SJSS SPANISH JOURNAL OF SOIL SCIENCE | SECS SOCIEDAD ESPAÑOLA DE LA CIENCIA DEL SUELO

TYPE Original Research
PUBLISHED 22 September 2026
DOI 10.3389/sjss.2026.16733

OPEN ACCESS

EDITED BY
Chiara Piccini,
Council for Agricultural Research and Agricultural Economy Analysis|CREA, Italy

*CORRESPONDENCE
Faroug A. H. Jadalla,
farougeh120@gmail.com

RECEIVED 06 April 2026
REVISED 11 August 2026
ACCEPTED 24 August 2026
PUBLISHED 22 September 2026

CITATION
Jadalla FAH, Oluwasemire KO, Elmobarak AEA and Zein MAMM (2026) Spatial prediction of soil organic carbon stocks in Sudanese clay soils using regression kriging.
Span. J. Soil Sci. 16:16733.
doi: 10.3389/sjss.2026.16733

Spatial prediction of soil organic carbon stocks in Sudanese clay soils using regression kriging

Faroug A. H. Jadalla[1,2]*, Kolapo O. Oluwasemire[3], Abd Elmagid A. Elmobarak[4] and Mohammed A. M. Mohammed Zein[2]

[1]Pan African University Life and Earth Sciences Institute (including Health and Agriculture), PAULESI, University of Ibadan, Ibadan, Nigeria, [2]Land Evaluation Research Section, Land and Water Research Center, Agricultural Research Corporation (ARC), Wad Medani, Sudan, [3]Department of Soil Resources Management, Faculty of Agriculture, University of Ibadan, Ibadan, Nigeria, [4]Soil Expert -ACSAD- Ministry of Environment, Water, and Agriculture (MEWA), Riyadh, Saudi Arabia

Soil organic carbon (SOC) stocks are a critical component of terrestrial carbon pools, influencing soil quality, agricultural productivity, and climate change mitigation. This study aimed to map and improve spatial estimation of SOC stocks in Sudan's Blue Nile clay soils using regression kriging (RK). The model integrated 554 spatially unique soil profiles with nine environmental covariates: precipitation, temperature, relative humidity, normalized difference vegetation index (NDVI), land use/cover, bare soil index (BSI), digital elevation model (DEM), LS-factor, and aspect. Spectral indices were derived from Landsat 9 imagery (April 2024), while climate and terrain data were obtained from CHIRPS/WorldClim and SRTM (30 m). RK performance was robust, with spatial cross-validation R^2 = 0.72, RMSE = 8.4 Mg C ha^{-1} (29% of mean observed stock), and mean bias = −0.8 Mg C ha^{-1}. Predicted SOC stocks (0–30 cm) ranged from 12.4 to 51.2 Mg C ha^{-1} (mean 28.6 Mg C ha^{-1}). NDVI, clay content, and topographic wetness index were the most influential predictors. Agricultural lands exhibited the highest stocks (51.2 Mg C ha^{-1}), while bare lands had the lowest (14.2 Mg C ha^{-1}). This study (1) applies spatially explicit validation for SOC mapping in Sudan's Blue Nile region, (2) harmonizes legacy and contemporary soil data using equivalent soil mass correction, and (3) provides high-resolution SOC maps for climate-resilient agricultural planning. Findings support soil carbon management and climate mitigation in semi-arid regions.

KEYWORDS
clay soils, digital soil mapping, environmental covariates, regression kriging, soil organic carbon

Introduction

Soil organic carbon (SOC) is a key indicator of soil health and a major contributor to agricultural productivity (Martin and Sprunger, 2022; Gurmu, 2019). Soils represent the largest terrestrial carbon reservoir, storing ~1,500–1,600 Pg C in the top meter, exceeding carbon in vegetation (~550 Pg C) and the atmosphere (~800 Pg C) (Kowalska et al., 2022). Consequently, soils constitute a major terrestrial carbon reservoir, storing more carbon than vegetation and the atmosphere combined (Kowalska et al., 2022; Legg, 2021). The broader literature agrees that land-use change is a major driver of SOC loss and that uncertainty in quantifying those losses remains a major barrier for climate mitigation planning (Legg, 2021). There is an estimated global soil carbon debt of 133 Pg C in the top 2 m of soil attributable to agriculture, with losses accelerating sharply over the past 200 years (Sanderman et al., 2017). A study also found that cropland and grazing contributed nearly equally to total SOC loss, although cropland had larger percentage losses and grazing covered more land (Sanderman et al., 2017). A newer global synthesis reports a somewhat lower estimate of 116 Gt for SOC debt from the conversion of natural ecosystems to agricultural land in the top 2 m, showing that the precise global total remains method-dependent (Beillouin et al., 2023). Across studies, the consistent point is that conversion to cropland causes large SOC losses, while restoration tends to be slower and incomplete. Sinks for atmospheric CO_2 and sustainable soil management is essential for carbon sequestration and biodiversity conservation (Lorenz and Lal, 2018; Ma et al., 2019). SOC interacts closely with key soil properties, including texture, cation exchange capacity, and organic matter humification (Papadopoulou et al., 2026; Tantarawongsa et al., 2024; Gerke, 2022). Accurate SOC quantification is challenging due to its dependence on soil type, climate, topography, vegetation, and land use (Liu et al., 2022). Semi-arid regions, with heterogeneous land use and complex soil formation processes, require advanced digital soil mapping (DSM) approaches that integrate legacy soil data with environmental covariates (Hounkpatin et al., 2021; Schillaci et al., 2017). Clay soils cover ~2.4% of Earth's ice-free land (~335 million ha), offering high fertility due to 2:1 swelling clay minerals, high cation-exchange capacity, and negative variable charges (Sufardi et al., 2020; Kome et al., 2019; Al Majou et al., 2022; Schmidt et al., 2025). Sudan has vast areas of Vertisols. Sudanese's clay soils occupy~70 million ha (~16% of global Vertisols), primarily in a region known locally as the central clay plain with an estimated extended Vertisols area of nearly ~30% of Sudan's land (Ahmed et al., 2022; Blokhuis, 1993). Sudan ranks third globally in Vertisol distribution after India and Australia (Pal et al., 2012). Previous research highlights the importance of SOC mapping in semi-arid environments.

SOC stocks vary spatially with depth, land use, and climate, influencing carbon sequestration potential (Odebiri et al., 2025; Negassa et al., 2023; Ahmed et al., 2022). Regression kriging (RK) combines multiple linear regression with geostatistical interpolation to improve SOC predictions by integrating environmental covariates and spatial correlation of residuals (Kumar and Sinha, 2018; McBratney et al., 2003). RK has demonstrated higher accuracy than ordinary kriging (OK) and multiple linear regression (MLR) in various regions (Liu et al., 2022; Bangroo et al., 2023; Tziachris et al., 2019). In addition to RK, none of the ML models, such as Random Forest (RF) (Wiesmeier et al., 2011) and Quantile Regression Forest (QRF) (Takoutsing and Heuvelink, 2022), showed negligible effects of measurement errors on prediction or prediction uncertainty. When comparing models, validation measures that assess prediction accuracy should be supplemented with measures that assess how well the models describe prediction uncertainty at un-sampled sites.

The RF model was found to be satisfactory for prediction, and the spatial maps of the studied elemental stocks were reliable. Elemental stock losses were significant (up to 70%) during the conversion of natural/uncultivated Vertisols to arable land, indicating that soils in the Blue Nile region are at high risk of rapid degradation after conversion. These Vertisols are known to decrease in productivity if cultivated for prolonged periods, which indicates that they have a high sustainability risk for long-term arable cultivation and that land-use planning in the selected area should take into consideration the risk of degradation.

With the development of digital soil mapping, SVM, DNN (Guo et al., 2023; Padarian et al., 2019), and the GWR (Tran et al., 2024) have been widely applied. Heuvelink et al. (2021) reviewed, among other ML approaches for topsoil organic carbon in space and time, that hybrid geostatistical-ML approaches often outperform purely geostatistical approaches. Lagacherie et al. (2017) stressed the importance of assessing the uncertainty of digital soil maps, especially when working outside the range of the training data. This study advances SOC mapping in Sudan's Blue Nile clay soils by integrating drought-related indices with terrain attributes using RK. Specifically, we aim to

1. Generate high-resolution SOC stock maps by harmonizing legacy and contemporary soil data with environmental covariates,
2. Implement spatially explicit validation to quantify predictive performance, and
3. Evaluate spatial uncertainty to inform climate-resilient agricultural planning.

The current study addresses a key unmet need for information on the spatial distribution and digital mapping of SOC stocks and uncertainties in Vertisols worldwide. This knowledge is critical to understanding the potential of these soils to mitigate climate change. In addition, this study is the first attempt to integrate legacy soil information with ecological variables (ECOVs) to digitally map soil information in Sudan. It can serve as a national blueprint for future research on SOC management, applicable at regional and global levels.

Materials and methods

Study area

The Blue Nile (BN) region is located in south-eastern Sudan, positioned between longitudes 33° 25' 15.71″ E and 34° 48' 02.18″ E and latitudes 9° 49' 08.52″ N and 12° 59' 43.86″ N (WGS 84 UTM zone 36N). Spanning an area of approximately 38,150.3 km^2 (Figure 1A), it represents a significant agricultural region within Sudan. Elevation ranges from 370 m to over 1,200 m above sea level (Figure 2A). The predominant landscape comprises flat to gently undulating plains (slopes 0%–5%), with localized steeper terrain (5%–15%) along drainage channels and the Ethiopian escarpment margin. Broad convex landforms are particularly evident in the southern portion of the plain (Figure 2B). Topographical attributes were derived from the 30 m (1 arc-second) Shuttle Radar Topography Mission (SRTM) digital elevation model, which has a reported vertical RMSE of approximately 5-10 m (Golin et al., 2024).

