https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Enhancing the reliability of baseline C stock digital inventory of Sri Lanka through a dense sampling

Udaya W.A. Vitharana[1], Sagar Gautam[2], Chang Gyo Jung[2], Umakant Mishra[2]

[1]Department of Soil Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, 20400, Sri Lanka
[2]Bioscience Division, Sandia National Laboratory, Livermore, CA 94550, USA

Correspondence to: Udaya W.A. Vitharana (uvithara@agri.pdn.ac.lk)

Abstract. Accurate national inventories of soil organic carbon (SOC) are essential for quantifying terrestrial carbon storage and supporting land-management and climate-mitigation strategies. However, estimates are often constrained by sparse SOC measurements and global products that inadequately represent regional environmental heterogeneity. Here, we developed a high-resolution (30 x 30 m inventory of surface SOC stocks (0-30 cm) across Sri Lanka using a SOC data obtained using a spatially balanced network of 1,488 soil sampling locations, environmental covariates, and machine-learning-based digital soil mapping. Random Forest provided the best predictive performance, explaining 85 % of the training variability (Coefficient of determination (R^2) = 0.85; Root mean square error (RMSE) = 1.39 kg C m^{-2}, Lin's concordance correlation coefficient (CCC) = 0.90) and retaining prediction accuracy when evaluated on an independent validation dataset (R^2 = 0.51; RMSE = 2.36 kg C m^{-2}; CCC = 0.67). SOC stocks were highly heterogeneous, with a median of 5.4 kg C m^{-2} and an estimated national stock of 382.4 Tg C. The largest stocks occurred in the central highlands, whereas warmer and drier lowlands stored substantially less carbon. Shapley additive explanations (SHAP) analysis identified aridity index, mean annual precipitation, potential evapotranspiration, mean temperature of warmest and driest quarters, and solar radiation as dominant controls. These drivers operated through contrasting pathways: cooler temperatures constrained decomposition in the highlands, while moisture availability regulated carbon inputs across the lowlands. Tea and rubber plantations supported comparatively high SOC stocks. Benchmarking showed that the updated map substantially outperformed the previous Sri Lankan inventory developed in 2019 using sparse soil SOC data, SoilGrids v2.0, and GSOCmap v1.5.0. Our results demonstrate that dense, representative sampling combined with machine learning can improve national SOC inventories and provide a robust baseline for monitoring soil-carbon change and assessing climate vulnerability in heterogeneous tropical landscapes.

1 Introduction

Soil organic carbon (SOC) represents the largest actively cycling carbon reservoir in terrestrial ecosystems and plays a central role in regulating exchanges of carbon between the land and atmosphere. When deeper soil horizons are considered, global SOC stocks may reach approximately 3,100 Pg C (Friedlingstein et al., 2026). Because of its large pool size, even relatively small changes in this reservoir can substantially influence the terrestrial carbon balance. SOC stocks are sensitive to changes

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

in temperature, precipitation, soil moisture, vegetation, and land management, and soils can therefore function either as a net carbon sink or source depending on the balance between organic carbon inputs and losses. Estimates suggest that improved soil management could offset approximately 1–16 % of annual anthropogenic greenhouse gas emissions, although the achievable magnitude varies widely among ecosystems, soils, and management systems (Skea et al., 2022). Whether SOC accumulates or declines is ultimately governed by interactions among intrinsic soil characteristics, including mineralogy, texture, SOC composition and stabilization mechanisms (Zhang et al., 2023), and extrinsic controls such as hydrothermal conditions, vegetation, land use, climatic conditions and management practices (He et al., 2024). Beyond its role in the global carbon cycle, SOC is a fundamental indicator of soil health because it influences nutrient availability, soil structure, aggregate stability, water retention, microbial activity, and overall ecosystem functioning. Maintaining or increasing SOC is therefore closely linked to sustainable agricultural production, resilience to environmental stress, and long-term food security (Kopittke et al., 2022). However, SOC is characterized by pronounced spatial heterogeneity resulting from interactions among soil-forming factors, topography, vegetation, environmental conditions, and human management. This heterogeneity occurs from field to continental scales and presents a major challenge for accurately estimating SOC stocks. Resolving such variability at the national scale is particularly important because reliable baseline estimates provide the foundation for detecting future SOC changes, identifying areas vulnerable to carbon loss, targeting management interventions, and evaluating opportunities for additional carbon sequestration.

Robust national SOC inventories also have important implications for environmental policy and carbon accounting. Spatially explicit baseline SOC information is required to support monitoring, reporting, and verification of land-sector carbon changes and can contribute to national greenhouse-gas inventories developed under Intergovernmental Panel on Climate Change (IPCC) reporting frameworks. Such information can also support assessment of progress toward nationally determined contributions and land degradation neutrality targets under the United Nations Convention to Combat Desertification (UNCCD). In addition, observational SOC datasets provide important benchmarks for evaluating and constraining Earth system models, in which uncertainty in the magnitude and spatial distribution of soil carbon remains a major source of uncertainty in simulated terrestrial carbon-cycle feedbacks (Dai et al., 2019). Consequently, improving the spatial representation of SOC stocks has relevance extending from local soil management to national policy and global Earth system prediction. Obtaining reliable national SOC inventories, however, remains challenging. Capturing fine-scale spatial heterogeneity requires sufficiently dense and spatially representative soil sampling, harmonized analytical procedures, and coordinated field and laboratory programs. Such national-scale campaigns are costly and labor intensive, particularly in regions characterized by strong environmental gradients and diverse land-use systems. Consequently, many countries rely partly or entirely on regional or global soil databases to establish baseline SOC stocks. Several publicly available global databases of SOC stocks are currently available, including the Harmonized World Soil Database version 2.0 (HWSD v2.0, Nachtergaele et al., 2023), SoilGrids version 2.0 (SoilGrids v2.0, Poggio et al., 2021), and the FAO Global Soil Organic Carbon Map (GSOCmap v1.5.0, http://54.229.242.119/GSOCmap/). These products have substantially advanced global soil assessment and are particularly valuable where national observations are sparse.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Nevertheless, global SOC products may perform inconsistently at national and subnational scales because they often cannot fully capture local environmental variability and data limitations. Their accuracy depends on the density and representativeness of the underlying soil observations, harmonization of legacy measurements, environmental covariates, and the ability of statistical or machine-learning models to represent complex interactions controlling SOC distributions. Global models must also accommodate highly heterogeneous environmental conditions across continents and biomes, which may limit their ability to resolve relationships specific to individual countries or regions. Consequently, substantial discrepancies can occur between globally predicted SOC stocks and locally observed values. Previous assessments in Sri Lanka, for example, have identified inconsistencies between global soil products and field observations even when evaluated using relatively limited number of SOC derived from soil profiles (Vitharana et al., 2019). Systematic evaluation against dense, independent national observations is therefore necessary to quantify uncertainties in global products and determine whether country-specific models can provide more reliable estimates.

Sri Lanka provides a particularly suitable setting for addressing these questions. Despite its relatively small land area, the island encompasses exceptionally strong gradients in elevation, precipitation, temperature, soil type, vegetation, and land use over short geographic distances. Conditions range from humid central highlands and south-western regions to substantially drier lowland environments, creating pronounced spatial variability in soil-forming processes and SOC accumulation. Such heterogeneity poses a major challenge for generalized global mapping approaches but also provides an opportunity to examine how interacting environmental controls regulate SOC distribution across a tropical landscape. A dense and spatially balanced national sampling network, combined with environmental covariates and machine-learning approaches, can therefore provide a stronger basis for resolving national SOC patterns and evaluating the performance of existing global products.

Here, we use an intensive and spatially balanced soil sampling network within a machine-learning framework to develop a high-resolution national assessment of SOC stocks across Sri Lanka. Specifically, we aim to (1) quantify the magnitude and spatial heterogeneity of SOC stocks across the country; (2) identify the dominant environmental controls and interactions governing the spatial distribution of SOC; (3) estimate the total national SOC stock; and (4) evaluate the accuracy and precision of widely used global SOC datasets, including SoilGrids v2.0 and GSOCmap v1.5.0, as well as an existing national legacy dataset, against field observations. By integrating dense national observations with spatial modeling and independent validation, this study provides an updated SOC baseline for Sri Lanka and demonstrates the importance of country-specific datasets for improving soil carbon inventories in environmentally heterogeneous tropical regions.

2 Materials and methods

2.1 Study area, soil sampling and analysis

This study covers the tropical Island Sri Lanka (central coordinates : 7°20′ N and 80°45′ E, Area: 65630 km^2) which falls mainly within the tropical rainforest climate (Af) category (Kottek et al., 2006). We used a spatially balanced sampling scheme with 1488 sampling locations (Fig. 1) to obtain surface soil samples (0-30 cm) representing different landcover and land use

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

(LULC) types. Our sample size corresponds to proposed sample size by Vitharana et al. (2019) for Sri Lanka through spatial analysis of SOC stocks using legacy soil profile data. The land area of Sri Lanka was divided into 20 km x 20 km square grid cells and within each cell, areas of major land use types were digitized visually using Google Earth satellite images. These LULC categories included, tree covered areas, grasslands, crop lands, wetlands, artificial surfaces and other lands categories identified by UNCCD (UNCCD, 2022). We used these classes because this study aimed to create a baseline soil organic carbon map to support UNCCD land degradation neutrality reporting.

