are recovered in the MAOM fraction, but this does not lead to different interpretations or conclusions regarding the overall functioning of the MAOM versus POM fraction[17,18].

After drying to constant weight in a 60 °C oven, each fraction was analysed for C and N concentration in an elemental analyser (LECO TruSpec CN). A few of the measured soils contained inorganic C, which was removed from the sample by acid digestion[38] before elemental analyses.

The measured MAOM C and N ($n = 186$) were used to estimate MAOM and POM C and N for the remaining 9,229 soils of the LUCAS database, using the RF approach (Supplementary Fig. 2), and estimated MAOM C and N were compared with the measured data for verification (Supplementary Fig. 3), as described in the data analyses section that follows. In our analyses, soil organic C is used as a proxy for SOM, avoiding the use of conversion factors.

Climate, net primary productivity and N deposition data. For each data point, long-term mean annual temperature (MAT, °C) and annual precipitation (cm) were obtained from the high-resolution WorldClim dataset (version 2), which has average monthly climate data for minimum, mean and maximum temperature and for precipitation for 1970–2000 (http://worldclim.org/version2).

An estimate of the yearly net primary production of LUCAS locations was derived overlaying those points with the 10 yr average Moderate Resolution Imaging Spectroradiometer (MODIS) 2000–2010 at 1 km resolution (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod17a3).

The average N depositions for the period 2006–2010 were derived from the European Monitoring and Evaluation Programme (EMEP) model (rv 4.5) (ref. [39]), providing wet and dry deposition spatially distributed across coniferous and deciduous forests and semi-natural areas (including all grassland).

Statistical analyses. The 'RandomForest' library of R core (https://www.r-project.org/) was used to predict the measured C and N in the MAOM fraction (g C kg^{-1} soil and g N kg^{-1} soil, respectively) from the following predictors: silt + clay content, pH, total SOC, total soil N, N deposition, MAT, land cover, extractable potassium. Those variables were selected among all the variables available to represent soil, climate and land-use characteristics, although only a few of them explained most of the variance (Supplementary Fig. 2). The RF explained 86.0 and 90.8 of the variance in predicting C and N in the MAOM fraction, respectively, with high agreement between measurements and predictions (Supplementary Fig. 3). We performed a cross-validation for the predictions of the C and N in MAOM by splitting the measured dataset in different thresholds of training and testing data (Supplementary Fig. 4). The model-training performances were very good with $R^2 > 0.8$ for all runs.

After the RF training, the model was used to predict the C and N in the MAOM fraction for the 9,229 LUCAS points without fractionation measurements, using the same covariates. The C and N in the POM fractions were calculated as a difference from the total soil organic C and N, respectively. In a few LUCAS points, the estimated C ($n = 109$) and N ($n = 176$) in the POM fraction were negative (Supplementary Fig. 5) and were excluded as considered model failure. With this approach, we are being conservative on the potential to store C in the MAOM fraction. As some points were overlapping, a total of 222 were excluded, representing only 2% of the entire dataset. The points measured spanned the full range of the estimated points. However, only a few were on the high end of the SOM values (Supplementary Fig. 5). Thus, to evaluate the significance of the curvature in the relationship between SOC stocks and MAOM, which is at the base of the saturation concept, we fitted a generalized additive model (GAM) with P-splines smoother using the GAM function of the 'gamm4' R library. The estimated curve uses 2.84 d.f. if all points are included, or 2.61 d.f. if the last two are removed. In either case, the approximate significance of the smooth term is very high ($P < 0.0001$; Supplementary Fig. 6), confirming C saturation of the MAOM fraction (Fig. 4).

We used 1,000 bootstrap replicates to obtain a measure of the uncertainty of the predicted means for C and N in the MAOM and in the POM fractions by land use and mycorrhizal association obtained with RF. Given that the 186 measured sample points were representative of the overall population of LUCAS points, we implemented a direct finite population bootstrap that is a mixture of a simple random sampling without replacement and a one–one design (algorithm 3, ref. [40]). Given the natural unbalance between mycorrhizal types, mycorrhizal distribution has been considered a classification variable. The effect of the different sample sizes has been incorporated into the 95% confidence intervals in Supplementary Table 2, obtained using the bootstrap distribution. This analysis confirmed the significantly lower soil C/N ratio and higher MAOM fraction in the AM forests compared with ECM forests (Fig. 3).

We conducted a path analysis on the 186 measured soil samples to assess the direct and indirect effects on SOC stocks of silt + clay content, pH, MAT, N deposition, net primary production, forests and grasslands. In particular, we fitted a structural equation model with observed variables using maximum likelihood and including f_{MAOM}, C/N$_{POM}$ and C/N$_{MAOM}$ as mediation variables. To this end, we have used the structural equation model function of the 'lavaan' R library[41]. The final set of factors was assessed by retrieving only those with a $P < 0.05$ in at least one regression. We limited this analysis to the 186 measured soil samples to avoid circularity in our reasoning and have this analysis entirely independent of the RF analyses.

Data availability

Data from the LUCAS database can be accessed at http://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data. The fractionation data will be made available at the European Soil Data Centre of the EU Joint Research Centre website (https://esdac.jrc.ec.europa.eu/). The tree occurrence dataset can be also downloaded from https://doi.org/10.6084/m9.figshare.c.3288407, associated with the paper by Mauri et al.[21]. The final dataset is available from the European Soil Data Centre of the European Commission: https://esdac.jrc.ec.europa.eu/content/soil-organic-matter-som-fractions. Requests for data can be addressed to E.L. (Emanuele.LUGATO@ec.europa.eu).

Code availability

The R scripts are available from the European Soil Data Centre (ESDAC) of the European Commission: https://esdac.jrc.ec.europa.eu/content/soil-organic-matter-som-fractions.

References

37. Wang, B. & Qiu, Y. L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**, 299–363 (2006).
38. Harris, D., Horwáth, W. R. & van Kessel, C. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. *Soil Sci. Soc. Am. J.* **65**, 1853–1856 (2000).
39. Simpson, D. et al. The EMEP MSC-W chemical transport model—technical description. *Atmos. Chem. Phys.* **12**, 7825–7865 (2012).
40. Antal, E. & Tillé, Y. A direct bootstrap method for complex sampling designs from a finite population. *J. Am. Stat. Assoc.* **106**, 534–543 (2011).
41. Rosseel, Y. Lavaan: an R package for structural equation modeling and more v.0.5–12 (BETA). *J. Stat. Softw.* **48**, 1–36 (2012).

Author contributions

M.F.C., E.L. and J.S. developed the research concepts. E.L. and M.G.R. conducted the data and statistical analyses. M.L.H. performed the soil analyses. M.F.C., E.L. and J.S. interpreted the data and wrote the paper with contributions from M.G.R. and M.L.H.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41561-019-0484-6.

Correspondence and requests for materials should be addressed to M.F.C.

Peer review information Primary Handling Editor(s): Xujia Jiang.

Reprints and permissions information is available at www.nature.com/reprints.