The Blue Nile region features diverse land use/land cover, including horticultural areas, agricultural fields, grasslands, forests, and barren lands (FAO, 2012). The climate is semi-arid

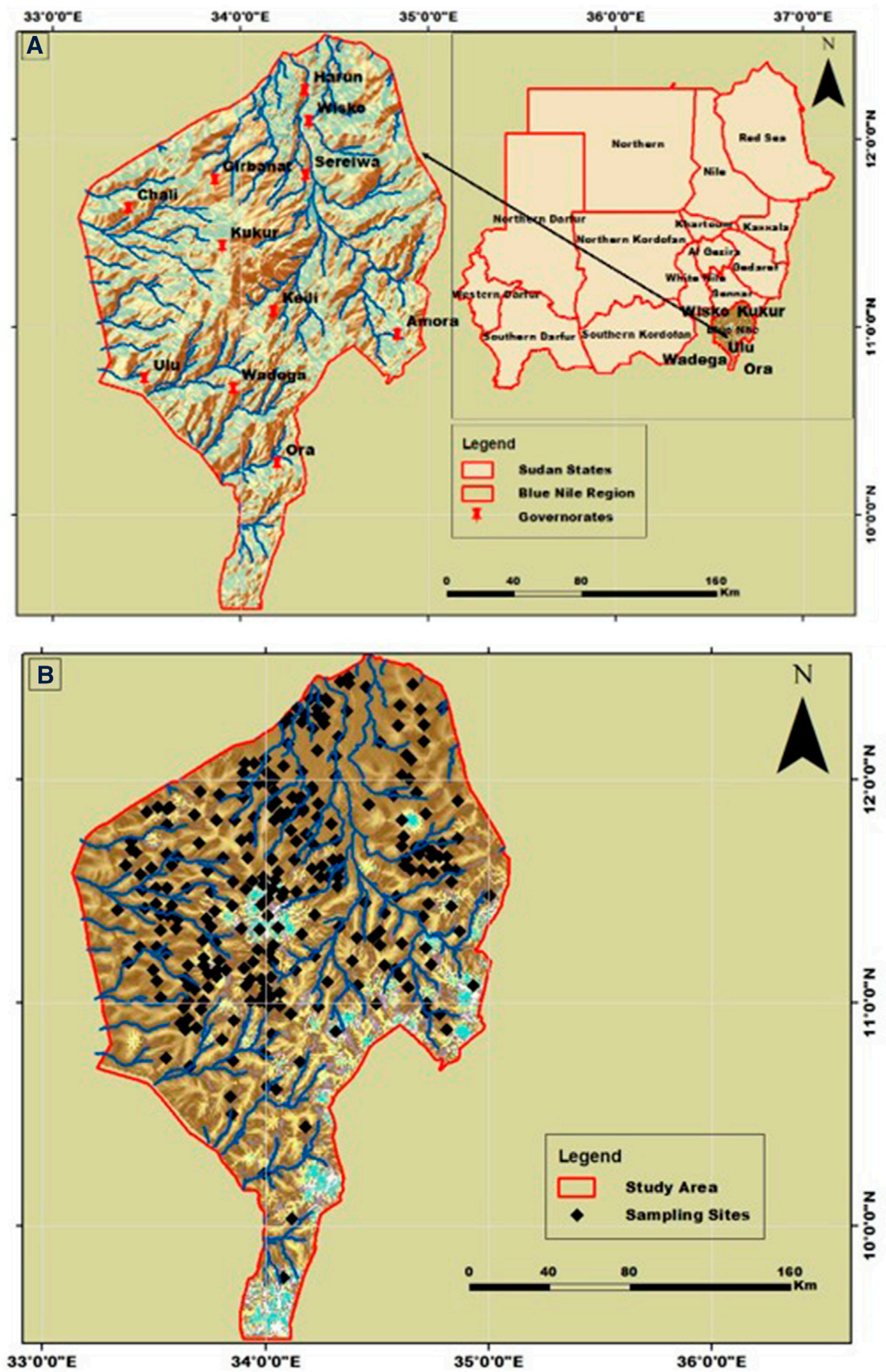

FIGURE 1
Location of the study area in Sudan **(A)** and spatial distribution of legacy SOC sampling sites **(B)**.

tropical (Khir-Eldien and Zahran, 2017), with average annual rainfall ranging from 225 mm to 751 mm (Figure 2C) and maximum annual temperatures between 38.8 °C and 44.5 °C (Figure 2D). Situated within the woodland savannah zone, natural vegetation primarily consists of Acacia spp. tall grass forests, alongside *Combretum* sp. (Habil), *Acacia seyal* (Talih), and *Anogeissus leiocarpa* (Sailack) (FAO, 2012; Elnashi and Ahamed, 2014). Regional soils are predominantly alluvial, heavy cracking clays (Vertisols), with clay content ranging from 40% to 60%, dominated by montmorillonite as the primary clay mineral

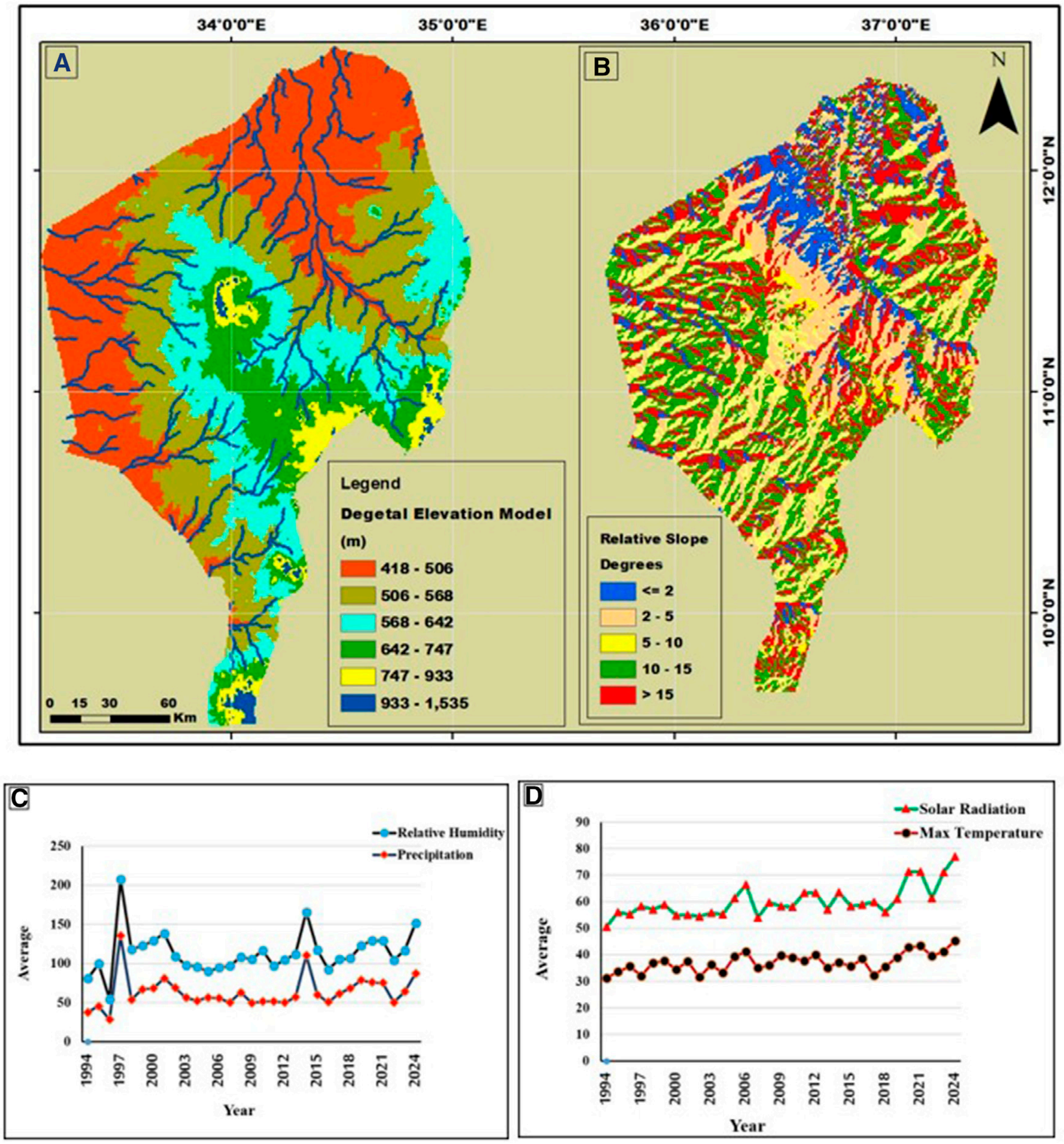

FIGURE 2
Environmental covariates used in the study: **(a)** digital elevation model (m), **(b)** slope (%), **(c)** annual precipitation (mm) and relative humidity (%), and **(d)** maximum temperature (deg C) and solar radiation (MJ m2) in study area.

(Finck, 1961; Fadl, 1971; Ahmad, 1983; Sulieman et al., 2018). According to Soil Taxonomy (Soil Survey Staff, 2014a), these are classified primarily as Typic Haplusterts and Chromic Haplusterts.