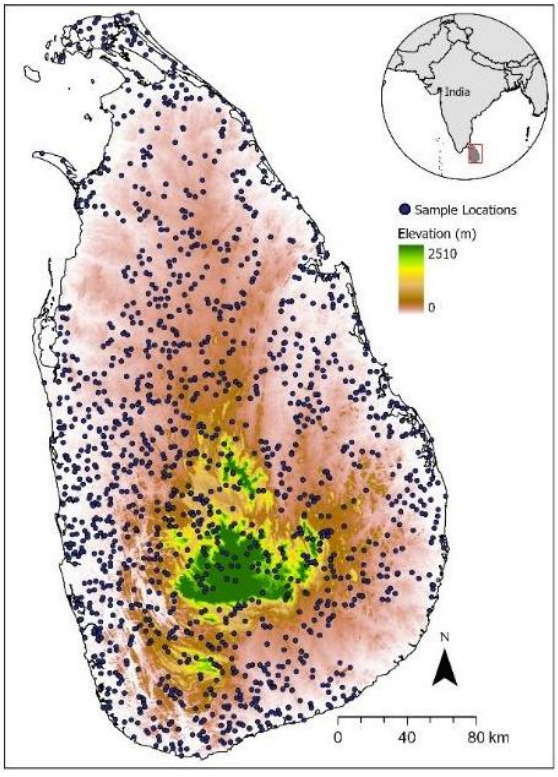


Figure 1: Soil profile observation locations (n = 1498) overlaid on the elevation map of Sri Lanka.

The total number of samples was distributed across grid cells proportionate to their total landcover within each cell. Then spatial locations of samples within each cell were found randomly by applying conditioned Latin hypercube sampling (Minasny and McBratney, 2006) while considering land-use type as the covariate for identification of sample locations. Field sampling was completed through a rigorous one-year campaign conducted from mid-2021. However, highly protected forest reserves were excluded from this study due to access restrictions (data gaps in the northwest and southeast sectors, as illustrated in Fig. 1). Coordinates and the present LULC type of each sampling point were recorded by the field team. We re-classified these LULC types into 12 classes considering the potential impacts of human intervention and vegetation on SOC storage (Table A1).

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

An undisturbed soil core sample was taken at each sampling site at 15 cm depth, representing the surface soil (0–30 cm). These samples were dried at 105 °C to a constant mass to find the bulk density (D_b, g cm^{-3}):

$$D_b = \frac{M_{dc}}{V_c} \quad (1)$$

where, M_{dc} and V_c are the dried soil mass of undisturbed core sample (g) and inner volume of the soil core (cm^{-3}), respectively. Subsequently, the gravel particles (>2 mm) were separated from the oven dried sample by wet sieving the sample through a 2 mm sieve. Then, gravel particles were oven dried to find the gravel percentage (G_p):

$$G_p = \frac{M_{dg}}{M_{dc}} \times 100 \quad (2)$$

where, M_{dg} is the oven dried mass of gravel particles. Subsequently, bulk density of fine earth fraction of soil ($D_{b<2\ mm}$) was computed using the following equation:

$$D_{b<2mm} = D_b \times \left(1 - \frac{G_p}{100}\right) \quad (3)$$

Three disturbed soil samples (0-30 cm) were taken within a circle of 1 m radius from each sample location using a Gauge auger and mixed to obtain a composite sample. Composite samples were air-dried, crushed and sieved through a 2 mm mesh sized sieve. Organic matter contents of samples were determined using the loss on ignition (LOI) method (Nelson and Sommers, 1996). This method relies on igniting organic matter particles at high temperature. Facilitating complete combustion, samples were ground to a particle size <0.4 mm using an agate mortar and pistil. Then, ground samples (3 g) were oven dried (105 °C) for 24 hr and subsequently ignited in a muffle furnace for 16 hours at a temperature of 400 °C. Then the percentage soil organic matter (SOM) content was computed using the following equation:

$$SOM = \left(\frac{M_{OD} - M_{IG}}{M_{OD}}\right) \times 100 \quad (4)$$

where, M_{OD} and M_{IG} are masses (g) of oven dried and muffle furnace ignited sample, respectively. The loss of mass after ignition was reported as the percentage organic matter content. Because a country-specific SOM-to-SOC conversion factor is unavailable for Sri Lanka, we adopted the factor of 0.5 recommended by Pribyl (2010), rather than the conventional factor of 0.58, to avoid potential overestimations. Soil organic carbon stock (SOC stock, kg m^{-2}) was computed using the following equation:

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$SOC\ stock = SOC\ \% \times D_{b<2mm} \times D\ \times \frac{1}{10} \quad (5)$$

where, D denotes the soil depth (30 cm) and $D_{b<2mm}$ is in g cm^{-3}.

1.1 Selection of intrinsic and extrinsic controllers of SOC stocks

We used a range (n = 38) of potential intrinsic and extrinsic controllers of SOC stocks (Table A2) representing the SCORPAN model (McBratney et al., 2003) to model and map the spatial variation of the SOC stocks. We used five layers of legacy soil information, order, suborder, great group, subgroup and series, representing explanatory variables for soils (S), parent material (P) and age (A). First, soil series maps of Sri Lanka (Mapa, 2020; Mapa et al., 2010) were aggregated to a single map and the soil series name of each map unit was extracted from the attribute database. In cases where map units consisted of soil complexes or associations, the soil series was considered representative of the respective map units. Subsequently, soil taxonomy (Soil Survey Staff, 2014) subgroup name of each map unit was extracted from the attribute database. Finally, these subgroup names were disaggregated to great groups, suborders and orders. Climatic (C) variables were obtained from the WorldClim version 2.1 climate data for 1970-2000, these included 19 high resolution biometric variables and average solar radiation (Fick and Hijmans, 2017). In addition, we incorporated the aridity index and potential evapotranspiration (Zomer et al., 2022). The heterogeneity of vegetation (or organisms "O") was integrated for prediction model by including average NDVI and NPP (2000–2022) derived from moderate resolution imaging spectro-radiometer (MODIS) satellite imagery data products MOD13Q1 V and MOD17A3HGF v6, respectively (Running and Zhao, 2015). Further, high resolution LULC data were obtained from European Space Agency (ESA) WorldCover 2021 data base (Zanaga et al., 2021). The topographic or relief (R) variations were represented by the digital elevation model (DEM) derived from NASA Shuttle Radar Topography Mission and seven primary and secondary topographic attributes derived from the DEM (Zevenbergen and Thorne, 1987; Moore et al., 1991; Gallant and Dowling, 2003; Böhner and Selige, 2006; Gruber and Peckham, 2009).

SOC stocks, SOC and $D_{b<2mm}$ data were subjected to exploratory data analysis. A Linear mixed-effects (lme) model (Pinheiro and Bates, 2002) in the nlme R package (Pinheiro et al., 2024) was used to analyze significance of field recorded land use on the spatial variability of SOC. This was followed by Tukey's post hoc test to determine statistical significance of mean differences in SOC stocks among LULC types.

1.2 Spatial modelling and mapping of SOC

The set of SOC stock observations was split to spatial prediction model training (n = 1,190, i.e. 80 % of data) and validation (n = 298) data sets. To assure the unbiasedness of the validation data set, we used conditioned Latin hypercube sampling to pick validation data by representing the distribution of SOC stock and spatial distribution of sample locations (Fig. 2).

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

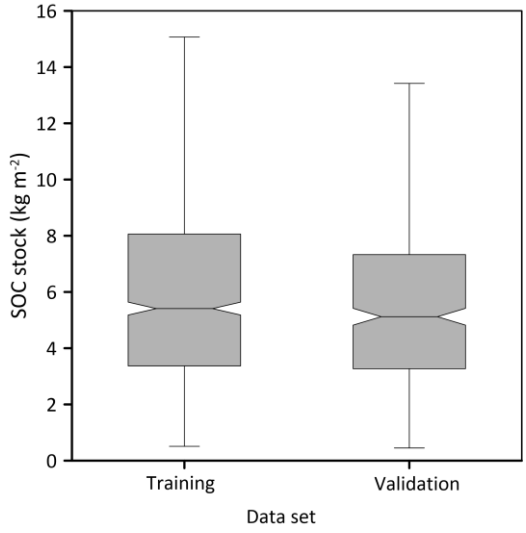


Figure 2: Boxplots showing the data distributions of SOC stocks of model training (*n* = 1190) and validation (*n* = 298) datasets. Whiskers extend to the interquartile range.

