Science of the Total Environment 951 (2024) 175476

Contents lists available at ScienceDirect

Science of the Total Environment

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

Major overlap in plant and soil organic carbon hotspots across Africa

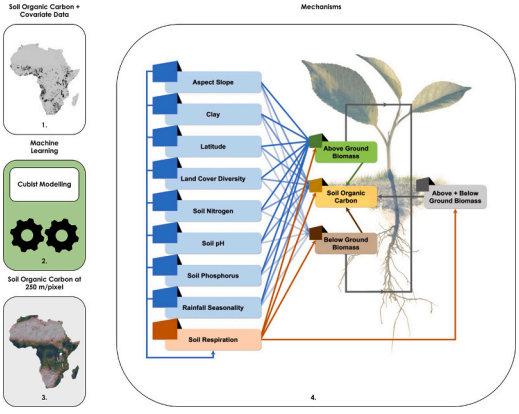
Ndiye Michael Kebonye [a,b,*], Kingsley John [c], Manuel Delgado-Baquerizo [d,**], Yong Zhou [e,f], Prince Chapman Agyeman [g], Zibanani Seletlo [h], Brandon Heung [c], Thomas Scholten [a,b,i]

[a] Department of Geosciences, Chair of Soil Science and Geomorphology, University of Tübingen, Rümelinstr. 19-23, Tübingen, Germany
[b] DFG Cluster of Excellence "Machine Learning: New Perspectives for Science", University of Tübingen, AI Research Building, Maria-von-Linden-Str. 6, 72076 Tübingen, Germany
[c] Department of Plant, Food and Environmental Sciences, Faculty of Agriculture, Dalhousie University, 50 Pictou Rd, Truro, NS B2N 5E3, Canada
[d] Laboratorio de Biodiversidad y Funcionamiento Ecosistémico, Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Sevilla, Spain
[e] Department of Wildland Resources, Utah State University, Logan, UT 84321, USA
[f] Ecology Center, Utah State University, Logan, UT 84321, USA
[g] Sustainable Resource Management, Memorial University of Newfoundland, Corner Brook A2H 6P9, Canada
[h] Department of Animal Science and Production, Faculty of Animal and Veterinary Sciences, Botswana University of Agriculture and Natural Resources, Private Bag 0027, Gaborone, Botswana
[i] CRC 1070 Resource Cultures, University of Tübingen, Tübingen, Germany

HIGHLIGHTS

- Modelling approaches delineate African terrestrial C hotspots.
- Scale-dependent SOC-plant C interplay
- The tropics host most of the terrestrial C in Africa.
- Terrestrial C hotspots are required for decision-making.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Abasiofiok Mark Ibekwe

Keywords:
Africa
Landscapes

ABSTRACT

Terrestrial plant and soil organic carbon stocks are critical for regulating climate change, enhancing soil fertility, and supporting biodiversity. While a global-scale decoupling between plant and soil organic carbon has been documented, the hotspots and interconnections between these two carbon compartments across Africa, the second-largest continent on the planet, have been significantly overlooked. Here, we have compiled over 10,000 existing soil organic carbon observations to generate a high-resolution map, illustrating the distribution pattern

* Correspondence to: N. M. Kebonye, Department of Geosciences, Chair of Soil Science and Geomorphology, University of Tübingen, Rümelinstr. 19-23, Tübingen, Germany.
** Corresponding author.
E-mail addresses: ndiye.kebonye@uni-tuebingen.de (N.M. Kebonye), m.delgado.baquerizo@csic.es (M. Delgado-Baquerizo).

https://doi.org/10.1016/j.scitotenv.2024.175476
Received 17 May 2024; Received in revised form 8 August 2024; Accepted 10 August 2024
Available online 13 August 2024

Soil organic carbon (SOC)
Spatial scale
Above-ground carbon
Below-ground carbon

of soil organic carbon in Africa. We then showed that above- and below-ground plant carbon are significantly and positively correlated with soil organic carbon across Africa. Both soil and plant carbon compartments shared major hotspots in the tropical regions. Our study provides critical insights into the spatial distribution of carbon hotspots across Africa, essential for soil conservation and safeguarding terrestrial carbon stocks amidst the challenges of climate change.

1. Introduction

Terrestrial plant carbon (C) and soil organic carbon (SOC) stocks are essential for enhancing soil fertility, regulating climate, and supporting biodiversity worldwide. While soils store two to three times more C than is found in plants, both plant and soil C compartments are, in theory, highly interconnected (Luo et al., 2019). Thus, plant litter inputs could potentially contribute to C stored in some soils (Grace et al., 2006). In turn, organic matter (OM) stored within soils is critical for supporting nutrient cycling, providing available nutrients for plant uptake and growth. However, the current literature suggests a global-scale decoupling of plant and soil C stocks, with the tropics exhibiting the highest plant C pools while boreal regions support the largest soil C reserves (Blais et al., 2005). Therefore, both understanding the interconnections between plant and soil C and precisely pinpointing C storage hotspots is crucial for developing effective strategies to enhance terrestrial C storage and conservation, thereby promoting ecosystem sustainability and mitigating climate change (Berhongaray et al., 2019; Luo et al., 2019; Rasse et al., 2005; Wang et al., 2022).

Many studies examine the contribution and influence of plant biomass C on soil organic carbon (SOC) through small-scale field experiments (Berhongaray et al., 2019; Dietzel et al., 2017; Kätterer et al., 2011) or studies related to land use types (Geng et al., 2017). Examples in Africa, such as those by Bukombe et al. (2022), show that the allocation of soil and plant C along a toposequence in Central Africa is largely driven by geochemical properties rather than topography. Zhou et al. (2023) demonstrates that C derived from grasses contributes more than half of the SOC in tropical savannas. At smaller spatial scales (e.g., ecosystem-scale), soil processes and properties can be better understood because of the fine details expected at this level (Eusterhues et al., 2003; Martens et al., 2023; Zech et al., 2022). However, studies investigating C hotspots and interlink between plant and soil C for vast spatial scales are largely lacking. In Africa, most large-scale studies only map the spatial distribution and variability of SOC (Hengl et al., 2021; Vågen et al., 2016), monitor the temporal dynamics and persistence of SOC in selected sub-Saharan regions (von Fromm et al., 2024), or attempt to understand the controls of SOC in the sub-Saharan region with fairly limited samples (von Fromm et al., 2021). Even so, the patterns of above- and below-ground plant biomass C, in conjunction with SOC, are critical to understand especially across Africa; however, such understanding is severely lacking when only relying on experimental research that targets a few sites.

Plants can significantly impact soil C through processes via litter inputs and root rhizodeposition; additionally, the formation and persistence of SOC are also shaped by additional factors, including climate and soil texture (Wiesmeier et al., 2019). In another instance, we know that soil microbes tend to decompose organic materials (e.g., plant litter) at different rates. However, when temperature increases, decomposition rates increase concurrently due to high microbial activity. Temperature rise has been shown to promote rapid heterotrophic soil respiration (Bond-Lamberty et al., 2018), resulting in a decrease in SOC stocks (Mayer et al., 2017). Meanwhile, nutrients, including phosphorus and nitrogen, are also crucial components of the soil expected to improve plant productivity, which in turn, should increase SOC stocks (Spohn et al., 2023) via litter input. Long-term experiments provide evidence for this outcome (Baethgen et al., 2021; Merbach and Schulz, 2013) although not for large spatial scales. All these environmental factors and processes play an integral role in the soil-plant C interactions. Aside from a scarcity of large-scale investigations on soil-plant C linkages, it remains uncertain whether regions with elevated plant biomass C always correspond to high soil C levels, especially across varying environmental conditions, such as Africa. Also, previous SOC mapping efforts in Africa have either limited their efforts by applying only remote sensing data (e.g., MODIS; Vågen et al., 2016) or did not account for spatial autocorrelation, limiting the reliability of the SOC predictions. Other African SOC maps had regions with missing data, especially in dryland regions (Hengl et al., 2021; Hengl et al., 2015). Meanwhile, the latest SOC map from SoilGrids 2.0 is a prediction of the entire world, not Africa (Poggio et al., 2021). This global map might not be able to capture the detailed local variations of SOC across Africa.

Here, we seek to investigate C hotspots and the interlinking of above- and below-ground plant biomass C and SOC across Africa. Despite its unique landscape configuration and varied climatic and vegetative distribution, the continent has been underrepresented in studies on plant and soil C distribution patterns and dynamics. Unlike previous global C modelling products, which have largely overlooked the distinct characteristics of the African landscape, our study provides a more nuanced understanding of the continental drivers of C dynamics. Africa contributes ~7 % of global C (Corbeels et al., 2019; Ontl and Schulte, 2012), yet the underlying mechanisms that require urgent attention to drive more C into the soil remain poorly understood. Through advanced statistical modelling, our research demonstrates a clearer understanding of soil and plant C dynamics and processes across Africa's varied environmental conditions. Furthermore, the substantial threats facing many ecosystems in Africa, such as deforestation and land degradation, this knowledge plays a pivotal role in determining whether both plants and soils share the same C hotspots or if separate protection strategies are required for the conservation of plant and soil C when their hotspots do not entirely coincide. For policymakers, the interest is to pinpoint terrestrial C hotspots for C balance, pricing, or budgeting efforts (Berretta, 2020; Fatichi et al., 2019; Mugabowindekwe et al., 2024). These hotspots need to be effectively and transparently communicated; as such, implementing large-scale C modelling approaches to generate digital interactive products (e.g., maps) is invaluable. Our objectives were to: (1) develop a high-resolution SOC map for Africa, (2) delineate the terrestrial C (plant C compartments and SOC) spatial hotspots across the continent, and (3) examine the interlinks between plant C and SOC in Africa.