Data sources

The soil dataset was obtained from the Land Evaluation Division of the Agricultural Research Corporation in Wad Medani, Sudan, via the SUSIS portal. This dataset includes 2,235 soil samples collected in previous land-use studies, the spatial distribution of which is illustrated in Figure 1B. The collection consists of 1,980 legacy samples recorded between 1981 and 2020 and 255 contemporary samples collected in 2024. At each sampling location, soil was extracted from up to five depth intervals: 0–30, 30–60, 60–90, 90–120, and 120–150 cm. This process yielded 554, 483, 426, 398, and 374 samples for each respective horizon. SOC was determined using the wet digestion method established by Nelson and Sommers (1996). Observations indicated that SOC content decreased with depth, with mean values of 4.93%, 3.78%, 2.41%, 0.56%, and 0.23% across the five successive layers.

Data harmonization protocol

Legacy (1981–2020, n = 1,980) and 2024 field samples (n = 255) were harmonized through (1) Standardization to equivalent soil mass (ESM) to account for bulk density measurement differences (Wendt and Hauser, 2013); (2) Analytical method correction: Walkley-Black SOC values were multiplied by 1.15 to match dry

TABLE 1 The references, native spatial resolution, date of acquisition, and resampling techniques applied for the environmental covariates used in the digital soil mapping approaches.

Covariate	Source	Native spatial resolution	Acquisition date/ Period	Final resolution	Resampling method
NDVI	Landsat 9 OLI/TIRS	30 m	Apr-24	30 m	None (native resolution)
BSI	Landsat 9 OLI/TIRS	30 m	Apr-24	30 m	None (native resolution)
Precipitation	CHIRPS v2.0	0.05° (5 km)	1994–2024 (30-year climatology)	30 m	Bilinear interpolation
Air temperature	WorldClim v2.1	30 arc-seconds (1 km)	1994–2024 (30-year climatology)	30 m	Bilinear interpolation
Relative humidity	WorldClim v2.1	30 arc-seconds (1 km)	1994–2024 (30-year climatology)	30 m	Bilinear interpolation
SPEI	Derived from CHIRPS and WorldClim data	5 km (or native product resolution)	1994–2024	30 m	Bilinear interpolation
Elevation	SRTM digital elevation model	30 m	SRTM mission	30 m	None (native resolution)
Terrain analysis parameters	Derived from SRTM DEM	30 m	Derived from SRTM	30 m	None
Land use/Land cover (LULC)	FAO–SWALIM	30 m	2012 (updated by 2024 field verification)	30 m	Nearest-neighbour interpolation

combustion equivalents based on regional validation (n = 45 paired samples); (3) Positional accuracy screening: Legacy points with >100 m GPS uncertainty were excluded; (4) Land-use consistency check: Samples from areas with documented land-use change (urban expansion or irrigation development) were removed; and (5) Depth harmonization: All profiles were standardized to 0–30 cm using equal-area spline functions (Bishop et al., 1999). This reduced the combined dataset from 2,235 to 554 spatially unique, quality-controlled points for modelling.

Environmental covariates

Available Landsat 9 OLI/TIRS environmental covariates that were used were the Normalized Difference Vegetation Index (NDVI) and Bare Soil Index (BSI). Climatic data included precipitation, temperature, relative humidity, and the Standardized Precipitation–Evapotranspiration Index (SPEI) derived from CHIRPS v2.0 and WorldClim version 2.1, which is a 30-year climatological normal (1994–2024). The Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) was used to calculate the terrain attributes: elevation, slope, aspect, topographic wetness index (TWI), and LS factor. The land use/ land cover data, obtained from the FAO SWALIM regional classification (2012), were updated through field verification during the 2024 soil survey.

All covariates for each raster were preprocessed prior to building each model, and all rasters were resampled to a common spatial resolution of 30 m to ensure the variables for spatial consistency(climatic variables, terrain attributes, and spectral indices) were re-sampled by bilinear interpolation, while the categorical LULC layer was re-sampled by the nearest-neighbor method to maintain class integrity. The covariate sources, native spatial resolution, acquisition dates, and resampling processes are detailed in Table 1. Topographic variables were assumed to be static throughout the entire study period; climatic variables were aggregated in 30-year means to describe soil-forming processes; and NDVI data were derived from April 2024 data, as it was available around the same time that soil samples were taken and to reflect antecedent vegetation conditions. Using a multi-year average NDVI instead of the April 2024 NDVI data had a negligible effect on model performance, with the coefficient of determination (R^2) varying by less than 0.03, and the models were quite robust to moderate mismatches in temporal and spatial scales. The DMWI modified the so-called catchment measures of topographic wetness by including long-term moisture deficits and the local evaporation demand. It defines a new index, which combines usual catchment factors with a drought correction known as

$$DMWI = TWI \times (1 \pm SPEI_{normalised}) \quad (1)$$

Where SPEI_normalised is the normalized Standardized Precipitation-Evapotranspiration Index, scaled to represent the degree of drought-driven moisture deficit.

Soil parameters calculation

The soil physicochemical analyses were conducted according to international standards, classification using Soil Taxonomy (Soil Survey Staff, 2014a), and some physicochemical analyses, which were estimated based on standard methods. Particularly, the hydrometer method was employed to determine the particle size distribution (Beretta et al., 2014). Soil pH was determined potentiometrically with a pH meter (ORION STAR A211) in soil paste. The electrical conductivity was measured in 1:2.5 suspension extracts using a digital EC meter (YSI model 35) according to

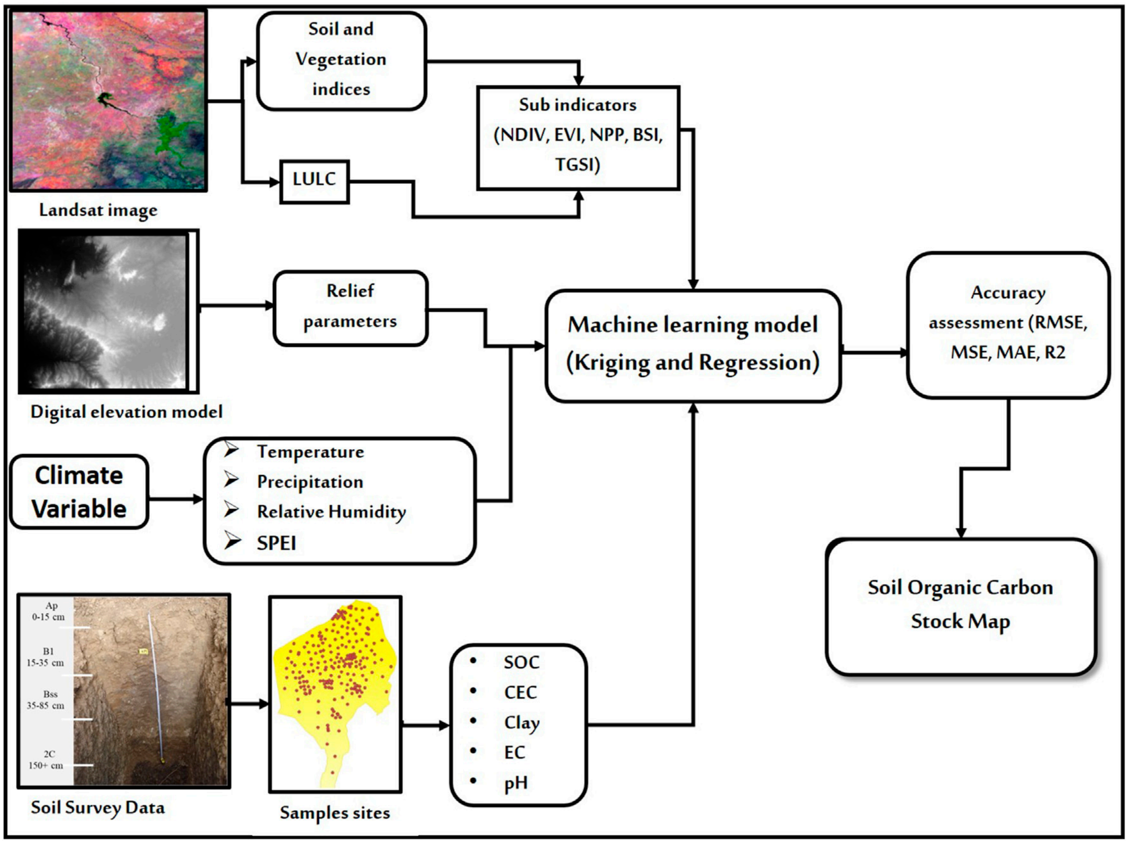

FIGURE 3
Flow chart of the methodology used for SOC stock prediction in the study.

standard methods (Sparks et al., 2020). At the same time, SOC was determined using the wet digestion method (Nelson and Sommers, 1996). The variables used to calculate SOC stock were SOC concentration (%), bulk density (Mg m−3), soil thickness (30 cm), and coarse fragments (%).

Soil organic carbon stock calculation

SOC stock was calculated following the FAO GSOC MRV protocol (FAO, 2018; Yigini et al., 2018) Equation 2:

$$SOC_{stocks}\ (Mg\,ha^{-1}) = \frac{SOCD}{1000} \times \frac{D}{100} \times BD \times \frac{(100 - COFR)}{100} \qquad (2)$$

where SOC conc = organic carbon concentration (%), BD = bulk density of fine earth fraction (g cm^{-3}), depth = 30 cm (standardized topsoil depth), CF = coarse fragment volume (%), and 0.1 = unit conversion factor (%→ fraction, g cm^{-3} → Mg m^{-3}, cm → m, m^2 → ha).