We trained four individual machine learning models, namely support vector machines, *K*-nearest neighbor (Mansuy et al., 2014), gradient boosting (Friedman, 2001), random forest (RF) (Breiman, 2001), with R package caret (Kuhn, 2008), kknn (Schliep and Hechenbichler, 2026), gbm (Ridgeway, 2026), ranger (Wright and Ziegler, 2017), respectively. 10-fold cross validation was used to tune model hyperparameters and train models. Further, to leverage the strengths of these models we used the stacking ensemble approach of machine learning by fitting a random forest model for individual model predictions. The independent validation dataset (*n* = 298) was used to compute validation indices namely, coefficient of determination (R^2), root mean square error (RMSE), mean estimation error (MEE) and Lin's concordance correlation coefficient (CCC).

$$R^2 = 1 - \frac{\sum_{i=0}^{n}(Z(\boldsymbol{X}_i)-\hat{Z}(\boldsymbol{X}_i))^2}{\sum_{i=0}^{n}(Z(\mathbf{X}_i)-m_Z)^2} \quad (6)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(Z(\mathbf{X}_i) - \hat{Z}(\mathbf{X}_i))^2} \quad (7)$$

$$\mathrm{MEE} = \sum_{i=0}^{n} Z(\mathbf{X}_i) - \hat{Z}(\mathbf{X}_i) \quad (8)$$

$$\mathrm{CCC} = \frac{2\,r\,\sigma_Z\,\sigma_{\hat{Z}}}{\sigma_Z^2 + \sigma_{\hat{Z}}^2 + (m_Z - m_{\hat{Z}})^2} \quad (9)$$

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

where, $Z(X_i)$ and $\hat{Z}(X_i)$ denote the observed and model predicted SOC stock at the i^{th} location, respectively. The terms m_Z and $m_{\hat{Z}}$ represent the means, σ_Z and $\sigma_{\hat{Z}}$ the standard deviations, σ_Z^2 and $\sigma_{\hat{Z}}^2$ the variances of observed and predicted SOC stock, respectively and r is the Pearson correlation between observed and model predictions. The random forest algorithm with hyper parameters, mtry = 3, min.node.size = 10 and num.trees = 750 (Probst et al., 2019) resulted the most accurate and precise predictions enabling us to use the same for mapping of SOC stocks in Sri Lanka at a resolution of 30 m x 30 m. To quantify uncertainty associated with predictions, we used quantile regression forest (Meinshausen, 2006), a variant of random forest using the R package Quantregforest (Meinshausen and Schiesser, 2026). We generated maps of 25th and 75th percentiles, then the difference between these percentiles, *i.e.* the map of interquartile range (IQR) of predicted SOC stocks was used for the interpretation of prediction uncertainty.

Shapley additive explanations (SHAP) analysis, was used to quantify marginal contributions of each input variable to predictions of SOC stock, thus to identify major drivers of its spatial variability. Using the iml package (Molnar, 2018) in R, SHAP analysis was performed on the training data with a hyperparameter-tuned random forest model, setting permutations to 100. Then the mean absolute SHAP values of the input variables were used to rank their importance (Lundberg et al., 2020).

1.3 Assessment of relative improvement of estimates of SOC stocks

We further used the validation dataset (n = 298) to evaluate how this study improved upon existing baseline spatial data for Sri Lanka's SOC stocks. These baselines included the most recent national SOC stocks map (SOC stocks: 2019) by Vitharana et al., 2019) and two global databases: SoilGrids v2.0, and GSOCmap v1.5.0. To compare performance, we evaluated the SOC stock values from each dataset at the 298 validation locations against field-measured values using R^2, RMSE, MEE, and CCC.

2 Results

2.1 Exploratory Data Analysis

The distributions of SOC stocks and the soil measurements used in their calculation, exhibited slight skewness (Table 1). Moreover, the majority of covariates including mean annual precipitation (MAP) and mean annual temperature (MAT) showed considerable deviation from the normal distribution. A skewed distribution of natural phenomena was anticipated on the Island, given its inherently heterogeneous climatic, topographic, and geological settings (Mapa, 2020b). SOC stocks showed a large heterogeneity (CV = 59 %) with a median stock of 5.4 kg m^{-2}. In contrast, SOC showed a greater heterogeneity (CV = 72 %) with median value of 1.3 %. Bulk density values of the dataset showed a range from 0.2 to 2.3 g cm^{-3}, reflecting soft soil materials in wetlands and compacted soils in abandoned agricultural lands, respectively. The broad range of elevation, MAP and MAT values across the sample locations (Table 1) reflects that our sampling scheme successfully captured the topographic and climatic variability across the Island.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1: Summary statistics of SOC stocks and other soil parameters and values of selected co-located environmental covariates (*n* = 1489). Db<2mm, MAP and MAT represent, bulk density of fine earth fraction, mean annual precipitation and temperature, respectively.

Statistic	SOC stock ($kg\ m^{-2}$)	SOC (%)	$D_{b<2mm}$ ($g\ cm^{-3}$)	Elevation (m)	MAP (mm)	MAT (°C)
Min.	0.5	0.1	0.2	4.0	911.0	15.7
Max.	24.8	8.7	2.3	1924.0	4809.0	28.4
Mean	6.0	1.5	1.4	161.0	2008.1	26.7
Median	5.4	1.3	1.4	62.0	1732.0	27.2
SD	3.5	1.1	0.3	282.1	813.4	1.8
CV,%	59.0	72.1	19.4	175.2	40.5	6.7
Skewness	1.1	1.7	-0.7	3.1	1.2	-2.8

We found different levels of SOC across LULC (Fig. 3). The lme model showed a significant impact of LULC types on the spatial variability of SOC stocks. Tea cultivated lands had the highest SOC stock (11.3 $kg\ m^{-2}$) followed by rubber lands (9.2 $kg\ m^{-2}$). Mean comparison revealed SOC stocks of these two land uses were significantly larger than other land uses ($p<0.05$) no significant differences of SOC stocks were observed among other lands.

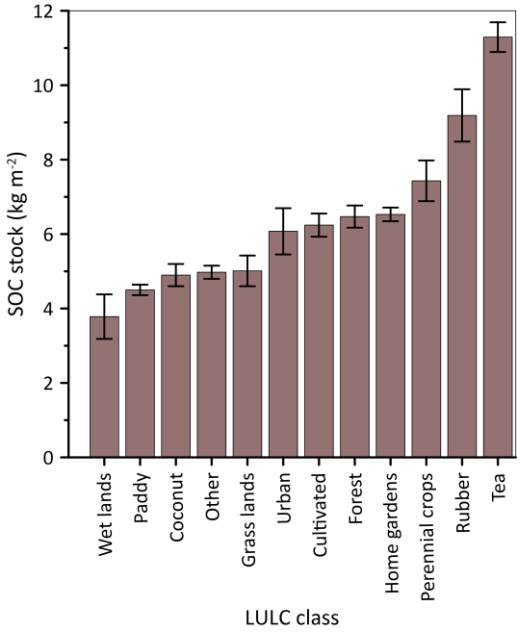


Figure 3: Comparison of average soil organic carbon (SOC) stocks across different land use and land cover (LULC) classes. Error bars represent standard errors of the mean.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.2 Random forest model fitting and controllers of SOC stocks

The trained random forest model explained 85 % of the variability in the training dataset and was characterized by low RMSE and MEE and a high CCC (Fig. 4a). Performance declined in the independent validation dataset, but the model still explained a substantial proportion of the variation in SOC stocks across the island (R^2 = 0.51; Fig. 4b). The validation RMSE was 0.97 kg m^{-2}, and the near-zero MEE indicated little overall bias. A high CCC further showed strong agreement between observed and predicted values. These patterns are illustrated in the scatter plots for the training and validation datasets (Fig. 4 a,b).

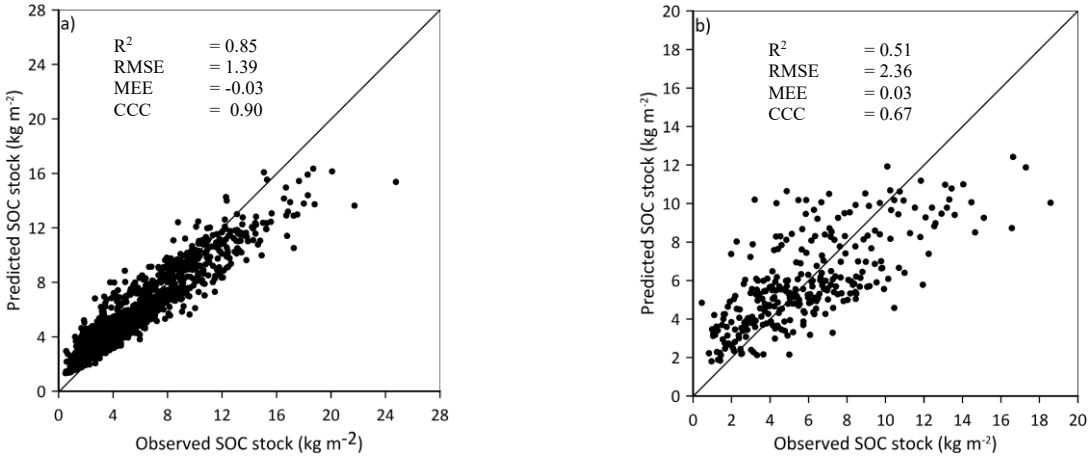


Figure 4: Scatter plots and validation indices showing relationship between random forest model predictions versus a) training data and b) validation data. Validation indices: coefficient of determination (R^2), root mean square error (RMSE), mean estimation error (MEE), and concordance correlation coefficient (CCC).