2. Methodology

To address the research objectives, we assembled a soil database of >10,000 SOC concentration field observations. We then applied machine learning (i.e., Cubist) to generate a high-resolution (~250 m/pixel) SOC map across Africa. Since the emphasis here was SOC, we first generated our own continental SOC map rather than rely on existing products (e.g., Africa Soil Information Service: AfSIS 250 m SOC map). We integrated spatial features (latitude and longitude) in our machine learning model. This way, our model would account for vast heterogeneity in sample locations across Africa. Moreover, the spatial features would allow our SOC model to learn about geographic priors rather than simply rely on the distributions of environmental covariates (e.g., climatic variables, and terrain) to generate a prediction. Having generated our high-resolution SOC map, we also obtained detailed above- and below-ground plant biomass C maps (Spawn et al., 2020). Based on the SOC, plant biomass C, and selected environmental driver (e.g., soil

nutrients, land cover diversity) maps spanning Africa, we extracted spatially representative data throughout the continent. This data was used as input for our statistical and mechanistic models to understand soil and plant C interlinks. It is worth noting that, environmental drivers used to explain soil-plant C dynamics were selected based on expert knowledge as some of the well-known drivers of the terrestrial C cycle. Since soil processes vary greatly across scales, especially over a vast landscape like Africa, we also studied similar interactions across a continuum of spatial scales by aggregating all maps from high to low scale. Newly aggregated maps were generated for the district and national levels which we refer to as intermediate and coarse scales throughout this study. Aggregating maps into varying scales would help with continental decisions and policy formulation affecting soil and plant C. This is a similar strategy used by the European Union in several areas (e.g., education, agriculture, economics) (European Union, 2018). To achieve this, we follow a series of stages outlined in a workflow (Supplementary Data Fig. 1). Detailed explanations for the different modelling phases are fully provided in the following subsections.

2.1. SOC data

Our study used the latest harmonized World Soil Information Service (WOSIS) SOC database at a scale of 1:100,000 (https://maps.isric.org/mapserv?map=/map/wosis_latest.map). This database contains thousands of profiles collected globally at various depths, ranging from the surface to the subsoil, collected between 1920 and 2013 (Batjes et al., 2020). Because the goal was to understand the relationships between SOC and plant C in Africa, we exclusively used SOC sampling locations obtained in Africa, especially the topsoil samples at the 0–20 cm depth increment, as these are more likely to be influenced by recent vegetation and environmental conditions. This selection resulted in $n =$ 10,750 observations, used to estimate SOC in grams per kilogram (g/kg) throughout Africa.

2.2. Environmental drivers for predicting SOC

The spatial variability, storage, and organic C sequestration potential of soils are all driven by a combination of environmental factors that may be represented as spatial predictors. Therefore, SOC was modeled based on environmental drivers derived from terrain, Landsat, soil, and climatic data. Some of these drivers, such as terrain and climatic data (temperature and rainfall), have previously been used to successfully model SOC across Africa (Hengl et al., 2021). A reference potential evapotranspiration image at a spatial scale of 30 arc sec was obtained from Zomer et al. (2022). Digital elevation data was based on Shuttle Radar Topography Mission (SRTM) data, acquired at a spatial scale of 250 m, and downloaded via the Google Earth Engine (GEE) platform. Hydrologic soil groups data (HYSOGs250m) (Ross et al., 2018); Landsat

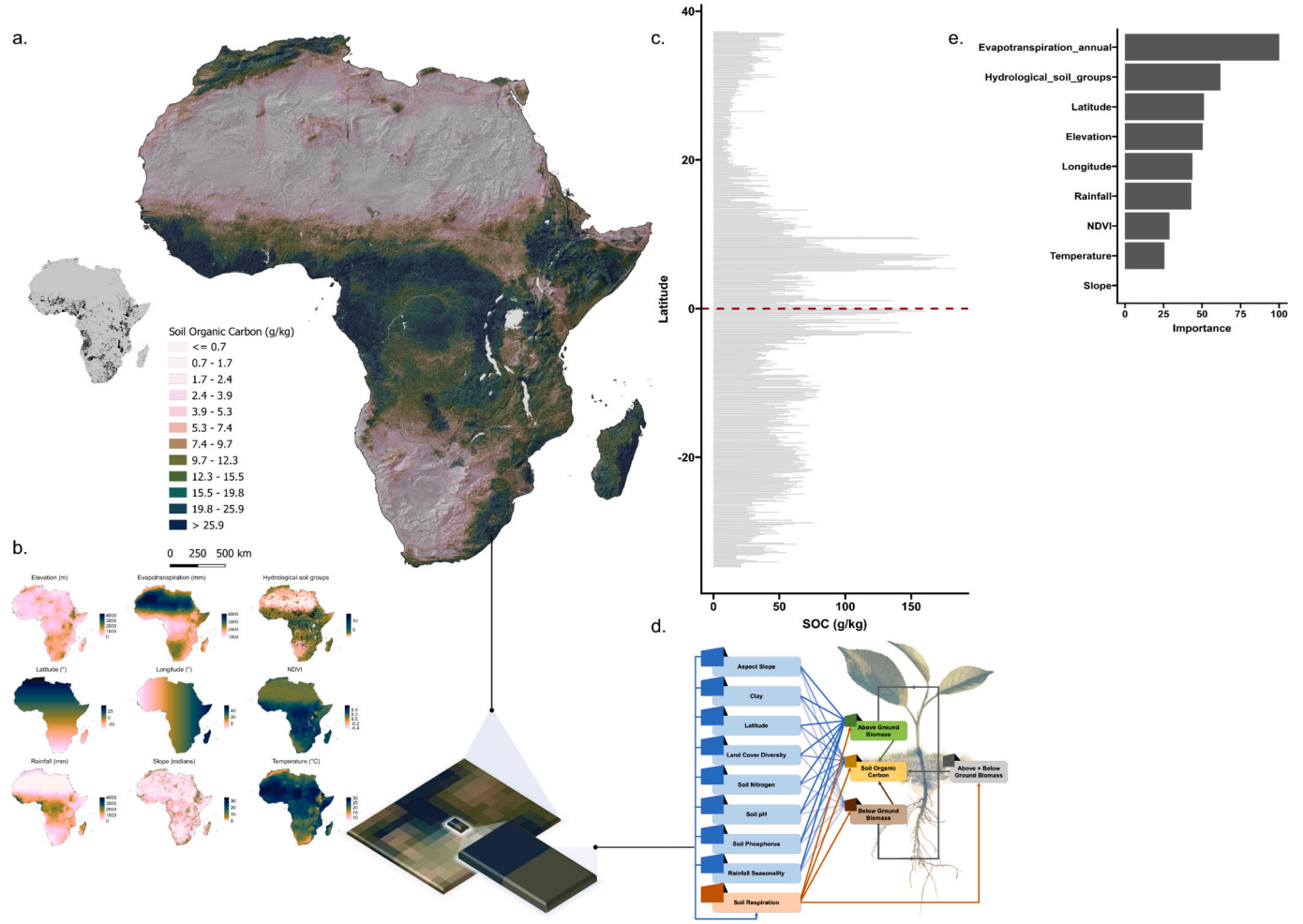


Fig. 1. Soil organic carbon (SOC) model and map for Africa in g/kg. (a) SOC sampling locations and 250 m/pixel prediction. (b) The first round of environmental drivers was used to model and map SOC. (c) SOC levels across a latitudinal gradient (plot uses the approximately $n = 100{,}000$ systematic grid synthetic sampling locations proposed in the workflow), the red dotted line represents the equator which also aligns well with the SOC spatial distribution map. (d) Hypothetical a priori pathways between secondary environmental drivers while assessing SOC and plant biomass C relationships for all spatial scales. (e) Variable importance plot showing each environmental driver's contribution or influence on the Cubist algorithm.

8 image collection 1 Tier 18-day normalized difference vegetation index (NDVI) composite for the period 2014 to 2022 (mean estimate), produced at a 250 m spatial scale, was downloaded via the Google Earth Engine (GEE) platform. Hydrologic soil groups are derived from textural classes; hence, they also provide insights into different soil types within the modelling channel. For instance, hydrologic soil group 'A' corresponds to soils that are primarily sandy (Ross et al., 2018). Hydrological soil groups significantly influence SOC variability by determining the rates of water infiltration, runoff, and erosion potential associated with different soil groups. In essence, these influence processes linked to OM decomposition by soil microbes and C input into soils owing to variations in water infiltration and runoff capacities (Védère et al., 2022). We obtained the long-term mean annual rainfall and temperature data (1970 to 2000). Such historical data allows us to explicitly account for past climatic conditions and their potential influence on recent SOC variability. Climatic legacies have a major effect on SOC in terrestrial ecosystems (Delgado-Baquerizo et al., 2017). To incorporate spatial context the oblique geographic coordinates approach by Møller et al. (2020) generated only the latitude and longitude coordinate equivalences. All environmental drivers for the African continent were aggregated to a common 250 m spatial scale. The full details on the first-round drivers can be found in Supplementary Data Table 1.

2.3. High-resolution predictive modelling of SOC

Soil observational data were gathered from $n = 10{,}750$ locations and combined with environmental drivers using a digital soil mapping (DSM) framework (McBratney et al., 2003). The dataset was cleaned and partitioned into training (90 %) and validation (10 %) sets. The SPlit approach, which considers the distribution of the data, was used to ensure representative subsamples for both sets (Joseph and Vakayil, 2022; Kebonye, 2021; Kebonye et al., 2023a). This was implemented using the *SPlit* package in *R* (RStudio Team, 2020). In addition, a Cubist model was utilized for SOC prediction across Africa, leveraging its computational efficiency, better interpretability, and adaptability to diverse data types (Quinlan, 1992) than the common Random forest model. This model has been widely applied in different DSM projects (e.g., Henderson et al., 2005; Viscarra Rossel et al., 2015; Walden et al., 2023). The model's hyperparameters were optimized using 10-fold cross-validation with 10 repeats, selecting the model with the lowest RMSE as optimal. This process was executed using the *caret* package in *R* (Kuhn et al., 2023). To evaluate the accuracy of our final SOC model, we used the coefficient of determination (R^2), concordance correlation coefficient (CCC), root mean square error (RMSE), and bias.