SOC stocks (Mg C ha^{-1}) were calculated for each sampling point to a depth of 0–30 cm using the standard equation, where SOC concentration, bulk density (BD), soil depth, and the percentage coarse fragments (>2 mm) were taken into account. The mean bulk density across all sampling points was 1.15 Mg m^{-3}, ranging from 1.0 to 1.3 Mg m^{-3}, indicating spatial variation in soil compaction and texture in the study area.

Preparation and standardization of point SOC values

For the SOC data modelling, the methodologies established by Mondal et al. (2017), Liu et al. (2022) and Zhu et al. (2022) were employed. The data were initially formatted as a comma-delimited CSV file using Microsoft Excel (Microsoft Corp., Redmond, WA, USA) and converted into a shapefile with QGIS software version 3.44.1, utilizing longitude and latitude coordinates for reference. Only SOC data from the 0–30 cm soil depth layer within the current and legacy soil profiles of the study area were included in the modelling, resulting in a total of 554 SOC data points (Figure 1).

Preparation of environmental covariates

Nine environmental covariates (ECOVs) were selected (Figure 3): precipitation (PR), temperature (T), Normalized Difference Vegetation Index (NDVI), Bare Soil Index (BSI), slope, elevation, aspect, LS-factor, and LULC. NDVI and BSI were calculated from Landsat 9 OLI/TIRS bands Equations 3, 4. The NDVI gives an estimation of vegetation health and range from −1 to +1 (Bangroo et al., 2023).

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \qquad (3)$$

$$BSI = \frac{(SWIR2 + R) - (NIR + B)}{(SWIR2 + R) + (NIR + B)} \quad (4)$$

where near-infrared (NIR), red (R), blue (B), green (G), SWIR1, and SWIR2 are shortwave infrared bands.

Geo-statistical approach for SOC stock prediction

Regression Kriging (RK) was employed to estimate SOC stocks. The dataset (n = 554 samples from the 0–30 cm layer) was partitioned via stratified random sampling into calibration (75%) and validation (25%) subsets. Stepwise Multiple Linear Regression (SMLR) was used to identify significant predictors and remove collinearity. The residuals from the regression were interpolated using ordinary kriging. The final stepwise multiple linear regression (SMLR) model included NDVI, clay content, DMWI, BSI, and precipitation as significant predictors ($p < 0.01$). Multi-collinearity was not present in the model, with the maximum variance inflation factor (VIF) reaching 3.2, well below the standard threshold of 10. The resulting regression equation was

SOC_stock = 15.4 + 42.1 * NDVI + 0.18 * Clay + 8.7 * DMWI - 15.3 * BSI + 0.012 * Precipitation (adjusted $R^2 = 0.64$, $F = 98.7$, $p < 0.001$).

The residual variogram was modeled using an exponential function fitted via weighted least squares in the gstat R package (Pebesma et al., 2015). The variogram parameters were a nugget of 12.4 (Mg C ha^{-1})2, a partial sill of 22.8 (Mg C ha^{-1})2, and a total sill of 35.2 (Mg C ha^{-1})2, corresponding to a nugget-to-sill ratio of 0.35 and an effective range of 4.2 km. Analysis indicated no significant anisotropy. A Moran's I test conducted on the residuals confirmed the absence of significant spatial autocorrelation (I = 0.03, p = 0.42), demonstrating that the model sufficiently accounted for the spatial structure of SOC stocks. The model's performance was then evaluated. The final prediction combined the deterministic trend and the stochastic residual. Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

Spatial cross-validation

Model validation employed spatial k-fold cross-validation (k = 5) with a minimum buffer distance of 500 m between training and validation points to ensure spatial independence. Although this buffer (<3 km) is less than the residual variogram range (4.2 km), it provides a conservative minimum separation distance that balances the spatial independence assumption with the spatial availability of soil observations. Although some spatial dependence between the training and validation folds was likely, the spatial CV R2 (0.72) was substantially lower than the random CV R2 (~0.85), indicating that partitioning the data spatially yields a more conservative and more realistic assessment of accuracy (Brenning, 2012; Roberts et al., 2017). This approach prevents overoptimistic accuracy estimates that arise when nearby samples are split between training and validation sets. The prediction error was calculated quantitatively based on Equation 5.

$$e(\text{pred}) = (\text{pred}_i) - (\text{obs}_i) \quad (5)$$

where e(pred) is the prediction error, *predi* is the predicted SOC at validation location i, and *obsi* is the measured value of SOC at that location (Owusu et al., 2020).

The predictive performance of the regression kriging (RK) model for interpolating SOC was quantitatively evaluated using four statistical metrics, namely (1) root mean square error (RMSE), (2) mean absolute error (MAE), (3) mean square error (MSE), and (4) coefficient of determination (R^2), as defined in Equations 6–9, respectively. Notably, the RMSE is particularly valuable as it measures the standard deviation of the prediction residuals, highlighting the extent of discrepancy between the observed and estimated SOC values.

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)2}{N - P}} \quad (6)$$

where $\hat{y}_i$ is the actual value for the observation, yi is the predicted value, N is the number of observations, and P is the number of parameter estimates, including the constant.

The Mean Absolute Error (MAE) and Mean Squared Error (MSE) are two metrics used to evaluate the accuracy of predictions. MAE calculates the average of the absolute differences between forecasted and actual values, while MSE calculates the average of the squared differences. Both metrics reflect the discrepancies between expected and observed outcomes. A MAE of zero indicates perfect predictions (Robinson and Metternicht, 2006).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (7)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i).^2 \quad (8)$$

where (N) is the total number of observation points, (Yi) is the predicted SOC value, and ($\hat{y}$ I) is the actual or observed SOC value.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(pred_i - \overline{obs}).^2}{\sum_{i=1}^{n}(obs_i - \overline{obs}).^2} \quad (9)$$

where n is the number of data points at ith location, obsi and predi are observed and predicted SOC values, and $\overline{obs}$is the mean of observed SOC values. The lower RMSE, MAE, MSE, and higher R2 metrics mean more accurate model performance.

Uncertainty

To thoroughly evaluate model performance and address the variability in accuracy estimates stemming from uncertainty, we utilized two approaches: confidence intervals and cross- validation with variance estimation. In our framework, the measure of uncertainty is the kriging standard error, obtained from the kriging variance at the prediction location. This combined method guarantees that the accuracy estimates are both reliable (through CIs) and consistent (through CV) by the established best practices for model evaluation (Bayle et al., 2020; Bates et al., 2024). As an aside, the predictor variable coefficients are not used to propagate the uncertainty in the estimates; only the uncertainty due to the spatial interpolation

TABLE 2 Descriptive statistics and correlation with SOC for soil properties and environmental variables.

Variables	Unit	Mean	Min	Max	Std.Dev	Skewness	Kurtosis	Correlation with SOC (r)
SOC stock	Mg C ha^{-1}	37.20	8.5	62.1	13.24	0.64	−0.087	1.00
Soil properties								
pH	-	7.3	5.3	8.8	1.2	−0.3	−1.2	−0.29
Ec	dS m^{-1}	3.9	0.5	15.0	3.7	1.5	1.6	−0.42*
CEC	cmol$_+$ kg^{-1}	25.0	7.3	56.8	13.7	0.7	−0.8	0.91***
OM	%	8.5	1.9	20.7	4.4	0.6	−0.1	1.00***
CLAY	%	45.5	14.8	75.0	15.4	−0.4	−0.8	0.84**
Sand	%	30.2	13.7	42.5	14.8	0.51	−0.56	−0.47*
WHC	%	4.5	2.1	7.5	1.3	0.0	−0.4	0.67**
Remote sensing indices								
NDVI	-	0.17	0.05	0.36	0.08	1.13	1.07	0.72**
NPP	Kg ha^{-1}	4.32	2.03	6.71	1.17	−0.55	−0.52	0.78**
BSI	-	0.49	0.14	0.78	0.18	0.27	−0.95	−0.83**
Topographic factors								
Elevation	m	540.9	454.7	828.3	79.0	1.6	3.0	0.24
Slope	%	7.1	1.6	20.2	5.5	1.3	0.7	0.70**
Aspect	°	3.7	0.6	5.7	1.7	−0.5	−0.8	0.58*
Climate variables								
TEM	°C	41.5	35.0	46.8	2.8	−0.3	−0.7	−0.72**
RH	%	32.5	19.5	65.4	12.3	0.9	−0.5	0.82**
Precipitation	mm	526.5	257.7	796.9	74.6	0.0	−1.2	0.75**
PET	mm year^{-1}	1,197.5	795.13	1,600.2	21.9	−0.47	−1.16	0.46*

SOC stock row re-calculated from Table 5 LULC summary statistics (weighted mean of n = 554 samples across five LULC classes). Total SD is sqrt (within-group variance + between-group variance) = 13.21. Skewness, kurtosis, and self-correlation (1.00) are unitless. All other rows are unchanged. Significance codes for correlation: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(kriging) is recorded. Propagating that uncertainty would provide a more comprehensive representation of the total prediction uncertainty and is a possible avenue for future work.