Fig. 5 illustrates the 15 most influential controllers of SOC stocks, as identified through Shapley analysis. Here, mean absolute SHAP values quantify the marginal contributions of each predictor variable to the predictions of stocks of the training data set. Within the SCORPAN factors included into the random forest model, climate, spatial position and relief have played critical roles in determining the heterogeneity of SOC stocks across the Island. Among these controllers, climate emerged as the dominant influence, with the aridity index identified as the most significant factor (absolute SHAP = 0.21). This is followed by almost equal contributions from MAP, precipitation of warmest quarter and potential evapotranspiration (0.15). Additionally, solar radiation (0.14) also significantly contributed to the model, with both spatial position (Y coordinate) and temperature making nearly equal contributions (0.11). In contrast, elevation demonstrated a relatively lower impact (0.08) on the variability of SOC stocks.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

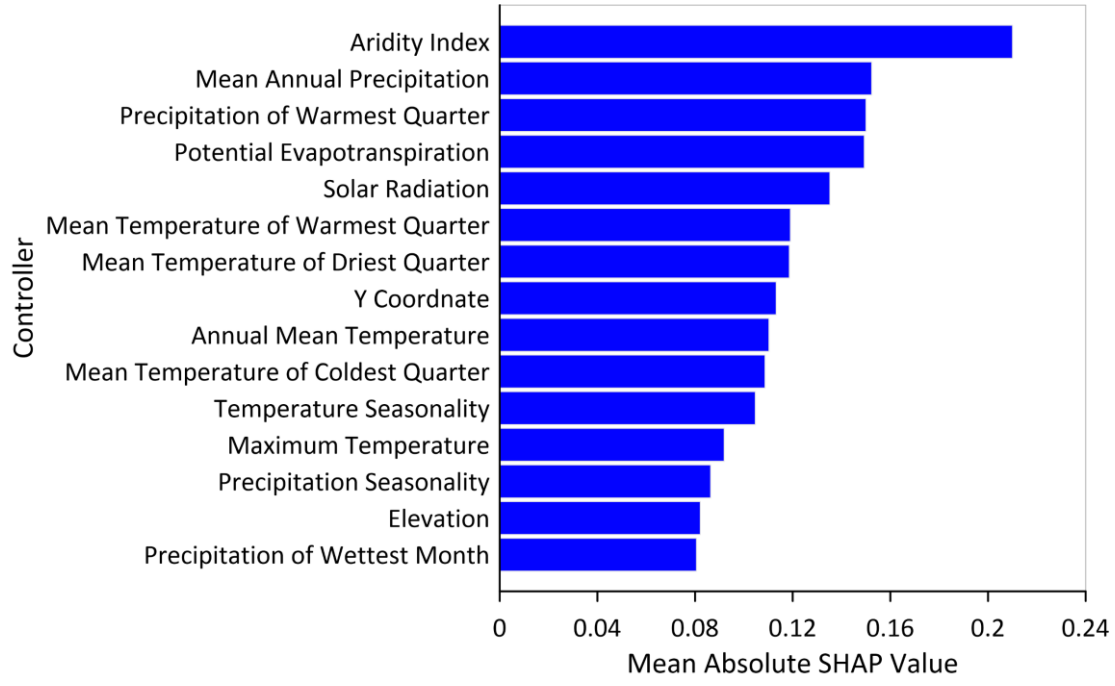


Figure 5: Mean absolute Shapley additive explanations (SHAP) values of the 15 most influential controllers of SOC stocks.

265

2.2 Spatial heterogeneity of SOC stocks

Predicted SOC stocks (Fig. 6a) exhibited contiguous spatial patterns across the Island. The highest SOC stocks, ranging from 12-16.4 kg m^{-2}, were predicted in the central region corresponding to the central highlands (elevation of 1800-2400 m). A substantial area extending from the center of the country towards the southwest, consisted of soils retaining comparatively
270 larger SOC stocks, between 6-12 kg m^{-2}. This area corresponds to the wet and intermediate zones of the country, which receive a mean annual rainfall exceeding 1,750 mm. In contrast, the remaining areas of the study, representing the low-lying (elevation 0-270 m) dry zone of the country, exhibited relatively lower SOC stocks, less than 6 kg m^{-2}. The northern, northwestern, eastern and southeastern regions, had soils with lowest SOC stock. Overall, the surface soils of this tropical island represent a substantial carbon reservoir, containing a total national SOC stock of 382.4 Tg. The IQR map (Fig. 6b) illustrates uncertainty
275 of prediction by quantifying the range of predicted SOC stocks across different spatial locations. Wider IQRs show greater prediction uncertainty. Compared to the range of observed SOC stocks, the IQR values determined across the spatial extent were notably lower. However, the analysis revealed relatively higher uncertainty associated with the prediction of larger SOC stocks as compared to smaller values. Notably, the predictions made for the forest areas that could not be sampled due to accessibility restrictions (Fig. 1) were associated with small IQRs. However, this observation does not necessarily suggest

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

accurate estimates of SOC stocks in highly protected forest reserves, as our sampling scheme did not adequately cover forest land cover. Overall, the surface soils of this tropical island represent a substantial carbon reservoir, containing a total national SOC stock of 382.4 Tg. Furthermore, when prediction uncertainties are accounted for (Fig. 6b), the total SOC stock ranges from 269.88 to 472.25 Tg.

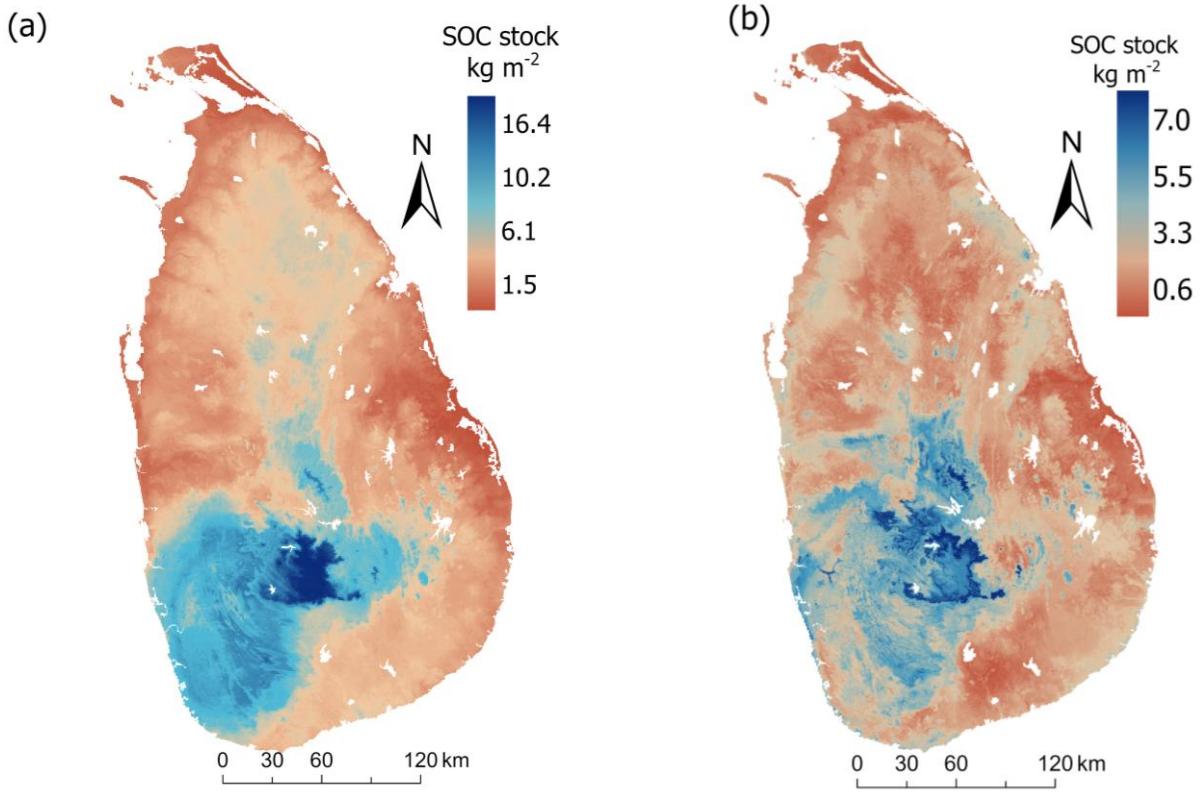


Figure 6: (a) predicted SOC stocks across the study area and, (b) associated prediction uncertainty, represented by the interquartile range.