Our final model (i.e., with the lowest RMSE), fitted with all the data, and incorporating geographic positions (i.e., latitude and longitude), was used to create a SOC map with a resolution of ~250 m/pixel. We also computed entropy estimates per pixel for our SOC prediction based on a co-occurrence matrix approach (Haralick et al., 1973). These estimates, based on spatial relationships between pixel values, provided a general indication of regions to expect high and low uncertainty. High and low entropy values correspond to high and low uncertainty estimates, respectively. Entropy estimates account for uncertainty in soil classification maps (i.e., discrete) (Zhu, 1997). According to Dai and Chen (2012), entropy can also be estimated for continuous cases. By continuous here we mean spatially continuous variables like SOC. Finally, to ensure the reliability of our map product, we assessed spatial autocorrelation (SAC) in the predictions using model residuals. Assessing spatial autocorrelation allowed us to validate if our model accurately represented reality, which is critical for map interpretation. This method has been used in several DSM projects to account for SAC across different spatial scales (Chevalier et al., 2021; Craven et al., 2020). Here, we used a Monte Carlo simulation (Wadoux and Heuvelink, 2023) to run a spatial autocorrelation function for Euclidean distances of 500 km, resulting in a correlogram plot (Nol et al., 2010). We used a sample size of $n = 500$ runs, deemed optimal for achieving a near 1:1 line of best fit on a scatter plot (Nol et al., 2010). A 500 km distance was considered relevant since it captures the detailed changes at the landscape scale which are of interest in this study. In this study, we consider the 250 m/pixel spatial scale as the fine-scale SOC map for Africa, which was further aggregated to district and national levels, the intermediate and coarse scales respectively. This aggregation aids in the development of C conservation policies. All SOC modelling and mapping analyses were conducted using *R* and *QGIS*, both of which are Free and Open-Source Software (FOSS).

2.4. Linking soil and plant C across Africa

Following the creation of our SOC map, we explored the link between plant and soil C in Africa, focusing on identifying hotspots under different environmental conditions. We considered plant biomass C alongside secondary drivers, including regional factors such as soil properties and land cover, to fully understand these links and underlying soil processes. The plant biomass C estimates, including above-ground biomass carbon (AGBC) and below-ground biomass carbon (BGBC), both at a spatial scale of 300 m, were obtained from Spawn et al. (2020). For total plant biomass carbon analysis, AGBC and BGBC estimates were combined (AGBC + BGBC). The aspect was derived along with previous terrain derivatives (elevation and slope). Latitude was also used to capture landscape changes associated with variations in latitudinal gradients. A 100 m scale land cover image for 2015 (Buchhorn et al., 2020) was converted to alpha diversity via the Shannon–Wiener index in *R* (Rocchini et al., 2021), generating a continuous variable to represent land cover diversity. The land cover diversity is now represented as a new image with pixel values ranging from zero (i.e., homogeneous or no diversity regions like the Sahara) to at most two representing the highest (i.e., heterogeneous or diverse regions like Central Africa) (Martínez-Núñez et al., 2023). Soil properties, including clay, nitrogen, pH, and phosphorus were obtained from the ISRIC – World Soil Information and ISDA soil databases (Hengl et al., 2021; Hengl et al., 2015). Rainfall seasonality, which captures the irregularities in rainfall for a normal year, was obtained from the WorldClim database (Fick and Hijmans, 2017). Soil respiration, which is the amount of CO_2 released from the soil surface into the atmosphere, was also obtained to account for soil microbial and plant root activities (Warner et al., 2019). Supplementary Data Table 1 contains complete information on these secondary drivers.

Lastly, the previously created SOC map was used as a response variable for each subsequent mechanistic and statistical modelling process. To capture as much spatial variability of the landscape as possible, we initially implemented regular sampling without replacement to generate approximately $n = 100{,}000$ synthetic sampling locations across the entire African continent. These sampling locations were then intersected with a raster stack that included all the secondary environmental drivers, above- and below-ground plant biomass C, and the SOC map. This process was first executed for the fine scale, followed by the intermediate and coarse scales, in that order (refer to Supplementary Data Fig. 1, stage 4). This approach facilitated the extraction of new sampling location datasets from the raster stacks for each spatial scale (see Supplementary Data Figs. 2, 3, and 4).

2.5. Mechanistic multivariate modelling

Structural equation models (SEMs) are mechanistic multivariate approaches used to describe pairwise relationships between variables while also delivering an intuitive graphical output illustrating networks of these relationships (Eisenhauer et al., 2015). In our case, this was the pairwise relationships between SOC and each of the plant biomass C (AGBC, BGBC, and the AGBC + BGBC = total plant biomass C) in the presence of other selected drivers (i.e., secondary drivers) across spatial scales. In each case, a priori hypothetical model (Fig. 1d), considered to be a baseline saturated model, was used to derive all the SEMs in this study. It is considered a saturated model since it is comprised of fully connected paths. However, such a saturated model usually results in a

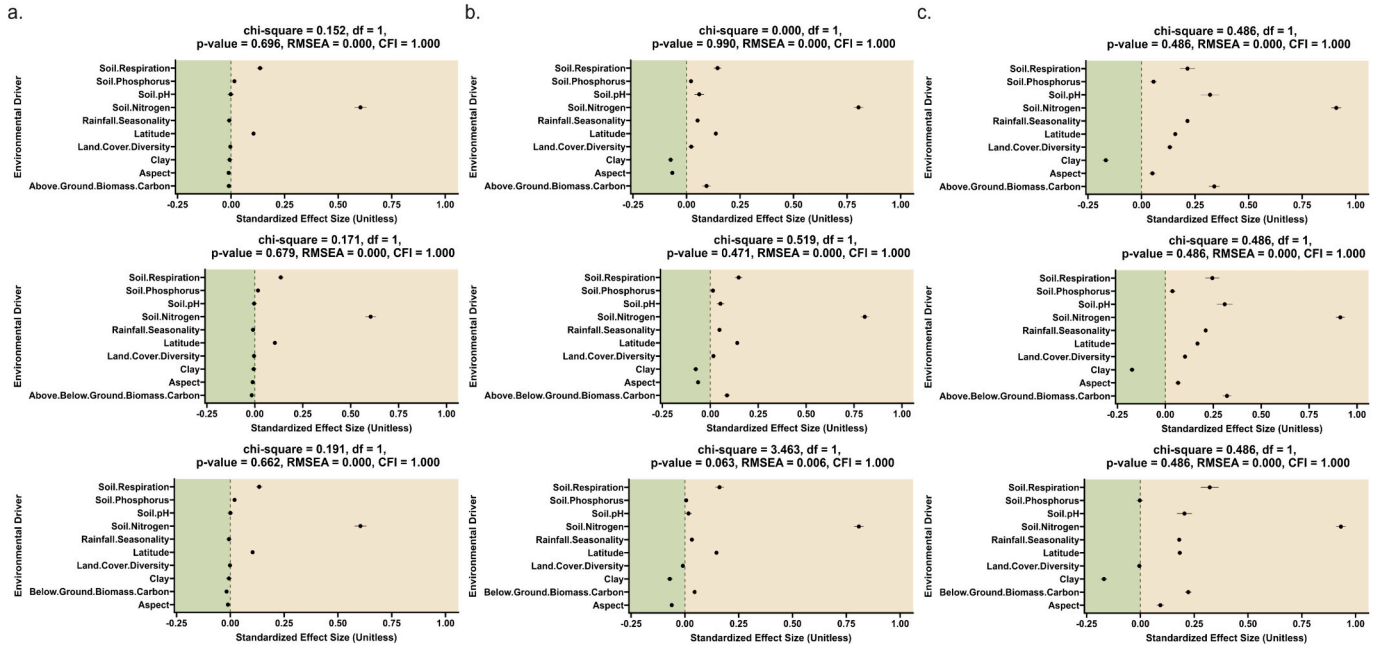


Fig. 2. Robust standardized effect sizes from structural equation models (SEMs) showing the pairwise relationships between soil organic carbon (SOC) and the different secondary environmental drivers across the various spatial scales, 250 m/pixel (fine), district (intermediate) and national (coarse). [Note: All plots in the same columns a, b, and c represent results for fine, intermediate, and coarse scales separately. Moreover, across rows, the first plots represent results associated with above-ground biomass carbon (AGBC) followed by total plant biomass carbon and below-ground biomass carbon (BGBC). The SEM colours only differentiate the negative (olive) from the positive (tan) standardized effect sizes. All through, the error bars on the effect sizes are the 95 % confidence intervals (CI).]

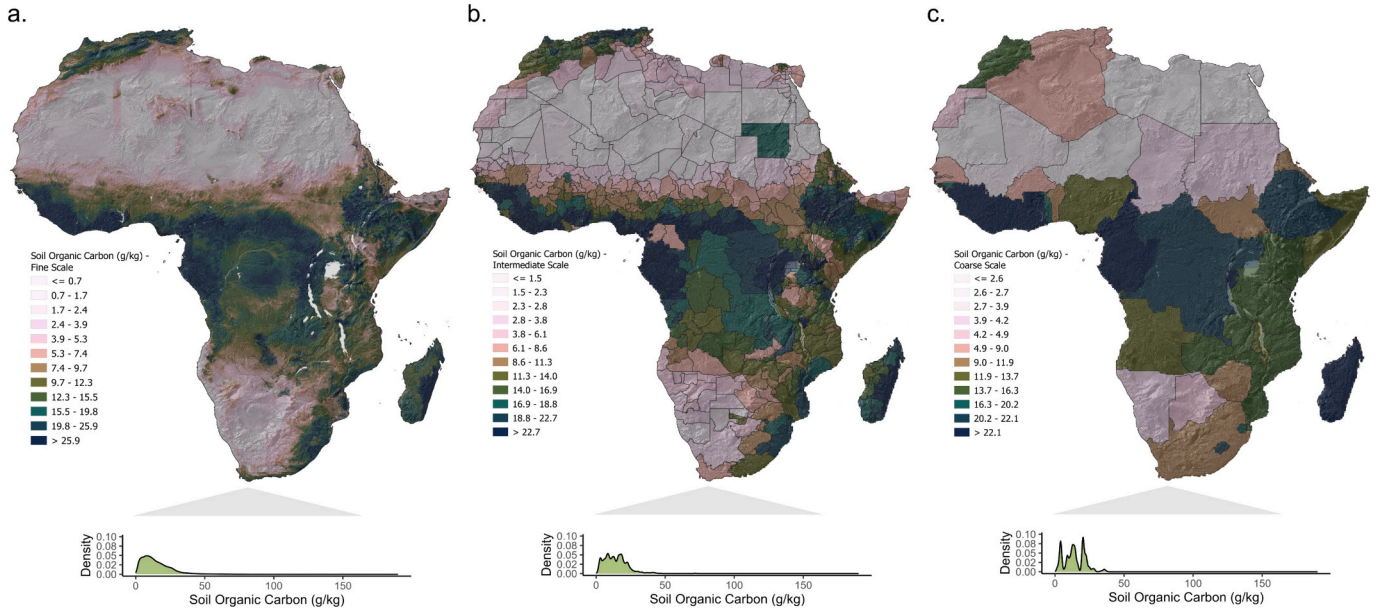


Fig. 3. The soil organic carbon (SOC) maps for fine (a), intermediate (b), and coarse (c) scales used as response variables in the Generalized Additive Models (GAMs). [Note: Underneath the maps are their density plots showing the differentiation in SOC level distributions for each map. The density plots per spatial scale were generated from the approximately $n = 100{,}000$ systematic grid synthetic sampling locations proposed in the workflow.]