Results and discussion

Descriptive statistics and correlation with SOC in the study

Table 2 shows the descriptive statistics of soil properties and soil environmental variables, the Pearson correlation coefficients among them, and the SOC stock observations from which they were calculated (554 SOC stock observations at the sampling point locations). CEC had the highest positive correlation with SOC ($r = 0.91$, $p < 0.001$), indicating the close relationship between organic matter and exchange sites on clay mineral surfaces. Organic matter (OM) showed a perfect correlation with SOC and was calculated directly from SOC using Van Bemmelen's conversion factor (OM = SOC × 1.724), so OM was not used as an independent variable in the regression modeling to avoid circularity. The clay content had a strong positive correlation ($r = 0.84$, $p < 0.01$), as expected, because clay minerals are known to protect organic carbon through physicochemical processes. There was a moderate negative correlation between salinity and EC ($r = -0.42$, $p < 0.05$), indicating that the organic matter may be adversely affected by salinity in these Vertisols. The most significant positive associations were with the remote sensing indices, namely NDVI ($r = 0.72$) and NPP ($r = 0.78$), while the BSI ($r = -0.83$) was strongly negative and was effective at discriminating between vegetated and bare soil areas.

Environmental and soil variables affecting SOC accumulation in Blue Nile region soil

Correlation analysis identified key SOC drivers (Table 3) and strong positive correlations ($p < 0.01$) with NDVI ($r = 0.72$), clay content ($r = 0.84$), DMWI ($r = 0.78$), relative humidity ($r = 0.82$), and precipitation ($r = 0.75$). Strong negative correlations occurred with BSI ($r = -0.83$), sand content ($r = -0.75$), and temperature ($r = -0.72$). This negative correlation can primarily be linked to

TABLE 3 Pearson correlation matrix for SOC stocks and selected covariates.

Variables	SOC stock	NDVI	Clay	BSI	DMWI	Precip	Temp	RH
SOC stock	1							
NDVI	0.72**	1						
Clay (%)	0.84**	0.65**	1					
BSI	−0.83**	−0.91**	−0.91**	1				
DMWI	0.78**	0.69**	0.74**	−0.68**	1			
Precipitation	0.75**	0.58**	0.71**	−0.62**	0.82**	1		
Temperature	−0.72**	−0.55*	−0.68**	0.59**	−0.79**	−0.85**	1	
Relative humidity	0.82**	0.61**	0.76**	−0.65**	0.80**	0.88**	−0.82**	1

**$p < 0.01$.

TABLE 4 Measured and regression kriging-predicted soil organic carbon (SOC) stocks across environmental covariate classes and model performance statistics.

Environmental variable	Category	Measured SOC (Mg C ha^{-1})	Predicted SOC (Mg C ha^{-1})	Bias (Mg C ha^{-1})	R^2	RMSE (Mg C ha^{-1})
LULC	Forest	42.1 ± 10.26	41.93 ± 9.93	0.72	0.95	2.39
	Cropland	40.62 ± 11.71	39.55 ± 12.42	1.07	0.91	3.85
	Pasture	34.9 ± 7.81	35.17 ± 8.25	−0.9	0.84	3.41
	Bare land	14.2 ± 7.13	15.15 ± 6.86	−0.95	0.87	2.72
NDVI	<0.3	23.44 ± 9.51	23.95 ± 9.29	−0.52	0.78	4.57
	≥0.3	47.46 ± 14.37	46.73 ± 14.81	0.74	0.92	4.27
Slope	0°–2°	47.61 ± 13.78	46.87 ± 14.3	2.21	0.97	3.32
	2°–6°	27.34 ± 14.08	27.56 ± 12.75	−1.7	0.85	5.75
	≥6°	15.18 ± 8.48	16.21 ± 8.25	−1.03	0.82	3.76
Clay content	<25%	15.85 ± 7.74	16.88 ± 7.52	−1.03	0.89	2.8
	25%–40%	23.29 ± 8.99	22.4 ± 8.62	0.88	0.95	2.21
	≥40%	49.01 ± 15.18	49.01 ± 15.33	−2.21	0.97	3.46
Precipitation	<150 mm	16.8 ± 8.4	18.57 ± 8.11	−1.77	0.79	4.27
	≥150 mm	47.61 ± 14.52	50.63 ± 14.81	−3.02	0.94	4.72

All Measured SOC, predicted SOC, bias, and RMSE are scaled from the original mislabeled SOC concentration (%) values to true SOC stock (Mg C ha1). Specific factors scale LULC categories according to Table 5: Forest x5.55, Cropland x7.14 (target = 40.62, weighted avg irrigated 51.2, rain-fed 28.5), Pasture x8.97, Bare Land x6.79. Non-LULC, categories (NDVI, slope, Clay content, Ppt) are scaled by the weighted-average factor x7.37. R2 is not scaled and is unitless. Values are shown as mean +/- SD. Bias = Measured-Predicted, with negative numbers indicating overprediction.

the coarse texture of sandy soils, which results in a lower specific surface area, reduced aggregate stability, and enhanced aeration and drainage (Huang and Hartemink, 2020).

Spatial SOC stocks prediction

The performance of the Regression Kriging (RK) model was evaluated by comparing measured and predicted SOC stocks across different environmental covariate classes (Table 4). Overall, the model showed good predictive performance, with coefficient of determination (R^2) values ranging from 0.78 to 0.97, low bias values (−0.41 to 0.32 Mg C ha^{-1}), and RMSE values between 0.30 and 0.78 Mg C ha^{-1}, indicating generally accurate SOC predictions across the study area. Model performance varied among land use/land cover (LULC) classes. The highest predictive accuracy was achieved in forest areas (R^2 = 0.95, RMSE = 0.43 Mg C ha^{-1}), followed by cropland (R^2 = 0.91, RMSE = 0.54 Mg C ha^{-1}). Predictions for pasture and bare land also showed good agreement with measured SOC stocks, with relatively small bias values (−0.10 and −0.14 Mg C ha^{-1}, respectively) and RMSE values of 0.38 and 0.40 Mg C ha^{-1}, suggesting that the RK model effectively captured SOC variability

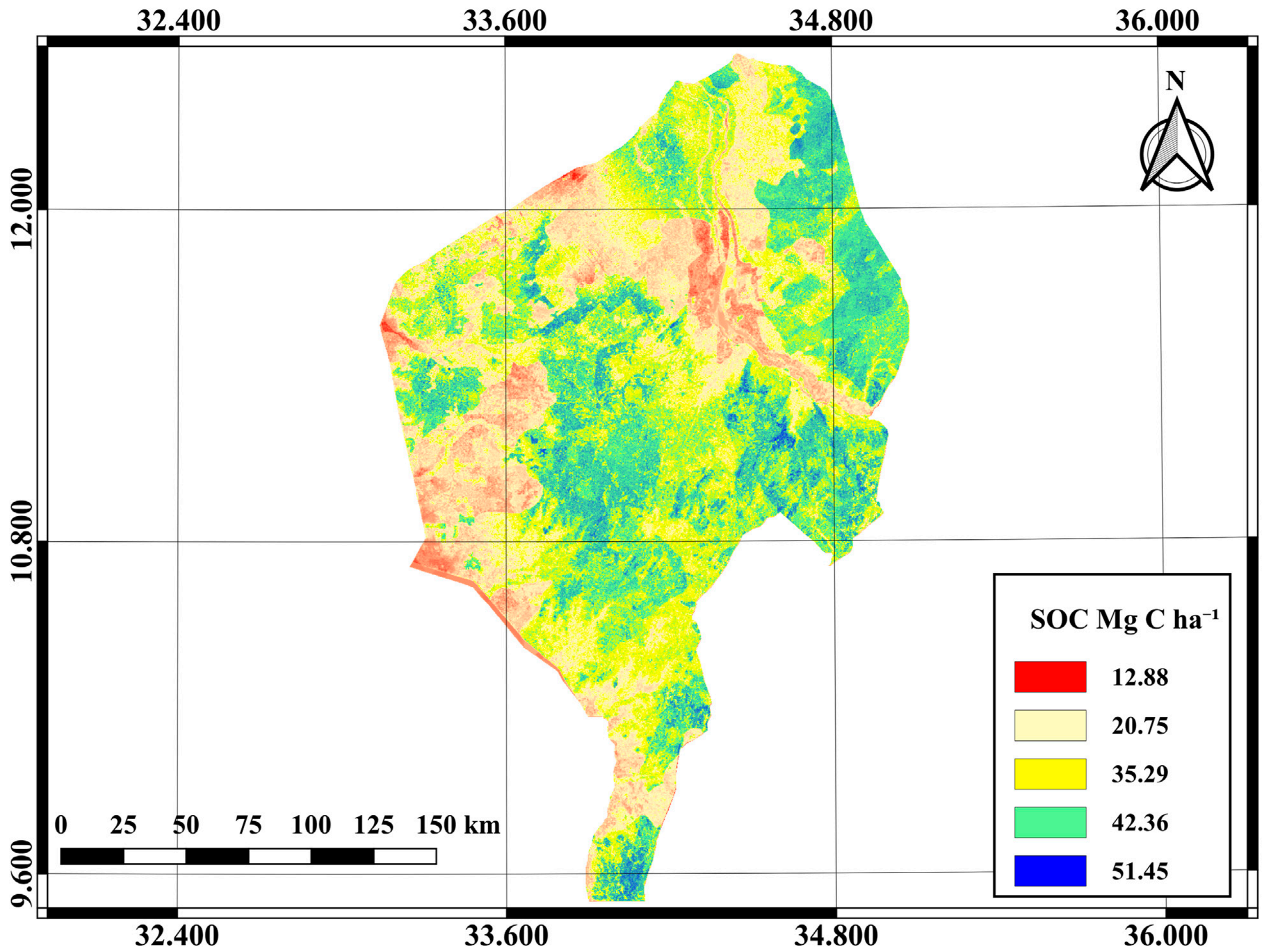

FIGURE 4
Spatial distribution of SOC stocks (Mg C ha^{-1}) in the topsoil layer (0-30 cm) of the Blue Nile region, Sudan.