2.2 Improvement of estimates of SOC stocks by the updated data base

The accuracy of the existing SOC stock baselines, evaluated using the validation dataset, is presented in Fig. 7a-c. The existing SOC stock datasets poorly explained the variability of the actual SOC stocks. Overall, the R^2 values indicate that all evaluated datasets exhibited poor agreement with the validation data. GSOCmap v1.5.0 performed the poorest, followed by SOC stocks: 2019 and SoilGrids v2.0. Two of these baselines yielded negative R^2 values indicating substantial residual error where spatial models performed worse than applying the mean of the validation dataset. Furthermore, SoilGrid v2.0 produced a near-zero R^2, suggesting it approximately predicts the dataset average and fails to capture the actual spatial heterogeneity across the Island.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

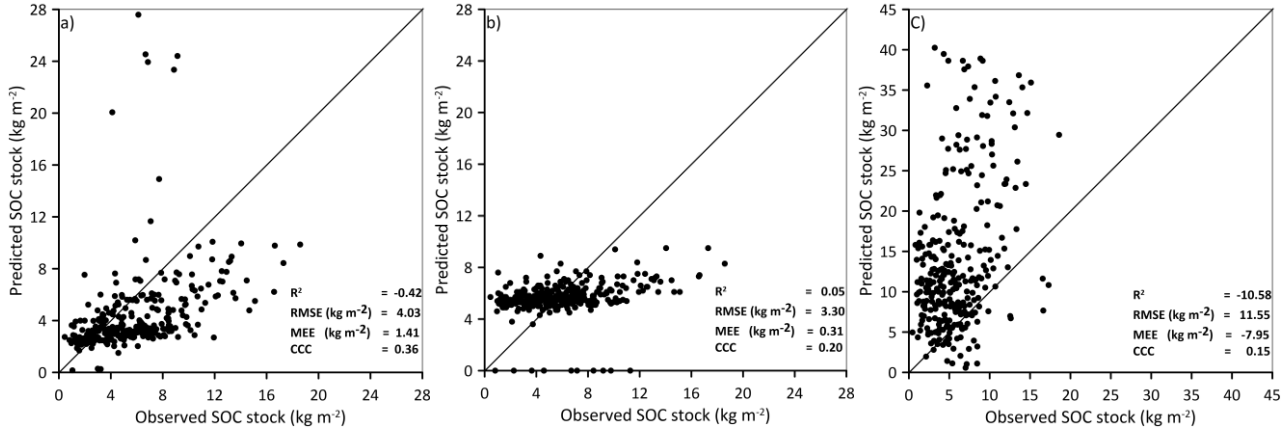


Figure 7: Scatter plots and validation indices illustrating the accuracy of existing baseline SOC stock datasets against the validation data set. (a) SOC stocks: 2019 (Vitharana et al., 2019), (b) SoilGrids v2.0, and the (c) Global Soil Organic Carbon Map (GSOCmap v1.5.0). Validation indices: coefficient of determination (R^2), root mean square error (RMSE), mean estimation error (MEE), and concordance correlation coefficient (CCC).

The low accuracy of these datasets is further evidenced by their high RMSE values. Notably, the RMSE of GSOCmap v1.5 dataset far exceeded the mean value of the validation data, highlighting highly erratic SOC stock estimates. While SoilGrids v2.0 showed comparatively lower RMSE, its error remained higher than the updated dataset developed in this study. The MEE values revealed a strong positive bias (overestimation) by GSOCmap v1.5, whereas SOC stocks: 2019 and SoilGrids v2.0 exhibited relatively lower bias. The CCC values substantiated the overall poor accuracy and precision of existing datasets, although SOC stocks: 2019 demonstrated comparatively better overall performance. This is further supported by the scatter plots, which show the observed and predicted values for the SOC stocks: 2019 dataset aligning more closely along the 1:1 line. In contrast, the updated baseline SOC stock dataset generated through denser spatial sampling and machine learning predictions driven by diverse environmental covariates clearly outperformed all existing baseline datasets (Figure 4b). This establishes it as a robust spatial inventory for Sri Lanka's SOC stocks. Nevertheless, as validation indices indicated, fine-scale localized variations in SOC stocks remain difficult to capture.

2 Discussion

The relatively low median SOC stock observed across Sri Lanka is broadly consistent with expectations for tropical environments, where warm temperatures promote rapid organic matter decomposition and, in humid regions, intense rainfall can enhance dissolved organic carbon transport and leaching. Compared with soils in cool and wet climates, tropical soils therefore often accumulate smaller SOC stocks despite supporting high rates of primary productivity (Chiti et al., 2010). In the

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

drier tropics, SOC storage can be further constrained by water limitation, which suppresses vegetation productivity and reduces the quantity of plant-derived carbon entering the soil (Naorem et al., 2022). Consistent with these mechanisms, Don et al., (2010) identified temperature and precipitation as major climatic controls on SOC changes across tropical regions. Sri Lanka spans pronounced gradients in rainfall, temperature, elevation, vegetation, and land use over a relatively small geographic area (Galagamage et al., 2026). The combination of these environmental gradients likely explains why national SOC stocks are relatively modest on average but highly heterogeneous spatially. Land use further modifies this climatic control on SOC storage. A global meta-analysis by Don et al. (2010) showed that tropical forest soils generally contain substantial SOC stocks, with values of approximately 8–10.5 kg C m^{-2}, and that conversion of native vegetation can reduce SOC stocks by approximately 32 % in dry tropical regions and 24 % in wet tropical regions. In contrast to the general expectation that relatively undisturbed forests should contain the greatest SOC stocks, our analysis identified particularly high SOC stocks under tea and rubber plantations. Previous studies from Sri Lanka provide some support for this pattern. Hemamali et al. (2020), for example, reported SOC stocks of approximately 6.5 kg C m^{-2} in subsurface soils under tea, comparable with approximately 6.7 kg C m^{-2} under tropical forests. Similarly, Nayanarangani et al. (2025) found comparable SOC levels in tea plantations and adjacent tropical forests, while vegetable-growing soils contained substantially lower SOC. Nevertheless, high SOC stocks under tea should not be interpreted as evidence that plantation establishment generally increases SOC relative to native forests. In montane central Taiwan, Chen et al. (2016) reported lower SOC concentrations in tea plantations than in adjacent forest soils, illustrating the importance of local climate, soil properties, management history, and plantation age. Our total SOC stock assessment is constrained by inadequate sampling coverage in forest reserves. Consequently, refining future estimates will require expanding sampling efforts to adequately cover protected forest reserves in Sri Lanka. Our spatial analysis further demonstrates that the geographic distribution of SOC across Sri Lanka arises from strongly non-linear climatic and topographic controls. Water availability emerged as one of the most important predictors, as indicated by the strong influence of the aridity index and WorldClim bioclimatic variables associated with precipitation and potential evapotranspiration. This result is consistent with observations from other tropical ecosystems showing that hydroclimatic conditions strongly regulate both vegetation productivity and soil carbon cycling (Esquivel-Muelbert et al., 2016). However, SOC patterns did not simply follow the national precipitation gradient. In particular, regions receiving the greatest rainfall did not invariably contain the largest SOC stocks, indicating that water availability interacts with temperature, elevation, radiation, vegetation, soil properties, and land use to determine the net balance between carbon inputs and decomposition. The 220-km transect across contrasting climatic and topographic regions provides additional insight into these interacting controls (Fig. 8a). Low-elevation areas in the southwestern and northeastern portions of the transect exhibited markedly different SOC stocks, with considerably greater SOC storage in the southwest, while the largest stocks overall occurred within the central highlands (Fig. 8b). The relatively high SOC stocks in southwestern Sri Lanka coincided with greater moisture availability, as represented by a higher aridity index and greater precipitation (Fig. 8c,d). The southwestern region receives substantial rainfall associated with the southwest monsoon as well as additional precipitation during inter-monsoonal and northeast-monsoonal periods (Mapa, 2020b), resulting in comparatively favorable year-round moisture conditions. Greater water availability can enhance

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

vegetation cover, productivity, and biomass inputs to soils, thereby increasing carbon inputs that contribute to SOC accumulation (Doetterl et al., 2015). Importantly, however, abundant moisture may simultaneously stimulate microbial decomposition. The high SOC stocks observed in this region therefore suggest that, under these conditions, enhanced carbon inputs and other stabilizing factors are sufficient to offset relatively rapid decomposition.