poor Chi-square fit ($p < 0.05$) because of the complexity of variable interactions thus, requiring further simplification to improve fitting. Initially, based on the saturated model, plant biomass C estimates were individually related to SOC while also incorporating other secondary drivers per spatial scale. Each SEM model would initially be assessed based on standard goodness of fit metrics including the Chi-square, *p*-value, root mean square error of approximation (RMSEA), and comparative fit index (CIF; Siles et al., 2023). Since the aim is to obtain a good-fitting SEM model, the goodness of fit metrics guided the need to perform a re-run of each analysis to obtain optimal results based on $n = 1000$ bootstraps. For generating optimal results, SEMs need to have $p > 0.05$, RMSEA close to 0, and CFI close to 1 (Eisenhauer et al., 2015). The process of obtaining optimal SEM results involved the systematic removal and or replacement of predictor variables (i.e., secondary drivers) while ensuring the SEM model does not vary much from the initial hypothetical model. For this, predictors with predictor variable modification indices (MI) of >10 were considered for either removal or replacement after each SEM re-run (Garrido et al., 2022). For full details

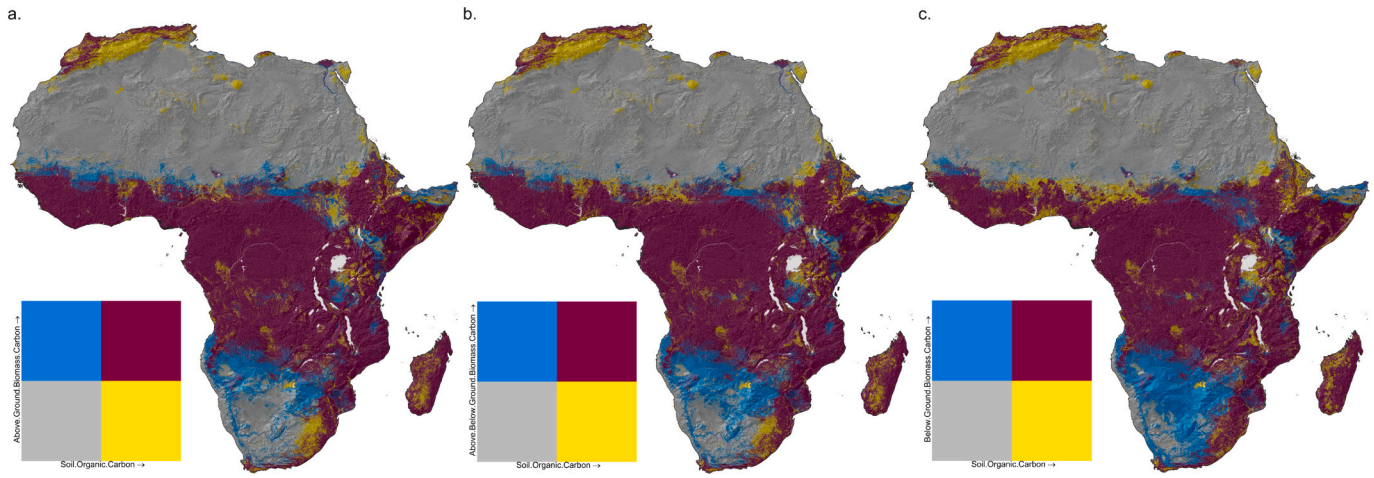


Fig. 4. Soil organic carbon (SOC) and plant biomass carbon (C) overlap. [Note: The maps represent results associated with above-ground biomass carbon (AGBC) followed by total plant biomass carbon and lastly below-ground biomass carbon (BGBC). Overlaps represent hotspots showing low–low (grey), low–high (blue), high–low (yellow) and high–high (pompadour) values respectively. The bivariate maps only show the fine-scale overlaps while the intermediate and coarse scale maps are shown in the Supplementary Data Fig. 10.]

on the SEM approach, we refer to previous studies, such as Eisenhauer et al. (2015) and Garrido et al. (2022). The *lavaan* package in *R* was used to run the SEM analyses on standardized datasets (with a zero mean) for each spatial scale to test the a priori hypothetical model (Rosseel, 2012; Siles et al., 2023). The standardized datasets ensured both response (SOC) and predictor variables ranged between 0 and 1 representing the smallest and largest values respectively.

2.6. Statistical modelling

For the statistical modelling, we applied Generalized Additive Models (GAMs), which are semi-parametric, non-linear, and able to characterize the complex relationships between the SOC and the respective plant biomass C, as well as the other environmental drivers. GAMs provide reasonable interpretability, allowing for coefficients to be expanded as smoothing functions for each of the predictors applied (de Brogniez et al., 2015). Like the SOC modelling, datasets for each spatial scale were partitioned into training (90 %) and independent validation (10 %) sets based on the SPlit approach. The GAMs were implemented using the *gam* function based on the *gam* package in *R* (Hastie, 2023). It is worth noting that, aside from collinearity, the selection of the secondary drivers was based on an a priori motive to particularly include already known SOC drivers. We did not employ the Cubist model in place of GAM since it is more difficult to interpret than a GAM. Also, despite its intuitive nature, Cubist necessitates further interpretation of results which are usually decisions or rules, typically in the form of individual multiple linear regression equations (Khaledian and Miller, 2020), whereas only a single 2D scatter plot associated with a GAM directly reveals the relationship between the response and each of the predictors of interest. As a result, GAMs may be considered to give first-hand insight into variable interactions. An additional benefit of the GAMs is that they are less computationally demanding than the Cubist model.

Despite the bents of GAMs, they are sensitive to multicollinearity; hence, variance inflation factor (VIF) analysis was carried out with a VIF threshold of 10, of which only plant biomass C would need to be excluded as per the results (Supplementary Data Table 2). All other predictors had their VIF values below the threshold of 10. Deleting the plant biomass C predictors following the GAM results would not allow us to explain how each relates to SOC in the context of other drivers. Therefore, on this account, we did not remove any predictors to ensure that we answered the current study objectives. In general, with SEM, we were able to develop and evaluate hypothesis-based models driven by expert knowledge that are mostly linear. The GAM, on the other hand, is a hypothesis-free machine learning technique that allows us to capture non-linear variable connections while acquiring knowledge on how each driver fully influences the SOC.

2.7. Major overlaps in the hotspots of soil and plant biomass C

A bivariate mapping approach that we recently introduced for DSM purposes (Kebonye et al., 2023b) was used to assess the spatial overlaps between SOC and plant biomass Cs across different spatial scales. This method simplifies the depiction of spatial relationships between SOC and plant biomass C, exposing the primary spatial C hotspots of interest between these variables that may be used in decision-making on a landscape scale. By C hotspots, we refer to regions across Africa where significantly high levels of C are expected, both plant and soil-based, than their surroundings. These hotspots are important for C cycling, climate change mitigation, and ensuring its balance in terrestrial ecosystems. Various studies have developed frameworks to identify and investigate C hotspots and their dynamics worldwide (Hobley et al., 2018; Mugabowindekwe et al., 2023; Ottoy et al., 2022; Timilsina et al., 2013). We applied a four-color palette matrix comprising four major hotspots between SOC and each plant biomass C. The four major hotspots denoted low–low (grey), low–high (blue), high–low (yellow), and high–high (pompadour) values respectively. We clarify that the low–low and high–high values are thresholds that reflect the hotspots likely to have the lowest and highest values for each scenario, rather than the fact that only the lowest and highest levels exist here. To set particular thresholds to determine what constituted low and high values for each variable, we used a quantile scale break for the two-by-two color palettes per variable for visual simplicity (Bivand et al., 2023). Despite the current grouping being coarse, considering a large-scale territory like Africa, which maintains a diversity of ecosystem types and shows considerable spatial heterogeneity, it was only fair to keep four main classes given the constraints of bivariate mapping already outlined (Kebonye et al., 2023b). We refer to our previous study on the full explanations regarding the bivariate mapping approach (Kebonye et al., 2023b). The *R* package *classInt* was important for this procedure (Bivand et al., 2017) while we applied the R code from: https://gist.github.com/scbrown86/2779137a9378df7b60afd23e0c45c188.

3. Results and discussion

Our model-based findings show a fundamental connection between

soil and plant biomass C (AGBC and BGBC) in Africa and further suggest that soil and plant C have similar hotspots explicitly delineated in the tropical zone. These findings are crucial for understanding the biogeochemical cycles. Additionally, they are essential for addressing concerns such as climate change mitigation, biodiversity conservation, sustainable agriculture practices, and landscape planning. We anticipate that given the fine-scale (~250 m/pixels) delineation of terrestrial carbon hotspots in digital form, it should be feasible to identify specific regions of interest (e.g., with levels higher than the 75th or 95th percentile) for initiating C or nutrient monitoring benchmarking efforts for various purposes, such as agriculture and landscape planning. This level of detail was not achievable with individual soil and plant carbon maps. Furthermore, our study has shed light on the mechanisms linking soil and plant C across the continent, which were previously unclear due to localised sampling campaigns.