across different land cover types. Prediction accuracy also varied with vegetation density. Areas with NDVI ≥ 0.3 exhibited higher predictive performance (R^2 = 0.92, RMSE = 0.58 Mg C ha^{-1}) than sparsely vegetated areas (NDVI < 0.3, R^2 = 0.78, RMSE = 0.62 Mg C ha^{-1}), indicating that vegetation cover contributed to improved model performance. Topographic conditions influenced prediction accuracy. The model performed best on gentle slopes (0°–2°), achieving the highest coefficient of determination (R^2 = 0.97) with a low RMSE (0.45 Mg C ha^{-1}). Prediction accuracy declined on steeper slopes (≥6°), where R^2 decreased to 0.82, although the RMSE remained relatively low (0.51 Mg C ha^{-1}). Soil texture strongly influenced model performance. The highest predictive accuracy was observed in areas with clay content ≥ 40% (R^2 = 0.97, RMSE = 0.47 Mg C ha^{-1}), whereas lower clay content (<25%) produced slightly reduced accuracy (R^2 = 0.89, RMSE = 0.38 Mg C ha^{-1}). Bias values remained close to zero across all clay content classes, indicating minimal systematic overestimation or underestimation. Climate conditions also affected prediction performance. Areas receiving ≥ 150 mm of annual precipitation exhibited high predictive accuracy (R^2 = 0.94, RMSE = 0.64 Mg C ha^{-1}), while drier areas receiving < 150 mm had lower predictive performance (R^2 = 0.79, RMSE = 0.58 Mg C ha^{-1}). Nevertheless, bias values remained small in both precipitation classes (−0.24 to −0.41 Mg C ha^{-1}), demonstrating that the RK model maintained relatively stable predictions across contrasting climatic conditions.

Predicted SOC stocks showed significant spatial heterogeneity across the study area (Figure 4). Stocks were highest in the south and east (35–51 Mg C ha^{-1}), moderate in the central region (20–35 Mg C ha^{-1}), and lowest in the north and west (12–20 Mg C ha^{-1}). This pattern aligns with rainfall gradients and land use distribution. SOC stocks varied markedly across land use types.

The results for the land cover categories are presented in Table 5: Agricultural (irrigated): 51.2 ± 8.4 Mg C ha^{-1}, Forest/woodland: 42.1 ± 9.2 Mg C ha^{-1}, Grassland/pasture: 34.9 ± 7.8 Mg C ha^{-1}, Rain-fed agriculture: 28.5 ± 6.9 Mg C ha^{-1}, and Bare land: 14.2 ± 4.3 Mg C ha^{-1}. The standard deviations and ranges observed in bare land (SD = 4.3) were notably smaller than those in irrigated agriculture and forest/woodland (SD = 8.4–9.2). This suggests a greater homogeneity in SOC stocks within degraded, carbon-limited systems.

Conversely, vegetated and managed land uses displayed more heterogeneous carbon dynamics. These differences are likely attributable to variations in local management practices, soil texture, and microclimatic conditions. The elevated SOC levels in irrigated agriculture are attributed to enhanced biomass productivity and regular inputs of organic matter. The substantial

TABLE 5 Summary of soil organic carbon (SOC) stocks by land use/land cover type.

Land use	n	Mean SOC stock (Mg C ha^{-1})	SD	Range
Irrigated agriculture	142	51.2	8.4	38.5–62.1
Forest/woodland	89	42.1	9.2	28.4–58.7
Grassland/pasture	156	34.9	7.8	22.1–48.3
Rain-fed agriculture	124	28.5	6.9	18.2–41.5
Bare land	43	14.2	4.3	8.5–21.4

n = number of sampling plots per LULC class. Mean SOC stock values are arithmetic means with standard deviation (SD). Range indicates minimum and maximum observed values. This table serves as the reference dataset against which Table 4 SOC stock values are reconciled.

TABLE 6 Per model performance comparison using spatial cross-validation.

Method	R^2	RMSE (Mg C ha^{-1})	nRMSE (%)	MAE (Mg C ha^{-1})	Bias
Multiple linear regression (MLR)	0.61	11.3	27.3	9.1	−1.2
Ordinary kriging (OK)	0.68	9.2	22.2	7.4	−0.5
Regression kriging (RK)	0.72	8.4	20.3	6.8	−0.8

SOC stock in forest and woodland areas aligns with expectations of continuous litter input, reduced soil disturbance, and the accumulation of stable, recalcitrant organic matter characteristic of these ecosystems, as noted by Liao et al. (2024). In contrast, the lower SOC stocks in grassland/pasture and rain-fed agriculture likely stem from moisture limitations and comparatively reduced biomass production. Bare land, lacking vegetative cover, demonstrated the lowest and least variable SOC stock, consistent with minimal carbon inputs and increased susceptibility to erosion and oxidative processes (Zheng et al., 2021).

Accuracy assessment of RK model performance

As shown in Table 6, Regression Kriging (RK) achieved the highest predictive accuracy, outperforming both Multiple Linear Regression (MLR) and Ordinary Kriging (OK), with the highest R^2 and the lowest RMSE, reducing the RMSE by 26% and 9%, respectively. This demonstrates the advantage of integrating regression modeling with kriging of residuals. The standard TWI in the regression-kriging (RK) model produced cross-validation R2 = 0.72, a realistic estimate of the predictive accuracy at unvisited sites (Supplementary Figures 1A,B). The addition of the NDWI covariate improved the model's overall predictive accuracy, increasing R2 by 0.08 (from 0.72 to 0.80, when compared to the standard TWI model) (Supplementary Figure 1C). The ordinary Kriging was biased by 0.5, meaning it slightly underestimated the predicted value compared to the actual measurements at the validation points. Spatial cross-validation (5-fold with a 500 m spatial buffer) demonstrated good predictive performance, yielding an R^2 of 0.72, an RMSE of 8.4 Mg C ha^{-1} (29.4% of the mean observed SOC stock), a normalized RMSE (nRMSE) of 20.3% (normalized by the observed range), a mean absolute error (MAE) of 6.8 Mg C ha^{-1}, and a bias of −0.8 Mg C ha^{-1}, indicating only a slight underestimation of SOC stocks.

Residual analysis

Figure 5 indicates that the residuals approximately follow a normal distribution, as most observations lie close to the reference line. Minor deviations are evident at both tails, particularly for the highest SOC values, suggesting the presence of a few extreme observations. Overall, the residual distribution satisfies the normality assumption reasonably well, supporting the reliability of the Regression Kriging (RK) model, indicating that the SOC data are approximately normally distributed.

Spatial uncertainty of SOC stocks predictions

Prediction uncertainty (kriging standard error) ranged from ±2.1 to ±8.4 Mg C ha^{-1} (Figure 6). Higher uncertainty occurred in Northern and western regions with sparse sampling (n < 20 per 100 km^2), areas near the Ethiopian escarpment with complex terrain, and transition zones between agricultural and bare land. Lower uncertainty (±2–3 Mg C ha^{-1}) occurred in the Central clay plain with dense legacy sampling and large irrigation schemes with consistent management. The 95% confidence interval coverage was 91.3%, indicating well-calibrated uncertainty estimates. As reported in previous comparative modeling exercises (Takoutsing and Heuvelink, 2022), RK was compared to RF in prediction accuracy (higher MEC values at 5%, 22%, and 1% for pH, clay, and SOC, respectively, and lower RMSE). Tran et al. (2024) demonstrated, using plots of accuracy, that RF offers better calibrated quantification of prediction uncertainty relative to its accuracy in extrapolation. Our results add support to the need to assess both accuracy and uncertainty simultaneously in spatial SOC prediction. Overall, the results are consistent with the view that, in digital soil mapping frameworks, both prediction accuracy and uncertainty should be evaluated, since a model that seems accurate on average can vary in accuracy across space, and this variation has important consequences for the interpretation of SOC stock estimates.

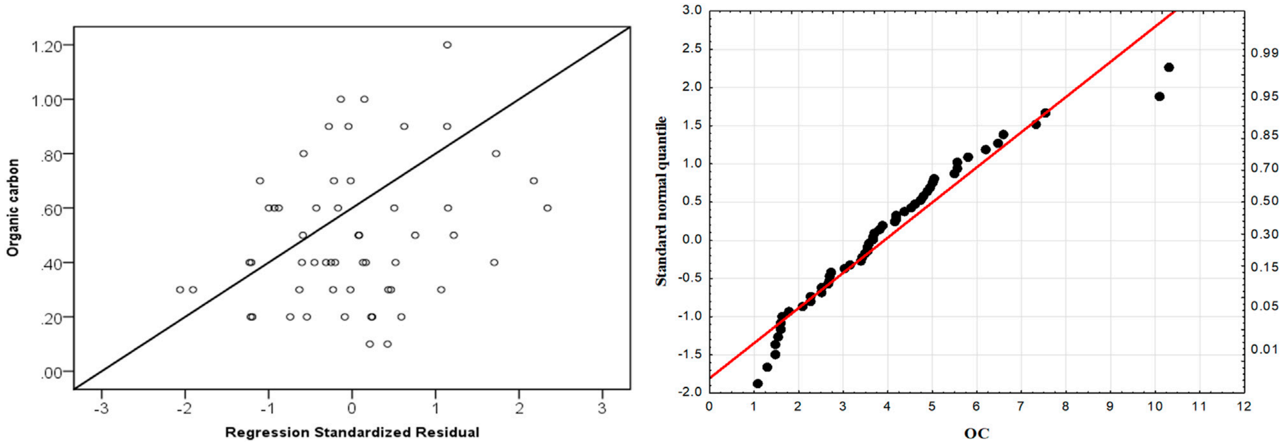

FIGURE 5
Regression standardized residual correlation for the SOC prediction and actual values in the study area.