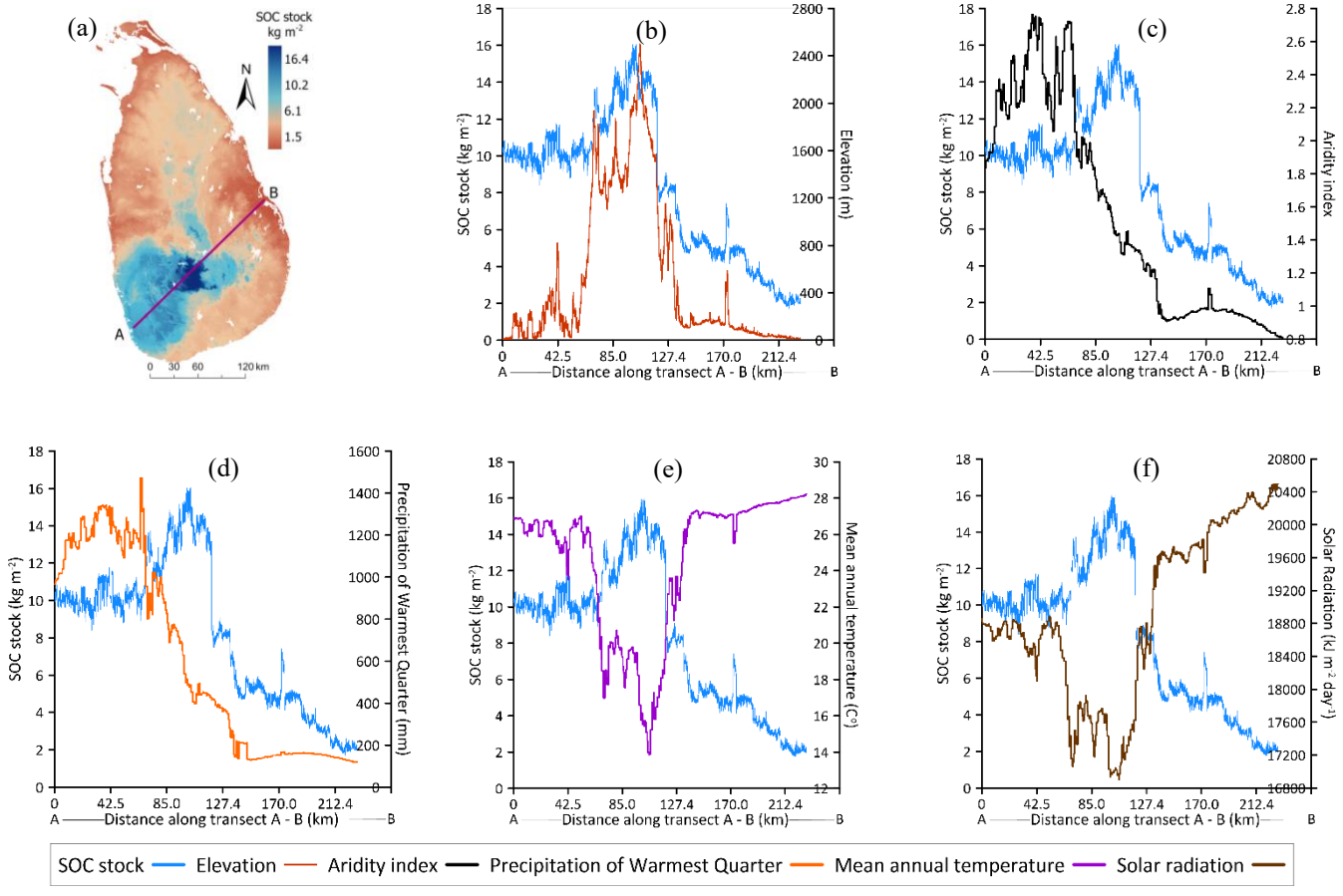


Figure 8: Location of the A–B transect across the SOC stock gradient in Sri Lanka, alongside corresponding variations in b) elevation, c) aridity index, d) precipitation of the warmest quarter, e) mean annual temperature, and f) solar radiation.

A different mechanism appears to dominate SOC accumulation in the central highlands. Here, high SOC stocks coincide with substantially lower atmospheric temperatures associated with increasing elevation (Fig. 8e). The negative relationship between temperature and SOC storage is well established because lower temperatures reduce microbial metabolic activity and decomposition rates, thereby increasing the residence time of organic matter in soil. Cooler conditions may also favor the persistence of relatively stable SOC fractions, including mineral-associated organic carbon (Ren et al., 2024). Interestingly,

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

the central highlands receive lower solar radiation than parts of the lowlands (Fig. 8f), which could limit photosynthesis and biomass production under some conditions. Nevertheless, SOC stocks remain high, suggesting that temperature-induced suppression of decomposition can outweigh potentially lower carbon inputs. This distinction is important because it demonstrates that similar SOC stocks can arise from different combinations of carbon inputs and losses across the landscape. The northeastern portion of the transect illustrates the opposite hydroclimatic constraint. Although this region receives relatively high solar radiation (Fig. 8f), low water availability associated with greater aridity restricts vegetation productivity and therefore limits organic carbon inputs to the soil (Fig. 8 c and d). High temperatures may simultaneously accelerate decomposition of the organic matter that is present. Together, these conditions provide a mechanistic explanation for the relatively low SOC stocks observed in northeastern Sri Lanka.

The transect analysis therefore reveals two major climatic pathways promoting SOC accumulation across the island: high moisture availability supports greater carbon inputs in the humid southwest, whereas low temperature restricts carbon losses in the central highlands. In drier and warmer regions, both reduced plant inputs and faster turnover constrain SOC accumulation. These contrasting mechanisms indicate that future SOC responses to climate change are unlikely to be spatially uniform. Warming may disproportionately affect high-elevation SOC stocks by weakening the temperature constraint on decomposition, whereas changes in precipitation and atmospheric water demand may exert stronger controls in lowland regions where SOC storage is primarily moisture limited. The importance of these non-linear environmental interactions is also evident from our benchmarking of the updated SOC dataset against existing global and regional products. Among the evaluated products, GSOCmap v1.5.0 showed the poorest agreement with the observations. GSOCmap was developed through a country-driven effort to produce a global SOC inventory at 1-km resolution, with individual countries applying methodologies according to available observations and technical capacity. For Sri Lanka, the product was generated from only 233 soil observations using multiple linear regression and without independent validation (FAO and ITPS, 2020). Its relatively poor performance in our analysis may therefore reflect both limited spatial sampling and the restricted ability of a linear modeling framework to represent the strongly non-linear relationships among climate, topography, land use, and SOC identified in the present study. The 2019 national SOC product substantially outperformed GSOCmap v1.5.0 and showed the best overall balance between agreement and precision among the three reference databases, as indicated by its CCC. This product was generated by integrating legacy information from 122 soil profiles with environmental covariates using geographically weighted regression. Its comparatively strong performance illustrates that an appropriately selected digital soil mapping framework can extract considerable spatial information even from a relatively limited number of observations, particularly when spatial non-stationarity is explicitly represented. Nevertheless, its lower predictive performance relative to our updated map also highlights the constraints associated with legacy datasets, especially where observations are sparse, unevenly distributed, or were collected using different sampling protocols. The comparison with SoilGrids provides additional evidence for the importance of both modeling approach and data volume. Vitharana et al. (2019) previously reported that the 2019 Sri Lankan SOC product outperformed SoilGrids250m v0.5.8 (Hengl et al., 2017). In our analysis, however, the updated SoilGrids250m v2.0 product (Poggio et al., 2021) exhibited performance comparable with the 2019 national SOC dataset,

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

although the latter retained a somewhat better balance between accuracy and precision according to CCC. The improved performance of the newer SoilGrids product likely reflects major methodological and data improvements, including implementation of a quantile random forest framework, incorporation of approximately 90,000 additional WoSIS soil profiles, and use of more than 400 environmental covariates (Batjes et al., 2020; Poggio et al., 2021). These advances allow a global model to learn transferable relationships between SOC and environmental conditions that can partly compensate for sparse observations within individual countries. This point is particularly relevant for Sri Lanka because the contemporary WoSIS database contains very limited local representation. Despite the small number of Sri Lankan profiles available to the global training dataset, SoilGrids was able to reproduce a substantial component of the observed national SOC heterogeneity. This performance demonstrates the value of large, globally harmonized soil databases combined with flexible machine-learning approaches, but it also exposes an important limitation: predictions in data-sparse regions necessarily depend heavily on relationships learned elsewhere. Such relationships may not fully represent locally specific soil-forming processes, management effects, parent materials, or climatic interactions. Consequently, global products remain valuable for broad-scale assessments, but locally calibrated SOC maps based on denser and more representative observations are preferable when national-scale carbon inventories, land-management decisions, or climate-mitigation assessments require greater spatial accuracy.

Collectively, our results demonstrate that SOC storage across Sri Lanka cannot be explained by a single climatic, topographic, or land-use factor. Instead, spatial SOC patterns reflect the balance between plant-derived carbon inputs and carbon losses, with the dominant control changing geographically. Moisture availability appears to exert a particularly strong influence in lowland tropical environments, whereas temperature becomes increasingly important in the central highlands. Land-use effects are superimposed on these environmental gradients and therefore need to be interpreted within their climatic and edaphic context. The superior performance of the updated SOC map further demonstrates the value of combining more representative soil observations with modeling approaches capable of resolving non-linear environmental controls. These findings provide an improved baseline for evaluating future changes in Sri Lankan SOC stocks and suggest that assessments of climate-change vulnerability should explicitly account for spatial differences in whether SOC storage is predominantly controlled by moisture, temperature, or their interaction.

2 Conclusion

This study demonstrates that high-density spatial sampling, coupled with machine-learning-based digital soil mapping, can substantially improve national-scale estimates of SOC stocks compared with products derived from sparse legacy observations or global mapping initiatives. The expanded sampling network not only improved predictive accuracy but also enabled a more detailed characterization of the environmental controls governing SOC distribution across Sri Lanka. Climate emerged as the dominant control on SOC spatial heterogeneity, interacting strongly with topography and highlighting the spatially variable nature of SOC responses to future climate change. In the central highlands, lower temperatures favor SOC accumulation by

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

limiting decomposition, whereas in the lowlands, water availability and solar radiation associated with monsoonal rainfall patterns exert stronger controls on carbon inputs and storage. The Random Forest framework effectively captured these complex, non-linear relationships and provided robust predictions across unsampled areas. Although land use and land cover were less influential than climatic variables at the national scale, tea and rubber plantations supported relatively high SOC stocks, demonstrating the potential importance of perennial cropping systems for maintaining soil carbon. Benchmarking further showed that the updated national SOC map substantially improved upon existing baselines. While methodological advances have enhanced the performance of SoilGrids v2.0, GSOCmap v1.5.0 showed relatively poor agreement with the new observations, indicating that its application for local-scale assessment and decision-making should be approached cautiously. Overall, our findings emphasize that reliable national SOC assessments require sufficiently dense and spatially representative observations combined with modeling approaches capable of resolving non-linear interactions among climate, topography, soils, and land use. The resulting high-resolution SOC dataset provides an improved baseline for national soil carbon inventories, identification of SOC-rich and vulnerable regions, and assessment of future climate impacts. More broadly, this study demonstrates the value of country-specific soil monitoring and tailored digital soil mapping frameworks for supporting evidence-based land management and soil carbon conservation in environmentally heterogeneous tropical regions.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix A

Table A1: Reclassified land use land cover types.