3.1. SOC spatial variability and model assessment

Our study suggests that the majority of soil C in Africa is located within tropical regions (Fig. 1). The spatial variability of SOC shows a substantial gradient along the southern to the northern coasts of Africa (Fig. 1a and c) and high levels of SOC are primarily found around the tropics or equator in Fig. 1c (marked by a red dotted line). Both the Saharan region in the northern part of Africa and the Kalaharian region in the southern part of Africa have relatively lower SOC levels owing to their arid to hyper-arid climatic conditions (Richards et al., 2023), which limits plant productivity and, consequently, C inputs into the soil. Moreover, these regions only have a few sampling locations since they are mostly deserts. The environmental drivers used to map SOC varied spatially across Africa (Fig. 1b). Overall, our model suggests that potential evapotranspiration is the most important variable explaining the spatial patterns of SOC across Africa (Fig. 1e). This is evident as the SOC map closely mirrors that of potential evapotranspiration, where regions with high potential evapotranspiration values correspond to low SOC levels. High potential evapotranspiration leads to high soil respiration and OM decomposition resulting in lower SOC, which is typical of drier regions with high temperatures (Nyberg and Hovenden, 2020). Meanwhile, in tropical settings with stable climatic conditions, constant litter input may raise SOC content (Bradford et al., 2016). This is due to the continuous addition of organic matter to the soil, which can enhance soil quality and water-holding capacity, and potentially contribute to C sequestration. The hydrological soil groups followed after potential evapotranspiration in influencing the spatial variability of SOC in Africa. This is an important driver of SOC because it reflects the regional characteristics of soils in general and conservation practices and water runoff potential in particular (Baumann et al., 2009).

We also found that NDVI, a metric reflecting vegetation cover and greenness and positively associated with vegetation vigor, played a role in predicting the spatial distribution of SOC, where higher NDVI values corresponded to higher SOC levels. Elevation played a more significant role in the SOC model than slope, confirming a previously established finding in an earlier study (Hengl et al., 2015). We observed some regions where a noticeable spatial correspondence between high SOC and elevation levels stands out, such as in southern (e.g., South Africa) and eastern Africa (e.g., Ethiopia).

Regarding the model accuracy, the results suggest a reasonable model fit with a concordance of 0.66 (Supplementary Data Fig. 5). Furthermore, the current model's RMSE of 17.17 is lower compared to the SoilGrids 250 m and SoilGrids 2.0's RMSE of 36.48 (Poggio et al., 2021). As for the Cubist model residuals, there was no evidence of spatial autocorrelation which means the deterministic model captured the true spatial variability of SOC (Supplementary Data Fig. 5). The SOC map generated in this study closely resembles the SoilGrids 2.0, which also includes Africa. Nonetheless, the present mapping effort incorporates geographical priors in the SOC prediction while also testing for spatial autocorrelation in SOC distribution across Africa, a factor not explicitly addressed in the previous attempt. Moreover, our study provides better details into the spatial variability of SOC compared to earlier efforts (Supplementary Data Fig. 6). As expected, we are more uncertain about very high SOC estimates for example (Supplementary Data Fig. 11), in the western (e.g., Côte d'Ivoire, Ghana, Guinea) and central parts of Africa (e.g., Democratic Republic of the Congo, Burundi, Rwanda). Therefore, while developing policies and initiating strategies to safeguard SOC, these regions should be cautiously handled. Overall, our uncertainty representation for Africa closely mirrors that of SoilGrids 2.0, as does the SOC map (Poggio et al., 2021). For the plant biomass C maps, uncertainty is relatively consistent in regions with very high soil C levels but is more obvious in regions between the Sahara desert and the tropical zone (Spawn et al., 2020). According to the Köppen climate classification, these regions between the Sahara desert and the tropical zone have semi-arid climatic conditions (i.e., BSh) (Beck et al., 2018).

3.2. Interlinking soil and plant biomass C across Africa

Our results revealed that plant biomass and soil C are highly correlated across Africa, offering promising prospects for the shared conservation of C storage in the African continent. Our structural equation models enabled us to explore the intricate relationships between plant biomass C and SOC, as well as various other environmental drivers (Fig. 2 and Supplementary Data Tables 3, 4, and 5). We identified soil nitrogen is highly correlated with SOC regardless of spatial scale since soil nitrogen generally controls C productivity (Vuichard et al., 2019) (Fig. 2 and Supplementary Data Fig. 7). Increased soil nitrogen availability can enhance plant growth and the amount of OM returned into the soil, leading to an overall increase in SOC levels (Wu et al., 2022). Conversely, high SOC levels can promote nutrient release through mineralization and create favorable conditions for plant access to soil nitrogen, contributing to improved soil fertility and plant productivity.

Our analyses further reveal that both AGBC and BGBC have a significant influence on SOC across Africa, especially for intermediate and coarse scales (Fig. 2, Supplementary Data Fig. 8). Thus, both are expected to contribute significantly to the observed SOC variability in Africa. Litter inputs from above-ground sources (e.g., leaves) have the potential to contribute to the accumulation of OM in soils through litter decomposition, consequently leading to an increase in SOC levels (Xu et al., 2022). Besides that, we observed that other environmental drivers, such as soil pH, land cover diversity, rainfall seasonality, and soil respiration, also play significant roles in facilitating the integration of C derived from litter into the soil, especially for coarse scale. For instance, we know that as pH increases (towards neutral or slightly alkaline), microbial activity tends to be more efficient, leading to higher levels of OM decomposition and, consequently, higher SOC levels (Zhou et al., 2020). Land cover diversity, however, influences regional microclimatic conditions, water circulation or availability, and nutrient cycling (Li et al., 2022), all of which we expect to affect AGBC.

Regions with diverse land cover types can potentially enhance SOC levels through varied C inputs or sources. This phenomenon is observed in countries like Morocco and Ethiopia, where diverse land cover types (e.g., croplands, forests, and shrublands) coupled with sustainable local land management practices (Etsay et al., 2019; Kusi et al., 2020), contribute to high SOC levels, as demonstrated in the earlier map (Fig. 1a). Similarly, BGBC enhances SOC storage through C inputs from root litter or rhizodeposits. Importantly, root litter tends to bind soil particles together, promoting the formation of soil aggregates. These aggregates act as a protective barrier against microbial degradation, thus enhancing SOC stability and long-term retention within the soil (Cotrufo et al., 2015). Like AGBC, the environmental drivers mentioned earlier (e.g., precipitation) also play crucial roles in shaping the influence of BGBC on SOC (Chen et al., 2018).

In addition to the influence of soil pH, as mentioned earlier, high rainfall seasonality positively affects SOC levels. The increased and

fluctuating moisture conditions associated with high rainfall seasonality promote robust plant growth, resulting in greater OM input into the soil through root litter and root exudates. However, increased moisture availability may also limit microbial activity, decelerating the decomposition of organic materials and the build-up of stable SOC (Stark and Firestone, 1995). Our SEM plots for total plant biomass C depicted expected common factors, considering both AGBC and BGBC influence SOC, albeit through distinct pathways. Nonetheless, we graphed the relationship of each with SOC separately to highlight marginal variations between the two.

We further illustrated how the influence of plant biomass C on SOC varies across different spatial scales by applying a simple and inherently intuitive GAM. As earlier mentioned, with the GAM, we precisely examined the non-linear relationships between the plant biomass C, environmental drivers, and their associations with SOC across Africa's vast environmental gradient (Supplementary Data Fig. 9). Our GAM results also confirmed the high dynamic and complex SOC processes across massive landscapes. We found that across all spatial scales, AGBC and BGBC shared consistent relationships with SOC. Because of the computation of the root-to-shoot ratios, we expected a strongly positive link between above- and below-ground biomass C, supporting such analogous relationships (Cairns et al., 1997). However, between the two plants biomass C and SOC, both the fine and coarse scales showed monotonous positive trends relative to monotonous negative trends observed for the intermediate scale (Supplementary Data Fig. 9). Our results ratified the already identified trade-offs between different plant biomass C relationships with SOC in literature (de la Cruz-Amo et al., 2020). We attributed these variations in trends to the complex, scale-dependent soil and ecosystem processes. In addition, we have established a consistent linear relationship between soil nitrogen and SOC across various spatial scales, reinforcing our earlier SEM results. Furthermore, our observations also indicate that environmental drivers, such as rainfall seasonality patterns, soil pH, land cover diversity, and soil phosphorus have distinct and influential effects on SOC, contingent upon the spatial scale. These findings show that soil and ecosystem processes vary greatly. The GAM accuracy results for each spatial scale in Supplementary Data Table 6 are all indicative of robust and generalizing models.

Our maps in Fig. 3 show how the distribution of SOC varies across scales while still maintaining fairly consistent spatial patterns, a comparable finding to earlier studies, for example (Holmes et al., 2005). We used each SOC map as a response variable in our GAMs. Our maps show relatively high SOC levels along the tropics while decreasing towards the north and south. As expected, we observed that regions with high SOC corresponded to high soil nitrogen levels (Supplementary Data Figs. 2, 3, and 4). While we examined the map at a coarse scale, we observed that variation in SOC becomes more evident, highlighting the cumulative low, moderate, and high SOC levels across different countries (as shown in Fig. 3c). For example, in the northern region of the continent, SOC ranged from $\leq$2.6 to 16.3 g/kg. We noticed Morocco exhibited a high SOC range between 13.7 and 16.3 g/kg, confirming the presence of active irrigation and improved cropping systems in this Mediterranean region (Ruiz-Peinado et al., 2013). We infer that in arid regions where SOC levels are low, implementing sustainable soil management practices locally may substantially enhance the overall SOC accrual and storage across landscapes. According to Baveye (2023), this is an endeavor to spark a new paradigm that focuses on SOC modelling following a bottom-up perspective or approach. That is, beginning at a highly detailed level (e.g., microscale, ecosystem-scale) and eventually expanding process understanding to encompass broader scales (e.g., massive landscapes, continents, etc.). This approach, if implemented in a sufficient number of study sites across Africa, should help address certain limitations associated with large-scale SOC models.