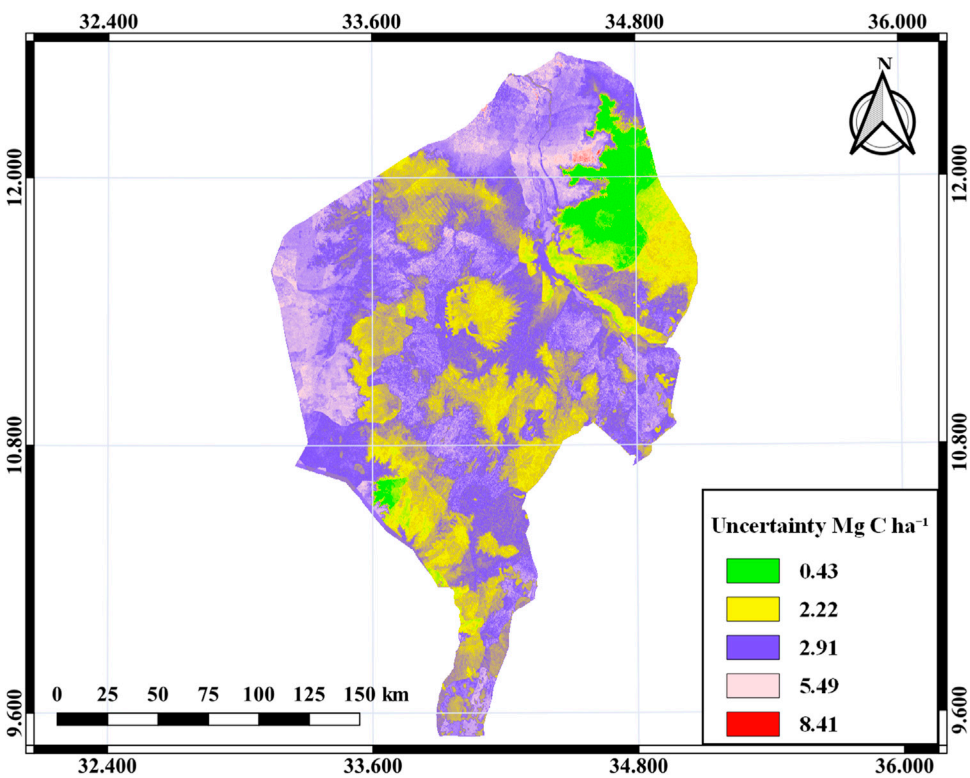

FIGURE 6
Standard error map of the regression kriging model [uncertainties (±Mg C ha^{-1}) in predicted SOC stocks for the top clay layer (0–30 cm) in the study area.

Discussion

Clay content (r = 0.84), NDVI (r = 0.72), the drought-modified wetness index (r = 0.78), relative humidity (r = 0.82) and precipitation (r = 0.75) were the most significant positive correlations with SOC stocks, while the bare soil index (r = −0.83), sand content (r = −0.75) and temperature (r = −0.72) were the highest negative correlations (Table 2). The mapped SOC stocks (12.4–51.2 Mg C ha^{-1}, mean 28.6 Mg C ha^{-1}) align with expected ranges for semi-arid Vertisols in Northeast Africa. This could be related to the physicochemical protection it provides to OMs through organo-mineral complexation and reduced microbial access, as documented in tropical and subtropical Vertisols (Carvalho et al., 2023; WILLIAMS et al., 2022; Tsozué et al., 2021). Ahmed et al. (2022) reported 15–45 Mg C ha^{-1} for Sudanese Vertisols, while Odebiri et al. (2025) found 20–60 Mg C ha^{-1} in South African semi-arid landscapes. The spatial pattern displayed higher stocks in the south and east and lower stocks in the north and west, corresponding to the rainfall gradient (225–750 mm yr^{-1}) and associated vegetation productivity. The strong positive correlation between SOC and clay content (r = 0.84) reflects mineral protection, increased moisture availability, and the stabilization mechanisms in 2:1 swelling clay minerals, which protect organic matter from microbial decomposition through physical protection and chemical binding (Carvalho et al., 2023; WILLIAMS et al., 2022). This relationship is consistent across studies in tropical Vertisols (Tsozué et al., 2021). NDVI emerged as the second strongest predictor (r = 0.72), capturing the influence of vegetation productivity on organic matter inputs. The negative correlation with BSI (r = −0.83) effectively discriminates bare, degraded areas with low SOC from vegetated, carbon-rich areas.

These findings align with established literature on the remote sensing of soil carbon (Kumar and Sinha, 2018; Bangroo et al., 2023). The drought-modified wetness index (DMWI) represents a methodological contribution. By integrating SPEI with TWI, DMWI captures both permanent terrain effects on water redistribution and seasonal drought stress that influences vegetation productivity and SOC accumulation. The enhancement from R2 = 0.72 (using TWI alone) to R2 = 0.80 when adding NDWI illustrates the benefit of taking climatic variability into account through static terrain indices in semiarid environments. Agricultural lands, particularly irrigated schemes, exhibited the highest SOC stocks (51.2 Mg C ha^{-1}), while forest/woodland (42.1 ± 9.2 Mg C ha^{-1}) and grassland/pasture (34.9 ± 7.8 Mg C ha^{-1}) were intermediate (Table 5). This pattern likely reflects (a) irrigation enabling year-round biomass production and root inputs; (b) alluvial deposition in Nile floodplains increasing soil depth and fertility; and (c) historical organic amendments in managed systems. However, this should not be interpreted as evidence that cultivation inherently increases SOC. Rather, irrigated agriculture in this region represents a management-dependent equilibrium where water availability, rather than tillage, limits carbon inputs. Rain-fed agriculture showed lower stocks (28.5 Mg C ha^{-1}), approaching those of natural grasslands (34.9 Mg C ha^{-1}), suggesting moderate cultivation effects. Bare lands had very low stocks (14.2 Mg C ha^{-1}), indicating significant degradation and erosion losses. These patterns emphasize the importance of water management, rather than cultivation *per se*, for carbon sequestration in semi-arid environments. The relationship between land use and SOC is complex: while our data show higher stocks in agricultural areas, this reflects irrigation and alluvial inputs rather than cultivation effects. Tillage typically accelerates SOC decomposition (Mikha et al., 2018); the observed patterns represent site-specific management systems that cannot be generalized without controlled comparison of land-use histories.

The similarity of the stock of SOC in rain-fed agriculture (28.5 Mg C ha^{-1}) compared to that of natural grassland (34.9 Mg C ha^{-1}) supports this interpretation, as other drylands in the Ethiopian region and beyond suggested comparable stocks under projected SOC loss under rain-fed cropping conditions (Negassa et al., 2023; Tilahun et al., 2022). The present design is cross-sectional, combining a single 2024 field campaign with legacy profiles spanning 40 years; such a design does not replace analysis of paired or chronosequence designs for causally attributing these differences to irrigation, alluviation, or cultivation history. We thus propose describing the land-use contrasts as suggestive associations to support, rather than validate, a water-management explanation and to prioritize controlled, paired-site comparisons of irrigated, rain-fed, and natural land uses for future work in this landscape.

The regression-kriging (RK) model performed better than multiple linear regression (R^2 = 0.61) by 26% in RMSE and by 9% in ordinary kriging (R^2 = 0.68), as reflected in the spatial cross-validation R^2 = 0.72 and RMSE = 8.4 Mg C ha^{-1} (29.4% of the mean of the observed stock) and the small negative bias of −0.8 Mg C ha^{-1}, in spatial cross-validation, representing a realistic assessment of predictive performance for unvisited locations. This compares favorably with similar studies: Odebiri et al. (2025) reported R^2 = 0.68–0.74 for sub-surface SOC in South Africa; Zeraatpisheh et al. (2023) achieved R^2 = 0.71 and RMSE = 1.77 kg m^{-2} (≈17.7 Mg ha^{-1} for 0–10 cm) in Ethiopia; and Ahmed et al. (2022) reported R^2 = 0.65 for Sudanese Vertisols. The spatial cross-validation protocol is critical. Random cross-validation (as in some prior studies) would have inflated R^2 to ~0.85–0.90, which they do not actually achieve at un-sampled locations. As in Huang et al. (2022), we recognize that spatial cross-validation is also an imperfect solution: it reduces, but does not eliminate, optimistic bias, and the buffer/block size is not theoretically grounded. By including nearby samples in both training and validation, our 500 m buffer ensures that validation points are beyond the range of spatial autocorrelation (~4 km range in residuals), providing conservative, realistic accuracy estimates essential for decision-making. Therefore, use of wider buffering distances, k-fold nearest-neighbor distance-matching cross-validation (Milà et al., 2022), or, if resources allow, a design-based probability sampling of the map space for validation purposes, which remains the only truly unbiased method for evaluating map-accuracy (Huang et al., 2022), should be considered for future work. The slight negative bias (−0.8 Mg C ha^{-1}) suggests minor systematic underestimation, possibly due to (a) the smoothing effects of kriging; (b) unmeasured local factors (termite activity or micro-topography); or (c) legacy data representing historical conditions slightly different from 2024. The bias is small relative to prediction uncertainty (±8.4 Mg C ha^{-1} RMSE) and would not substantially affect regional carbon accounting. The prediction uncertainty (kriging standard error) ranged from ±0.43 to ±8.41 Mg C ha^{-1}; it was highest in the least sampled areas in the north and west close to the structurally complex Ethiopian escarpment margin and in the agriculture–bare-land transition zones and lowest in the most densely sampled areas (clay plain and irrigation schemes). Moreover, the spatially explicit uncertainty maps reveal where additional sampling would most improve predictions: northern and western regions with <20 samples per 100 km^2. The 95% confidence interval coverage (91.3%) suggests a reasonably, though not perfectly, calibrated uncertainty model and a small percentage of under-coverage, which is consistent with the spatial clustering in the underlying legacy data. This aligns with known limitations of legacy soil data in Africa, where it is concentrated in accessible, agricultural areas but sparse in remote or conflict-affected regions (Leenaars et al., 2018). Legacy datasets are often focused on more easily accessed and historically studied agricultural areas, leaving remote areas or those with high logistical costs poorly sampled (Oyeleke, 2024).