Reclassed LULC name	Field observations	Number of samples
Paddy	Paddy	346
Tea	Tea	93
Rubber	Rubber	23
Coconut	Coconut	114
Perennial crops	Banana, Areca nut, Cinnamon, Cashew, Mango, etc.	39
Home gardens	Home gardens	348
Other	Abandoned, scrubland, unutilized and bare lands	211
Cultivated	Pulses, Vegetables, root crops, etc	101
Forest	Forest	126
Wet lands	Wet lands, Mangroves, flooded areas	23
Grass lands	Grass lands	46
Urban	Urban	18

Table A2: Potential intrinsic and environmental controllers of SOC used for modelling the spatial heterogeneity of SOC stocks.

Environmental Predictor	Resolution (m)	SCORPAN Variable	Data Source/ reference
Soil Order	30 m*	Soil	Aggregation of map units of soil series into soil orders by assigning the order of the dominant soil series within each map unit.
Soil Suborder	30 m*	Soil	Aggregation of map units of soil series into soil suborders by assigning the suborder of the dominant soil series within each map unit.
Soil Great Group	30 m*	Soil	Aggregation of map units of soil series into soil great group by assigning the great group of the dominant soil series within each map unit.
Soil Subgroup	30 m*	Soil	Aggregation of map units of soil series into soil sub group by assigning the sub group of the dominant soil series within each map unit
Soil Series	30 m*	Soil	Aggregation of map units of associations, complexes or consociations of soil series into soil series by assigning the dominant soil series.
BIO1 - Annual Mean Temperature (˚C)	1000 m	Climate	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.
BIO2 - Mean Diurnal Range (˚C)	1000 m	Climate	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.
BIO3 - Isothermality (%)	1000 m	Climate	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.
BIO4 - Temperature Seasonality	1000 m	Climate	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.
BIO5 - Max Temperature of Warmest Month (˚C)	1000 m	Climate	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.
BIO6 - Min Temperature of Coldest Month (˚C)	1000 m	Climate	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.
BIO7 - Temperature Annual Range (˚C)	1000 m	Climate	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.
BIO8 - Mean Temperature of Wettest Quarter (˚C)	1000 m	Climate	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.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

BIO9 - Mean Temperature of Driest Quarter (°C)	1000 m	Climate	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.
BIO10 - Mean Temperature of Warmest Quarter (°C)	1000 m	Climate	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.
BIO11 - Mean Temperature of Coldest Quarter (°C)	1000 m	Climate	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.
BIO12 - Annual Precipitation (°C)	1000 m	Climate	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.
BIO13 - Precipitation of Wettest Month	1000 m	Climate	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.
BIO14 - Precipitation of Driest Month (°C)	1000 m	Climate	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.
BIO15 - Precipitation Seasonality (%)	1000 m	Climate	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.
BIO16 - Precipitation of Wettest Quarter (mm)	1000 m	Climate	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.
BIO17 - Precipitation of Driest Quarter (mm)	1000 m	Climate	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.
BIO18 - Precipitation of Warmest Quarter (mm)	1000 m	Climate	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.
BIO19 - Precipitation of Coldest Quarter (mm)	1000 m	Climate	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.
Solar Radiation (kJ m^{-2} day^{-1})	1000 m	Climate	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.
Aridity Index	1,000 m	Climate	Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. Sci. Data 9. https://doi.org/10.1038/s41597-022-01493-1. Available online from the CGIAR-CSI GeoPortal at: https://cgiarcsi.community/2019/01/24/globalaridity-index-and-potential-evapotranspiration-climate-database-v3/ (1970-2000)
Potential evapotranspiration (mm day^{-1})	1,000 m	Climate	Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. Sci. Data 9. https://doi.org/10.1038/s41597-022-01493-1. Available online from the CGIAR-CSI GeoPortal at: https://cgiarcsi.community/2019/01/24/globalaridity-index-and-potential-evapotranspiration-climate-database-v3/ (1970-2000)
NDVI	250 m	Vegetation /Organisms	We used the MOD13Q1 Version 6 dataset (250 m, 16-day composite NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra satellite, accessed *via* Google Earth Engine. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.006
NPP (kgC m^{-2} $year^{-1}$)	500 m	Vegetation /Organisms	We used the MOD17A3HGF Version 6 dataset (250 m, 16-day composite NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra satellite, accessed *via* Google Earth Engine. Running, S. W., & Zhao, M., 2015. MOD17A3HGF MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. https://doi.org/10.5067/MODIS/MOD17A3HGF.006
Land Use and Land Cover	10 m	Vegetation /Organisms	We used European Space Agency (ESA) WorldCover 2021, accessed via Google Earth Engine. Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., … Arino, O. (2021). ESA WorldCover 10 m 2020 v100 (Version v100) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5571936
Elevation (m)	30 m	Topography	Digital elevation model (DEM) developed from NASA Shuttle Radar Topography Mission (SRTM) data: Shuttle Radar Topography Mission (SRTM) Global. Distributed by Open Topography. https://doi.org/10.5069/G9445JDF. Accessed: 2024-09-16
Topographic wetness index	30 m	Topography	Topographic wetness index map derived from DEM using SAGA: Boehner, J., Selige, T.,2006. Spatial Prediction of Soil Attributes Using Terrain Analysis and Climate Regionalisation, in: Boehner, J., McCloy, K.R., Strobl, J. (Eds), SAGA - Analysis and Modelling Applications, Goettinger Geographische Abhandlungen, Vol.115, pp.13-27.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Stream Power Index		Topography	Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrogical, geomorphological, and biological applications. Hydrol. Process., 5, 3-30.
Slope (%)	30 m	Topography	Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Earth Surface Process and Landforms 12, 47–56.
Multiresolution index of valley bottom flatness	30 m	Topography	Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resource Research 39, 1347–1359.
Multiresolution index of the ridge top flatness	30 m	Topography	Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resource Research 39, 1347–1359.
Slope length factor	30 m	Topography	Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrogical, geomorphological, and biological applications. Hydrol. Process., 5, 3-30.
Flow accumulation (m^2)	30 m	Topography	Gruber, S., Peckham, S., 2009. Land-Surface Parameters and Objects in Hydrology, in: Hengl, T., Reuter, H.I. (Eds.), Developments in Soil Science: Geomorphometry Concepts, Software, Applications. Elsevier, Amsterdam, The Netherlands, pp. 171–194.

*Denotes the resolution of raster maps derived from corresponding choropleth maps.

Code and data availability

The data sources supporting the findings of this study are available within the article and Appendix A. Additional data and custom code are available from the corresponding author upon reasonable request.

Author contributions

UV: investigation, Conceptualization, Writing (original draft), formal analysis, funding acquisition. UM: Conceptualization, writing (review and editing), Methodology, funding acquisition SG and CJ: Formal analysis, Software, Writing (review and editing).

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Copernicus Publications adds a standard disclaimer: “Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.”
Please feel free to add disclaimer text at your choice, if applicable.

Acknowledgements

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

We gratefully acknowledge the Department of Soil Science, Faculty of Agriculture, University of Peradeniya, for providing laboratory facilities. We also express our sincere appreciation to the Natural Resources Management Division of the Ministry of Environment for coordinating the funding that made this work possible.

Financial support

This study was supported by a grant from the Global Environment Facility (GEF) support to UNCCD 2018 National Reporting Process - Umbrella II project. Contributions of U. Mishra were also supported by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation Science Focus Area (RUBISCO SFA), which is sponsored by the Biological and Environmental Research (BER) program in the US Department of Energy Office of Science. Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc. for the US Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

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

Böhner, J. and Selige, T.: Spatial prediction of soil attributes using terrain analysis and climate regionalisation, in: SAGA – Analysis and Modelling Applications, edited by: Böhner, J., McCloy, K. R., and Strobl, J., Goettinger Geographische Abhandlungen, vol. 115, pp. 13–27, 2006.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Chiti, T., Certini, G., Grieco, E., and Valentini, R.: The role of soil in storing carbon in tropical rainforests: the case of Ankasa Park, Ghana, Plant Soil, 331, 453–461, https://doi.org/10.1007/s11104-009-0265-x, 2010.

Dai, Y., Shangguan, W., Wei, N., Xin, Q., Yuan, H., Zhang, S., Liu, S., Lu, X., Wang, D., and Yan, F.: A review of the global soil property maps for Earth system models, SOIL, 5, 137–158, https://doi.org/10.5194/soil-5-137-2019, 2019.

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., and Boeckx, P.: Soil carbon storage controlled by interactions between geochemistry and climate, Nat. Geosci., 8, 780–783, https://doi.org/10.1038/ngeo2516, 2015.

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Don, A., Schumacher, J., and Freibauer, A.: Impact of tropical land use change on soil organic carbon stocks – a meta-analysis, Glob. Change Biol., 17, 1658–1670, https://doi.org/10.1111/j.1365-2486.2010.02336.x, 2010.

Esquivel-Muelbert, A., Baker, T. R., Dexter, K. G., Lewis, S. L., ter Steege, H., López-González, G., Monteagudo Mendoza, A., Brienen, R., Feldpausch, T. R., Pitman, N. C. A., and et al.: Seasonal drought limits tree species across the Neotropics, Ecography, 40, 618–629, https://doi.org/10.1111/ecog.01904, 2016.

FAO and ITPS: Global Soil Organic Carbon Map (GSOCmap) Version 1.5: Technical Report [data set], FAO, Rome, Italy, https://doi.org/10.4060/ca7597en, 2020.