Lastly, we observed that the spatial overlaps between SOC and plant biomass C maintained similar patterns across different spatial scales (Fig. 4, Supplementary Data Fig. 10). Landscapes with high SOC and high plant biomass C levels were commonly found in the tropics (pompadour), which suggests that high plant biomass, including both AGBC and BGBC, is invariably linked to higher SOC levels. This relationship underscores the significance of vegetation and its role in the accumulation and storage of soil C. Furthermore, certain areas, particularly in the southern regions of Africa, exhibited high AGBC but had low SOC content (blue). These regions are characterized by predominantly sandy soils with less protected organic matter content, making them more susceptible to microbial decomposition and soil respiration. This vulnerability can potentially lead to reduced SOC levels, even in situations where there are substantial inputs of plant C. Furthermore, activities such as land clearance, fragmentation, or disturbance may be lowering SOC levels (Mani et al., 2021; Wang et al., 2023). Additionally, the absence of sustainable soil management practices, such as insufficient OM addition or erosion control, may limit SOC accumulation, even in the presence of ample vegetation. Again, we noted some regions with exclusively high SOC levels (yellow), which we assumed related to different C turnover processes. It would be interesting to investigate these regions further to understand the environmental conditions that might favor SOC accumulation. For example climate, soil texture, pH, and microbial activity (e.g., microbial carbon use efficiency – CUE, necromass, etc.) (Cotrufo et al., 2015; Tao et al., 2023) and others. For land managers and policymakers, these maps offer a chance to enhance their understanding of the spatial patterns, interlinks, and changes between the plant biomass C and SOC at the landscape level. Specifically, our results may support decisions concerning land use, ecosystem regeneration, and sustainable soil management practices geared towards improving SOC storage, accumulation, and long-term soil health. Additionally, our bivariate maps are instrumental for general ecosystem monitoring and conservation, as both plant biomass C and SOC in Africa call for joint conservation and safeguarding to mitigate challenges allied to climate change globally. It is worth noting that, despite the use of SOC content relative to stocks, the findings for either content or stocks mirror each other. We also present evidence from SEM analysis using SOC stocks (refer to Supplementary Data Fig. 12).

4. Conclusion

Overall, our modelling results provide compelling evidence that both plant and soil C exhibit major hotspots across Africa. This suggests that stakeholders—including policymakers, scientists, and environmentalists—must prioritize the simultaneous conservation of both plant biomass C and SOC, as this is crucial for mitigating global climate change. Furthermore, the advanced modelling techniques used in this study not only validate existing knowledge of Africa's tropical C hotspots but also underscore the power of statistical and mechanistic modelling in helping understand terrestrial C processes over large scales. This knowledge is indispensable for effective land-use planning, ecosystem restoration, and the advancement of sustainable soil management practices to ensure long-term soil health and conservation in Africa.

Funding

N.M.K was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645 which is highly appreciated. M.D-B. acknowledges support from TED2021-130908B-C41/AEI/10.13039/501100011033/Unión Europea NextGenerationEU/PRTR and the Spanish Ministry of Science and Innovation for the I + D + i project PID2020-115813RA-I00 funded by MCIN/AEI/10.13039/501100011033.

CRediT authorship contribution statement

Ndiye Michael Kebonye: Writing – review & editing, Writing –

original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kingsley John:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Manuel Delgado-Baquerizo:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Data curation, Conceptualization. **Yong Zhou:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Prince Chapman Agyeman:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation, Conceptualization. **Zibanani Seletlo:** Writing – review & editing, Writing – original draft, Validation, Data curation, Conceptualization. **Brandon Heung:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Thomas Scholten:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors in this work declare no competing interests.

Data availability

The SOC dataset used in this study was obtained from the World Soil Information Service (WoSIS): https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/2b643ef9-4bee-44d4-b50d-5020c9133b8b. The *R* codes used to run the analyses are available in an open repository: https://doi.org/10.5281/zenodo.10126384. These codes are systematically presented (Kebonye et al., 2023c).

Acknowledgements

We thank ISRIC World Soil Information for providing the standardized SOC data used in this study. Also, we would like to thank the anonymous reviewers for their time and effort in improving this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.175476.

References

Baethgen, W.E., Parton, W.J., Rubio, V., Kelly, R.H.M., Lutz, S., 2021. Ecosystem dynamics of crop–pasture rotations in a fifty-year field experiment in southern South America: century model and field results. Soil Sci. Soc. Am. J. 85, 423–437. https://doi.org/10.1002/saj2.20204.

Batjes, N.H., Ribeiro, E., van Oostrum, A., 2020. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth System Science Data 12, 299–320. https://doi.org/10.5194/essd-12-299-2020.

Baumann, F., He, J.-S., Schmidt, K., Kühn, P., Scholten, T., 2009. Pedogenesis, permafrost, and soil moisture as controlling factors for soil nitrogen and carbon contents across the Tibetan Plateau. Glob. Chang. Biol. 15, 3001–3017. https://doi.org/10.1111/j.1365-2486.2009.01953.x.

Baveye, P.C., 2023. Ecosystem-scale modelling of soil carbon dynamics: time for a radical shift of perspective? Soil Biol. Biochem. 184, 109112 https://doi.org/10.1016/j.soilbio.2023.109112.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 5, 180214. https://doi.org/10.1038/sdata.2018.214.

Berhongaray, G., Cotrufo, F.M., Janssens, I.A., Ceulemans, R., 2019. Below-ground carbon inputs contribute more than above-ground inputs to soil carbon accrual in a bioenergy poplar plantation. Plant Soil 434, 363–378. https://doi.org/10.1007/s11104-018-3850-z.

Berretta, A., 2020. Carbon Pricing in Sub-Saharan Africa. Konrad Adenauer Stiftung, Cape Town.

Bivand, R., Ono, H., Dunlap, R., Stigler, M., 2017. Choose Univariate Class Intervals [WWW Document]. Package 'classInt'. URL. http://cran.nexr.com/web/packages/classInt/classInt.pdf (accessed 12.9.23).

Bivand, R., Denney, B., Dunlap, R., Hernangómez, D., Ono, H., Parry, J., Stigler, M., 2023. classInt: Choose Univariate Class Intervals. [WWW Document]. classInt. URL. https://r-spatial.github.io/classInt/authors.html (accessed 10.26.23).

Blais, A.-M., Lorrain, S., Plourde, Y., Varfalvy, L., 2005. Organic carbon densities of soils and vegetation of tropical, temperate and boreal forests. In: Tremblay, A., Varfalvy, L., Roehm, C., Garneau, M. (Eds.), Greenhouse Gas Emissions — Fluxes and Processes: Hydroelectric Reservoirs and Natural Environments, Environmental Science. Springer, Berlin, Heidelberg, pp. 155–185. https://doi.org/10.1007/978-3-540-26643-3_7.

Bond-Lamberty, B., Bailey, V.L., Chen, M., Gough, C.M., Vargas, R., 2018. Globally rising soil heterotrophic respiration over recent decades. Nature 560, 80–83. https://doi.org/10.1038/s41586-018-0358-x.

Bradford, M.A., Berg, B., Maynard, D.S., Wieder, W.R., Wood, S.A., 2016. Understanding the dominant controls on litter decomposition. J. Ecol. 104, 229–238. https://doi.org/10.1111/1365-2745.12507.

Buchhorn, M., Smets, B., Bertels, L., Roo, B.D., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S., 2020. Copernicus Global Land Service: Land Cover 100m: Collection 3: Epoch 2015: Globe. https://doi.org/10.5281/zenodo.3939038.

Bukombe, B., Bauters, M., Boeckx, P., Cizungu, L.N., Cooper, M., Fiener, P., Kidinda, L.K., Makelele, I., Muhindo, D.I., Rewald, B., Verheyen, K., Doetterl, S., 2022. Soil geochemistry – and not topography – as a major driver of carbon allocation, stocks, and dynamics in forests and soils of African tropical montane ecosystems. New Phytol. 236, 1676–1690. https://doi.org/10.1111/nph.18469.

Cairns, M.A., Brown, S., Helmer, E.H., Baumgardner, G.A., 1997. Root biomass allocation in the World's upland forests. Oecologia 111, 1–11.

Chen, S., Wang, W., Xu, W., Wang, Yang, Wan, H., Chen, D., Tang, Z., Tang, X., Zhou, G., Xie, Z., Zhou, D., Shangguan, Z., Huang, J., He, J.-S., Wang, Yanfen, Sheng, J., Tang, L., Li, X., Dong, M., Wu, Y., Wang, Q., Wang, Z., Wu, J., Chapin, F.S., Bai, Y., 2018. Plant diversity enhances productivity and soil carbon storage. Proc. Natl. Acad. Sci. 115, 4027–4032. https://doi.org/10.1073/pnas.1700298114.

Chevalier, M., Mod, H., Broennimann, O., Di Cola, V., Schmid, S., Niculita-Hirzel, H., Pradervand, J.-N., Schmidt, B.R., Ursenbacher, S., Pellissier, L., Guisan, A., 2021. Low spatial autocorrelation in mountain biodiversity data and model residuals. Ecosphere 12, e03403. https://doi.org/10.1002/ecs2.3403.

Corbeels, M., Cardinael, R., Naudin, K., Guibert, H., Torquebiau, E., 2019. The 4 per 1000 goal and soil carbon storage under agroforestry and conservation agriculture systems in sub-Saharan Africa. Soil Tillage Res. 188, 16–26. https://doi.org/10.1016/j.still.2018.02.015. Soil Carbon and Climate Change: the 4 per Mille Initiative.

Cotrufo, M.F., Soong, J.L., Horton, A.J., Campbell, E.E., Haddix, M.L., Wall, D.H., Parton, W.J., 2015. Formation of soil organic matter via biochemical and physical pathways of litter mass loss. Nat. Geosci. 8, 776–779. https://doi.org/10.1038/ngeo2520.

Craven, D., van der Sande, M.T., Meyer, C., Gerstner, K., Bennett, J.M., Giling, D.P., Hines, J., Phillips, H.R.P., May, F., Bannar-Martin, K.H., Chase, J.M., Keil, P., 2020. A cross-scale assessment of productivity–diversity relationships. Glob. Ecol. Biogeogr. 29, 1940–1955. https://doi.org/10.1111/geb.13165.

Dai, W., Chen, X., 2012. Entropy of function of uncertain variables. Math. Comput. Model. 55, 754–760. https://doi.org/10.1016/j.mcm.2011.08.052.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A., Montanarella, L., van Wesemael, B., 2015. A map of the topsoil organic carbon content of Europe generated by a generalized additive model. Eur. J. Soil Sci. 66, 121–134. https://doi.org/10.1111/ejss.12193.

de la Cruz-Amo, L., Bañares-de-Dios, G., Cala, V., Granzow-de la Cerda, Í., Espinosa, C.I., Ledo, A., Salinas, N., Macía, M.J., Cayuela, L., 2020. Trade-offs among aboveground, belowground, and soil organic carbon stocks along altitudinal gradients in Andean tropical montane forests. Front. Plant Sci. 11.

Delgado-Baquerizo, M., Eldridge, D.J., Maestre, F.T., Karunaratne, S.B., Trivedi, P., Reich, P.B., Singh, B.K., 2017. Climate legacies drive global soil carbon stocks in terrestrial ecosystems. Sci. Adv. 3, e1602008 https://doi.org/10.1126/sciadv.1602008.

Dietzel, R., Liebman, M., Archontoulis, S., 2017. A deeper look at the relationship between root carbon pools and the vertical distribution of the soil carbon pool. SOIL 3, 139–152. https://doi.org/10.5194/soil-3-139-2017.

Eisenhauer, N., Bowker, M.A., Grace, J.B., Powell, J.R., 2015. From patterns to causal understanding: structural equation modeling (SEM) in soil ecology. Pedobiologia 58, 65–72. https://doi.org/10.1016/j.pedobi.2015.03.002.

Etsay, H., Negash, T., Aregay, M., 2019. Factors that influence the implementation of sustainable land management practices by rural households in Tigrai region, Ethiopia. Ecol. Process. 8, 14. https://doi.org/10.1186/s13717-019-0166-8.

European Union, 2018. Regions in the European Union. European Union.

Eusterhues, K., Rumpel, C., Kleber, M., Kögel-Knabner, I., 2003. Stabilisation of soil organic matter by interactions with minerals as revealed by mineral dissolution and oxidative degradation. Org. Geochem. 34, 1591–1600. https://doi.org/10.1016/j.orggeochem.2003.08.007.

Fatichi, S., Pappas, C., Zscheischler, J., Leuzinger, S., 2019. Modelling carbon sources and sinks in terrestrial vegetation. New Phytol. 221, 652–668. https://doi.org/10.1111/nph.15451.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086.

Garrido, M., Hansen, S.K., Yaari, R., Hawlena, H., 2022. A model selection approach to structural equation modelling: a critical evaluation and a road map for ecologists. Methods Ecol. Evol. 13, 42–53. https://doi.org/10.1111/2041-210X.13742.

Geng, Y., Ma, W., Wang, L., Baumann, F., Kühn, P., Scholten, T., He, J.-S., 2017. Linking above- and belowground traits to soil and climate variables: an integrated database on China's grassland species. Ecology 98, 1471. https://doi.org/10.1002/ecy.1780.
Grace, P., Ladd, J., Robertson, G., Gage, S., 2006. SOCRATES—A simple model for predicting long-term changes in soil organic carbon in terrestrial ecosystems. Soil Biol. Biochem. 38, 1172–1176. https://doi.org/10.1016/j.soilbio.2005.09.013.
Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314.
Hastie, T., 2023. Generalized Additive Models [WWW Document]. Package 'gam'. URL. https://cran.r-project.org/web/packages/gam/gam.pdf (accessed 10.16.23).
Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398. https://doi.org/10.1016/j.geoderma.2004.06.007.
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Jesus, J.M. de, Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10, e0125814. https://doi.org/10.1371/journal.pone.0125814.
Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G. E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.-M., Chamberlin, J., Silatsa, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci. Rep. 11, 6130. https://doi.org/10.1038/s41598-021-85639-y.
Hobley, E., Steffens, M., Bauke, S.L., Kögel-Knabner, I., 2018. Hotspots of soil organic carbon storage revealed by laboratory hyperspectral imaging. Sci. Rep. 8, 13900. https://doi.org/10.1038/s41598-018-31776-w.
Holmes, K.W., Kyriakidis, P.C., Chadwick, O.A., Soares, J.V., Roberts, D.A., 2005. Multi-scale variability in tropical soil nutrients following land-cover change. Biogeochemistry 74, 173–203. https://doi.org/10.1007/s10533-004-3544-x.
Joseph, V.R., Vakayil, A., 2022. SPlit: an optimal method for data splitting. Technometrics 64, 166–176. https://doi.org/10.1080/00401706.2021.1921037.
Kätterer, T., Bolinder, M.A., Andrén, O., Kirchmann, H., Menichetti, L., 2011. Roots contribute more to refractory soil organic matter than above-ground crop residues, as revealed by a long-term field experiment. Agric. Ecosyst. Environ. 141, 184–192. https://doi.org/10.1016/j.agee.2011.02.029.
Kebonye, N.M., 2021. Exploring the novel support points-based split method on a soil dataset. Measurement 186, 110131. https://doi.org/10.1016/j.measurement.2021.110131.
Kebonye, N.M., Agyeman, P.C., Biney, J.K.M., 2023a. Optimized modelling of countrywide soil organic carbon levels via an interpretable decision tree. Smart Agricultural Technology 3, 100106. https://doi.org/10.1016/j.atech.2022.100106.
Kebonye, N.M., Agyeman, P.C., Seletlo, Z., Eze, P.N., 2023b. On exploring bivariate and trivariate maps as visualization tools for spatial associations in digital soil mapping: a focus on soil properties. Precis. Agric. 24, 511–532. https://doi.org/10.1007/s11119-022-09955-7.
Kebonye, N.M., John, K., Delgado-Baquerizo, M., Zhou, Y., Agyeman, P.C., Seletlo, Z., Heung, B., Scholten, T., 2023c. Major overlap in plant and soil carbon hotspots across Africa (Version 1). In: Zenodo. https://doi.org/10.5281/zenodo.10126384.
Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model. 81, 401–418. https://doi.org/10.1016/j.apm.2019.12.016.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., 2023. Classification and Regression Training [WWW Document]. Package 'caret'. URL. https://cran.r-project.org/web/packages/caret/caret.pdf (accessed 12.9.23).
Kusi, K.K., Khattabi, A., Mhammdi, N., Lahssini, S., 2020. Prospective evaluation of the impact of land use change on ecosystem services in the Ourika watershed, Morocco. Land Use Policy 97, 104796. https://doi.org/10.1016/j.landusepol.2020.104796.
Li, Y., Liu, W., Feng, Q., Zhu, M., Yang, L., Zhang, J., 2022. Effects of land use and land cover change on soil organic carbon storage in the Hexi regions, Northwest China. J. Environ. Manag. 312, 114911 https://doi.org/10.1016/j.jenvman.2022.114911.
Luo, Z., Wang, G., Wang, E., 2019. Global subsoil organic carbon turnover times dominantly controlled by soil properties rather than climate. Nat. Commun. 10, 3688. https://doi.org/10.1038/s41467-019-11597-9.
Mani, S., Osborne, C.P., Cleaver, F., 2021. Land degradation in South Africa: justice and climate change in tension. People and Nature 3, 978–989. https://doi.org/10.1002/pan3.10260.
Martens, J., Mueller, C.W., Joshi, P., Rosinger, C., Maisch, M., Kappler, A., Bonkowski, M., Schwamborn, G., Schirrmeister, L., Rethemeyer, J., 2023. Stabilization of mineral-associated organic carbon in Pleistocene permafrost. Nat. Commun. 14, 2120. https://doi.org/10.1038/s41467-023-37766-5.
Martínez-Núñez, C., Martínez-Prentice, R., García-Navas, V., 2023. Land-use diversity predicts regional bird taxonomic and functional richness worldwide. Nat. Commun. 14, 1320. https://doi.org/10.1038/s41467-023-37027-5.
Mayer, M., Sandén, H., Rewald, B., Godbold, D.L., Katzensteiner, K., 2017. Increase in heterotrophic soil respiration by temperature drives decline in soil organic carbon stocks after forest windthrow in a mountainous ecosystem. Funct. Ecol. 31, 1163–1172. https://doi.org/10.1111/1365-2435.12805.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.
Merbach, I., Schulz, E., 2013. Long-term fertilization effects on crop yields, soil fertility and sustainability in the Static Fertilization Experiment Bad Lauchstädt under climatic conditions 2001–2010. Arch. Agron. Soil Sci. 59, 1041–1057. https://doi.org/10.1080/03650340.2012.702895.
Møller, A.B., Beucher, A.M., Pouladi, N., Greve, M.H., 2020. Oblique geographic coordinates as covariates for digital soil mapping. SOIL 6, 269–289. https://doi.org/10.5194/soil-6-269-2020.
Mugabowindekwe, M., Brandt, M., Chave, J., Reiner, F., Skole, D.L., Kariryaa, A., Igel, C., Hiernaux, P., Ciais, P., Mertz, O., Tong, X., Li, S., Rwanyiziri, G., Dushimiyimana, T., Ndoli, A., Uwizeyimana, V., Lillesø, J.-P.B., Gieseke, F., Tucker, C.J., Saatchi, S., Fensholt, R., 2023. Nation-wide mapping of tree-level aboveground carbon stocks in Rwanda. Nat. Clim. Chang. 13, 91–97. https://doi.org/10.1038/s41558-022-01544-w.
Mugabowindekwe, M., Brandt, M., Mukuralinda, A., Ciais, P., Reiner, F., Kariryaa, A., Igel, C., Chave, J., Mertz, O., Hiernaux, P., Tong, X., Rwanyiziri, G., Gominski, D., Li, S., Liu, S., Gasangwa, I., Hategekimana, Y., Ndoli, A., Nduwamungu, J., Saatchi, S., Fensholt, R., 2024. Trees on smallholder farms and forest restoration are critical for Rwanda to achieve net zero emissions. Commun Earth Environ 5, 1–10. https://doi.org/10.1038/s43247-024-01278-x.
Nol, L., Heuvelink, G.B.M., Veldkamp, A., de Vries, W., Kros, J., 2010. Uncertainty propagation analysis of an N2O emission model at the plot and landscape scale. Geoderma 159, 9–23. https://doi.org/10.1016/j.geoderma.2010.06.009.
Nyberg, M., Hovenden, M.J., 2020. Warming increases soil respiration in a carbon-rich soil without changing microbial respiratory potential. Biogeosciences 17, 4405–4420. https://doi.org/10.5194/bg-17-4405-2020.
Ontl, T.A., Schulte, L.A., 2012. Soil Carbon Storage. Nature Education Knowledge, 3, p. 35.
Ottoy, S., Truyers, E., De Block, M., Lettens, S., Swinnen, W., Broothaerts, N., Hendrix, R., Van Orshoven, J., Verstraeten, G., De Vos, B., Vancampenhout, K., 2022. Digital mapping of soil organic carbon hotspots in nature conservation areas in the region of Flanders, Belgium. Geoderma Regional 30, e00531. https://doi.org/10.1016/j.geodrs.2022.e00531.
Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7, 217–240. https://doi.org/10.5194/soil-7-217-2021.
Quinlan, J.R., 1992. Learning with Continuous Classes, in: In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence. Hobart, Australia, pp. 343–348.
Rasse, D.P., Rumpel, C., Dignac, M.-F., 2005. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. Plant Soil 269, 341–356. https://doi.org/10.1007/s11104-004-0907-y.
Richards, J., Brimblecombe, P., Engelstaedter, S., 2023. Modelling temperature-precipitation pressures on African timber heritage. Int. J. Climatol. n/a, 1–16. https://doi.org/10.1002/joc.8273.
Rocchini, D., Thouverai, E., Marcantonio, M., Iannacito, M., Da Re, D., Torresani, M., Bacaro, G., Bazzichetto, M., Bernardi, A., Foody, G.M., Furrer, R., Kleijn, D., Larsen, S., Lenoir, J., Malavasi, M., Marchetto, E., Messori, F., Montaghi, A., Moudrý, V., Naimi, B., Ricotta, C., Rossini, M., Santi, F., Santos, M.J., Schaepman, M. E., Schneider, F.D., Schuh, L., Silvestri, S., Šímová, P., Skidmore, A.K., Tattoni, C., Tordoni, E., Vicario, S., Zannini, P., Wegmann, M., 2021. rasterdiv—an information theory tailored R package for measuring ecosystem heterogeneity from space: to the origin and back. Methods Ecol. Evol. 12, 1093–1102. https://doi.org/10.1111/2041-210X.13583.
Ross, C.W., Prihodko, L., Anchang, J., Kumar, S., Ji, W., Hanan, N.P., 2018. HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. Sci Data 5, 180091. https://doi.org/10.1038/sdata.2018.91.
Rosseel, Y., 2012. lavaan: An R package for structural equation modeling. J. Stat. Softw. 48, 1–36. https://doi.org/10.18637/jss.v048.i02.
RStudio Team, 2020. RStudio: Integrated Development Environment for R.
Ruiz-Peinado, R., Bravo-Oviedo, A., López-Senespleda, E., Montero, G., Río, M., 2013. Do thinnings influence biomass and soil carbon stocks in Mediterranean maritime pinewoods? Eur. J. For. Res. 132, 253–262.
Siles, J.A., Vera, A., Díaz-López, M., García, C., van den Hoogen, J., Crowther, T.W., Eisenhauer, N., Guerra, C., Jones, A., Orgiazzi, A., Delgado-Baquerizo, M., Bastida, F., 2023. Land-use- and climate-mediated variations in soil bacterial and fungal biomass across Europe and their driving factors. Geoderma 434, 116474. https://doi.org/10.1016/j.geoderma.2023.116474.
Spawn, S.A., Sullivan, C.C., Lark, T.J., Gibbs, H.K., 2020. Harmonized global maps of above and belowground biomass carbon density in the year 2010. Sci Data 7, 112. https://doi.org/10.1038/s41597-020-0444-4.
Spohn, M., Braun, S., Sierra, C.A., 2023. Continuous decrease in soil organic matter despite increased plant productivity in an 80-years-old phosphorus-addition experiment. Commun Earth Environ 4, 1–10. https://doi.org/10.1038/s43247-023-00915-1.
Stark, J.M., Firestone, M.K., 1995. Mechanisms for soil moisture effects on activity of nitrifying bacteria. Appl. Environ. Microbiol. 61, 218–221. https://doi.org/10.1128/aem.61.1.218-221.1995.
Tao, F., Huang, Y., Hungate, B.A., Manzoni, S., Frey, S.D., Schmidt, M.W.I., Reichstein, M., Carvalhais, N., Ciais, P., Jiang, L., Lehmann, J., Wang, Y.-P., Houlton, B.Z., Ahrens, B., Mishra, U., Hugelius, G., Hocking, T.D., Lu, X., Shi, Z., Viatkin, K., Vargas, R., Yigini, Y., Omuto, C., Malik, A.A., Peralta, G., Cuevas-Corona, R., Di Paolo, L.E., Luotto, I., Liao, C., Liang, Y.-S., Saynes, V.S., Huang, X., Luo, Y., 2023. Microbial carbon use efficiency promotes global soil carbon storage. Nature 1–5. https://doi.org/10.1038/s41586-023-06042-3.
Timilsina, N., Escobedo, F.J., Cropper, W.P., Abd-Elrahman, A., Brandeis, T.J., Delphin, S., Lambert, S., 2013. A framework for identifying carbon hotspots and forest management drivers. J. Environ. Manag. 114, 293–302. https://doi.org/10.1016/j.jenvman.2012.10.020.