Fick, S. E. and Hijmans, R. J.: 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, 2017.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx, I. T., and et al.: Global carbon budget 2025, Earth Syst. Sci. Data, 18, 3211–3288, https://doi.org/10.5194/essd-18-3211-2026, 2026.

Friedman, J. H.: Greedy function approximation: A gradient boosting machine, Ann. Stat., 29, 1189–1232, https://doi.org/10.1214/aos/1013203451, 2001.

Galagamage, P., Payagala, C., Pathirana, U. P. G., Kug, J.-S., Atapaththu, K. S. S., Geng, X., Wu, Y.-K., Wang, D., Premarathne, U., Ranasinghe, P. N., and De Silva, P. H. C.: Rainfall dynamics in Sri Lanka: a review of present variability, trends and responses to climate change, https://doi.org/10.21203/rs.3.rs-9329627/v1, 2026.

Gallant, J. C. and Dowling, T. I.: A multiresolution index of valley bottom flatness for mapping depositional areas, Water Resour. Res., 39, 1347, https://doi.org/10.1029/2002WR001426, 2003.

Gruber, S. and Peckham, S.: Land-surface parameters and objects in hydrology, in: Geomorphometry: Concepts, Software, Applications, edited by: Hengl, T. and Reuter, H. I., Developments in Soil Science, vol. 33, Elsevier, Amsterdam, the Netherlands, pp. 171–194, https://doi.org/10.1016/S0166-2481(08)00007-X, 2009.

Hastie, T., Tibshirani, R., and Friedman, J.: Model assessment and selection, in: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, NY, USA, pp. 193–224, https://doi.org/10.1007/978-0-387-84858-7, 2009.

He, T., Ding, W., Cheng, X., Cai, Y., Zhang, Y., Xia, H., Wang, X., Zhang, J., Zhang, K., and Zhang, Q.: Meta-analysis shows the impacts of ecological restoration on greenhouse gas emissions, Nat. Commun., 15, 2415, https://doi.org/10.1038/s41467-024-46991-5, 2024.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: global gridded soil information based on machine learning, PLOS ONE, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Kopittke, P. M., Berhe, A. A., Carrillo, Y., Cavagnaro, T. R., Chen, D., Chen, Q., Román Dobarco, M., Dijkstra, F. A., Field, D. J., He, J., and et al.: Ensuring planetary survival: the centrality of organic carbon in balancing the multifunctional

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

nature of soils, Crit. Rev. Environ. Sci. Technol., 52, 4308–4324, https://doi.org/10.1080/10643389.2021.2024484, 2022.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15, 259–263, https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Kuhn, M.: Building predictive models in R using the caret package, J. Stat. Softw., 28, 1–26, https://doi.org/10.18637/jss.v028.i05, 2008.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S. I.: From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

Mapa, R. B.: Soils of the Northern Region of Sri Lanka: Morphology, Characterization & Classification. Special Publication No. 9. Soil Science Society of Sri Lanka, Nethwin Printers, Sri Lanka, 2020a.

Mapa, R. B., Somasiri, S., and Dassanayake, A. R. (Eds.): Soils of the dry zone of Sri Lanka: morphology, characterization and classification, Special Publication no. 7, Soil Science Society of Sri Lanka, Sarvodaya Wishva Lekha, Sri Lanka, 2010.

Mapa, R. B. (Ed.): The Soils of Sri Lanka, World Soils Book Series, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-030-44144-9, 2020b.

McBratney, A. B., Mendonça Santos, M. L., and Minasny, B.: On digital soil mapping, Geoderma, 117, 3–52, https://doi.org/10.1016/S0016-7061(03)00223-4, 2003.

Meinshausen, N.: Quantile regression forests, J. Mach. Learn. Res., 7, 983–999, https://www.jmlr.org/papers/v7/meinshausen06a.html, 2006.

Meinshausen, N. and Michel, L.: quantregForest: Quantile Regression Forests, R package version 1.4-0, available at: https://CRAN.R-project.org/package=quantregForest, https://doi.org/10.32614/CRAN.package.quantregForest, 2026.

Minasny, B. and McBratney, A. B.: A conditioned Latin hypercube method for sampling in the presence of ancillary information, Comput. Geosci., 32, 1378–1388, https://doi.org/10.1016/j.cageo.2005.12.009, 2006.

Molnar, C.: iml: An R package for interpretable machine learning, J. Open Source Softw., 3, 786, https://doi.org/10.21105/joss.00786, 2018.

Moore, I. D., Grayson, R. B., and Ladson, A. R.: Digital terrain modelling: a review of hydrological, geomorphological, and biological applications, Hydrol. Process., 5, 3–30, https://doi.org/10.1002/hyp.3360050103, 1991.

Nachtergaele, F., Verelst, L., Batjes, N. H., Dijkshoorn, J. A., van Engelen, V. W. P., Fischer, G., Jones, A. C., Montanarella, L., Petri, M., Prieler, S., Teixeira, E. N., Wilberg, D. E., and Shi, X.: Harmonized World Soil Database version 2.0 [data set], FAO and IIASA, Rome, Italy and Laxenburg, Austria, https://doi.org/10.4060/cc3823en, 2023.

Naorem, A., Jayaraman, S., Dalal, R. C., Patra, A., Rao, C. S., and Lal, R.: Soil inorganic carbon as a potential sink in carbon storage in dryland soils: a review, Agriculture, 12, 1256, https://doi.org/10.3390/agriculture12081256, 2022.

Nelson, D. W. and Sommers, L. E.: Total carbon, organic carbon, and organic matter, in: Methods of Soil Analysis: Part 3, Chemical Methods, Soil Science Society of America Book Series no. 5, edited by: Sparks, D. L., Page, A. L., Helmke,

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

P. A., Loeppert, R. H., Soltanpour, P. N., Tabatabai, M. A., Johnston, C. T., and Sumner, M. E., Soil Science Society of America, American Society of Agronomy, Madison, WI, USA, pp. 961–1010, https://doi.org/10.2136/sssabookser5.3.c34, 1996.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: 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, 2021.

Pribyl, D. W.: A critical review of the conventional SOC to SOM conversion factor, Geoderma, 156, 75–83, https://doi.org/10.1016/j.geoderma.2010.02.003, 2010.

Probst, P., Wright, M. N., and Boulesteix, A.-L.: Hyperparameters and tuning strategies for random forest, WIREs Data Mining Knowl. Discov., 9, e1301, https://doi.org/10.1002/widm.1301, 2019.

Ren, S., Wang, T., Guenet, B., Liu, D., Cao, Y., Ding, J., Smith, P., and Piao, S.: Projected soil carbon loss with warming in constrained Earth system models, Nat. Commun., 15, 123, https://doi.org/10.1038/s41467-023-44433-2, 2024.

Ridgeway, G. and GBM Developers: gbm: Generalized Boosted Regression Models, R package version 2.3.1, https://doi.org/10.32614/CRAN.package.gbm, 2026.

Schliep, K. and Hechenbichler, K.: kknn: Weighted k-Nearest Neighbors, R package version 1.4.1, https://doi.org/10.32614/CRAN.package.kknn, 2026.

Soil Survey Staff: Keys to Soil Taxonomy, 12th edn., USDA-NRCS, Washington, DC, USA, 360 pp., available at: https://www.nrcs.usda.gov/resources/guides-and-instructions/keys-to-soil-taxonomy, 2014.

Running, S. W. and Zhao, M.: MOD17A3HGF MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V006 [data set], NASA EOSDIS LP DAAC, https://doi.org/10.5067/MODIS/MOD17A3HGF.006, 2015.

UNCCD: PRAIS4 reporting manual and glossary, United Nations Convention to Combat Desertification (UNCCD) Secretariat, Bonn, Germany, https://prais4-reporting-manual.unccd.int/, 2022.

Vitharana, U. W. A., Mishra, U., and Mapa, R. B.: National soil organic carbon estimates can improve global estimates, Geoderma, 337, 55–64, https://doi.org/10.1016/j.geoderma.2018.09.005, 2019.

Wright, M. N. and Ziegler, A.: Ranger: A fast implementation of random forests for high dimensional data in C++ and R, J. Stat. Softw., 77, 1–17, https://doi.org/10.18637/jss.v077.i01, 2017.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N., Ramoino, F., and Arinò, O.: ESA WorldCover 10 m 2020 v100 [data set], Zenodo, https://doi.org/10.5281/zenodo.5571936, 2021.

Zevenbergen, L. W. and Thorne, C. R.: Quantitative analysis of land surface topography, Earth Surf. Process. Landf., 12, 47–56, https://doi.org/10.1002/esp.3290120107, 1987.

Zhang, B., Xu, C., Zhang, Z., Hu, C., He, Y., Huang, K., Pang, Q., and Hu, G.: Response of soil organic carbon and its fractions to natural vegetation restoration in a tropical karst area, southwest China, Front. For. Glob. Chang., 6,

https://doi.org/10.5194/egusphere-2026-6423
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.



https://doi.org/10.3389/ffgc.2023.1172062, 2023.

Zomer, R. J., Xu, J., and Trabucco, A.: Version 3 of the Global Aridity Index and Potential Evapotranspiration Database, Sci. Data, 9, https://doi.org/10.1038/s41597-022-01493-1, 2022.