Vågen, T.-G., Winowiecki, L.A., Tondoh, J.E., Desta, L.T., Gumbricht, T., 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. Geoderma 263, 216–225. https://doi.org/10.1016/j.geoderma.2015.06.023.

Védère, C., Lebrun, M., Honvault, N., Aubertin, M.-L., Girardin, C., Garnier, P., Dignac, M.-F., Houben, D., Rumpel, C., 2022. How does soil water status influence the fate of soil organic matter? A review of processes across scales. Earth Sci. Rev. 234, 104214 https://doi.org/10.1016/j.earscirev.2022.104214.

Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res. 53, 845. https://doi.org/10.1071/SR14366.

von Fromm, S.F., Hoyt, A.M., Lange, M., Acquah, G.E., Aynekulu, E., Berhe, A.A., Haefele, S.M., McGrath, S.P., Shepherd, K.D., Sila, A.M., Six, J., Towett, E.K., Trumbore, S.E., Vågen, T.-G., Weullow, E., Winowiecki, L.A., Doetterl, S., 2021. Continental-scale controls on soil organic carbon across sub-Saharan Africa. SOIL 7, 305–332. https://doi.org/10.5194/soil-7-305-2021.

von Fromm, S.F., Doetterl, S., Butler, B.M., Aynekulu, E., Berhe, A.A., Haefele, S.M., McGrath, S.P., Shepherd, K.D., Six, J., Tamene, L., Tondoh, E.J., Vågen, T.-G., Winowiecki, L.A., Trumbore, S.E., Hoyt, A.M., 2024. Controls on timescales of soil organic carbon persistence across sub-Saharan Africa. Glob. Chang. Biol. 30, e17089 https://doi.org/10.1111/gcb.17089.

Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., Peylin, P., 2019. Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary production. Geosci. Model Dev. 12, 4751–4779. https://doi.org/10.5194/gmd-12-4751-2019.

Wadoux, A.M.J.-C., Heuvelink, G.B.M., 2023. Uncertainty of spatial averages and totals of natural resource maps. Methods Ecol. Evol. 14, 1320–1332. https://doi.org/10.1111/2041-210X.14106.

Walden, L., Serrano, O., Zhang, M., Shen, Z., Sippo, J.Z., Bennett, L.T., Maher, D.T., Lovelock, C.E., Macreadie, P.I., Gorham, C., Lafratta, A., Lavery, P.S., Mosley, L., Reithmaier, G.M.S., Kelleway, J.J., Dittmann, S., Adame, F., Duarte, C.M., Gallagher, J.B., Waryszak, P., Carnell, P., Kasel, S., Hinko-Najera, N., Hassan, R., Goddard, M., Jones, A.R., Viscarra Rossel, R.A., 2023. Multi-scale mapping of Australia's terrestrial and blue carbon stocks and their continental and bioregional drivers. Commun Earth Environ 4, 1–12. https://doi.org/10.1038/s43247-023-00838-x.

Wang, M., Guo, X., Zhang, S., Xiao, L., Mishra, U., Yang, Y., Zhu, B., Wang, G., Mao, X., Qian, T., Jiang, T., Shi, Z., Luo, Z., 2022. Global soil profiles indicate depth-dependent soil carbon losses under a warmer climate. Nat. Commun. 13, 5514. https://doi.org/10.1038/s41467-022-33278-w.

Wang, Y., Luo, G., Li, C., Ye, H., Shi, H., Fan, B., Zhang, W., Zhang, C., Xie, M., Zhang, Y., 2023. Effects of land clearing for agriculture on soil organic carbon stocks in drylands: a meta-analysis. Glob. Chang. Biol. 29, 547–562. https://doi.org/10.1111/gcb.16481.

Warner, D.L., Bond-Lamberty, B., Jian, J., Stell, E., Vargas, R., 2019. Spatial predictions and associated uncertainty of annual soil respiration at the global scale. Glob. Biogeochem. Cycles 33, 1733–1745. https://doi.org/10.1029/2019GB006264.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - a review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026.

Wu, X., Wang, L., An, J., Wang, Y., Song, H., Wu, Y., Liu, Q., 2022. Relationship between soil organic carbon, soil nutrients, and land use in Linyi City (East China). Sustainability 14, 13585.

Xu, Y., Liu, K., Yao, S., Zhang, Y., Zhang, X., He, H., Feng, W., Ndzana, G.M., Chenu, C., Olk, D.C., Mao, J., Zhang, B., 2022. Formation efficiency of soil organic matter from plant litter is governed by clay mineral type more than plant litter quality. Geoderma 412, 115727. https://doi.org/10.1016/j.geoderma.2022.115727.

Zech, S., Schweizer, S.A., Bucka, F.B., Ray, N., Kögel-Knabner, I., Prechtel, A., 2022. Explicit spatial modeling at the pore scale unravels the interplay of soil organic carbon storage and structure dynamics. Glob. Chang. Biol. 28, 4589–4604. https://doi.org/10.1111/gcb.16230.

Zhou, W., Han, G., Liu, M., Zeng, J., Liang, B., Liu, J., Qu, R., 2020. Determining the distribution and interaction of soil organic carbon, nitrogen, pH and texture in soil profiles: a case study in the Lancangjiang River Basin, Southwest China. Forests 11, 532.

Zhou, Y., Bomfim, B., Bond, W.J., Boutton, T.W., Case, M.F., Coetsee, C., Davies, A.B., February, E.C., Gray, E.F., Silva, L.C.R., Wright, J.L., Staver, A.C., 2023. Soil carbon in tropical savannas mostly derived from grasses. Nat. Geosci. 16, 710–716. https://doi.org/10.1038/s41561-023-01232-0.

Zhu, A.-X., 1997. Measuring Uncertainty in Class Assignment for Natural Resource Maps under Fuzzy Logic.

Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the global aridity index and potential evapotranspiration database. Sci Data 9, 409. https://doi.org/10.1038/s41597-022-01493-1.