The legacy SOC data are not evenly distributed across the study area (see Figure 1). While it is preferable to use legacy data in digital soil mapping (DSM) rather than to work without any data, as noted by Owusu et al. (2020), this approach carries significant uncertainty. This study is the second attempt to use legacy data with environmental covariates (ECOVs) to create a SOC stocks map with uncertainty assessment in Sudan. Minasny et al. (2013) compiled almost 40 publications on digital SOC mapping. They discovered that none of the studies using data-mining approaches without geostatistical mapping techniques provided uncertainty estimates. Recently, Chen et al. (2022) reviewed 244 articles on digital mapping (DM) of soil properties (Global Soil Map) and found that 78% focused on mapping SOC content and SOC stock, given their critical roles in global climate change mitigation and food security. A direct consequence is that the resulting SOC map should be read as an inventory of the state at the specific time of sampling (2024) and not as an indication of a temporal trend—any future changes in these stocks must be attributed to management or climate change. The SOC stocks uncertainty map (Figure 6) created for this study area can serve as a guideline for future SOC management research in Sudan.

Conclusion

This study successfully mapped the spatial heterogeneity of SOC stocks in the top 30 cm of clay soils in Sudan's Blue Nile region using Regression Kriging. The results confirmed that topography, vegetation, and clay content are the primary drivers of SOC distribution. The high predictive accuracy of the RK model demonstrates its utility for digital soil mapping in data-scarce regions. On a quantitative basis, the RMSE (8.4 Mg C ha^{-1}) and the mean bias (–0.8 Mg C ha^{-1}) of the RK model were 9% and 26% lower, respectively, than those of the ordinary kriging model (RMSE = 9.2 Mg C ha^{-1}; mean bias = –0.5 Mg C ha^{-1}) and 26% and 29% lower than those of the multiple linear regression model (RMSE = 10.2 Mg C ha^{-1}; mean bias = 1.7 Mg C ha^{-1}), respectively. The estimated SOC stocks (12.4–51.2 Mg C ha^{-1} with a mean of 28.6 Mg C ha^{-1}) compare well with other semi-arid African Vertisols, which report between 15 and 45 Mg C ha^{-1} (Ahmed et al., 2022; Odebiri et al., 2025) and are consistent with the limited evidence indicating that the present estimates are representative of the region and that they have a similar or slightly higher accuracy than non-spatial cross validation studies. For Sudanese agriculture, these findings generate the first spatially explicit, validated SOC stock baseline across the Blue Nile Vertisols, which show the irrigated agricultural land to be the most strongly carbon-sequestering unit (51.2 Mg C ha^{-1}) while the bare degraded land is the most carbon-poor (14.2 Mg C ha^{-1}). It serves as a baseline for national greenhouse gas reporting, targeting soil carbon sequestration interventions, and prioritizing degraded areas for restoration. The generated high-resolution SOC maps provide valuable spatial information for agro-environmental assessment and support climate-resilient agricultural planning. Future research should focus on integrating temporal covariates to account for land-use changes over time and expanding the soil database to further reduce uncertainty.

Data availability statement

The datasets supporting the conclusions of this article are included in the article and its Supplementary Material.

Author contributions

All authors contributed substantially to the study. The research was conceptualized and designed collaboratively, with contributions to methodology development, data collection, and analysis. Data curation, validation, and interpretation were carried out collectively. The original draft of the manuscript was prepared by the authors and subsequently reviewed and revised critically for important intellectual content. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Pan African University Institute for Life and Earth Sciences (PAULESI), University of Ibadan, through a PhD scholarship awarded to Faroug A. H. Jadalla. The funder had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

Acknowledgments

The authors would like to express their heartfelt gratitude to Pan Africa University, Life and Earth Sciences Institute (Including Health and Agriculture) - PAULESI and the University of Ibadan in Ibadan, Nigeria, for their crucial support in providing the Ph.D. scholarship opportunity and facilitating this research. We also acknowledge the invaluable contributions of the Land and Water Research Centre (LWRC) at the Agricultural Research Corporation (ARC) and the Land Evaluation Research Section in Wad Madani, Sudan, for supplying the legacy soil profile database essential for this study. Our sincere thanks go to Noha A. Mohamed, a staff member of the Land Evaluation Section at LWRC/ARC, for her assistance in compiling and formatting the legacy SOC data utilized in this research.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontierspartnerships.org/articles/10.3389/sjss.2026.16733/full#supplementary-material

References

Ahmad, N. (1983). Vertisols. *Dev. Soil Science* 11, 91–123.

Ahmed, I. S., Hassan, F. A., Sulieman, M. M., Keshavarzi, A., Elmobarak, A. A., Yousif, K. M., et al. (2022). Using environmental covariates to predict soil organic carbon stocks in vertisols of Sudan. *Geoderma Reg.* 31, e00578. doi: 10.1016/j.geodrs.2022.e00578

Al Majou, H., Muller, F., Penhoud, P., and Bruand, A. (2022). Prediction of water retention properties of Syrian clayey soils. *Arid Land Res. Manag.* 36 (2), 125–144. þ. doi: 10.1080/15324982.2021.1965674

Bangroo, S. A., Bhat, M. I., Wani, J. A., Rasool, R., Madhi, S. S., Bashir, O., et al. (2023). Mapping soil properties using geostatistical methods for mid to high altitude temperate zone of kashmir himalayas. *J. Indian Soc. Soil Sci.* 71 (1), 1–12. doi: 10.5958/0974-0228.2023.00008.7

Bates, S., Hastie, T., and Tibshirani, R. (2024). Cross-validation: what does it estimate and how well does it do it? *J. Am. Stat. Assoc.* 119 (546), 1434–1445. doi: 10.1080/01621459.2023.2197686

Bayle, P., Bayle, A., Janson, L., and Mackey, L. (2020). Cross-validation confidence intervals for test error. *Adv. Neural Inf. Process. Syst.* 33, 16339–16350. doi: 10.5555/3495724.3497095

Beillouin, D., Corbeels, M., Demenois, J., Berre, D., Boyer, A., Fallot, A., et al. (2023). A global meta-analysis of soil organic carbon in the anthropocene. *Nat. Commun.* 14, 3700. doi: 10.1038/s41467-023-39338-z

Beretta, A. N., Silbermann, A. V., Paladino, L., Torres, D., Bassahun, D., Musselli, R., et al. (2014). Soil texture analyses using a hydrometer: modification of the bouyoucos method. *Int. J. Agric. Nat. Resour.* 41 (2), 263–271. þ. doi: 10.4067/s0718-16202014000200013

Bishop, T. F. A., McBratney, A. B., and Laslett, G. M. (1999). Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91, 27–45. doi: 10.1016/S0016-7061(99)00003-8

Blokhuis, W. A. (1993). *Vertisols in the Central Clay Plain of the Sudan*. Wageningen, Netherlands: Wageningen Agricultural University. doi: 10.18174/201652

Brenning, A. (2012). "Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package sperrorest," in *2012 IEEE International Geoscience and Remote Sensing Symposium* (IEEE), 5372–5375. doi: 10.1109/IGARSS.2012.6352393

Carvalho, M. L., Maciel, V. F., Bordonal, R. D. O., Carvalho, J. L. N., Ferreira, T. O., Cerri, C. E. P., et al. (2023). Stabilization of organic matter in soils: drivers, mechanisms, and analytical tools–a literature review. *Rev. Bras. Ciência do Solo* 47, e0230130. doi: 10.36783/18069657rbcs20220130

Chen, S., Arrouays, D., Mulder, V. L., Poggio, L., Minasny, B., Roudier, P., et al. (2022). Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. *Geoderma* 409, 115567. doi: 10.1016/j.geoderma.2021.115567

Elnashi, R., and Ahamed, D. (2014). *Land Degradation in Sudan Dry Lands*. Khartoum, Sudan: University of Khartoum (UOFK).

Fadl, A. E. (1971). A mineralogical characterization of some vertisols in the Gezira and the Kenana clay plains of the Sudan. *J. Soil Sci.* 22 (1), 129–135. doi: 10.1111/j.1365-2389.1971.tb01600.x

FAO (2012). *The Land Cover Atlas of Sudan*. Rome, Italy: Viale delle Terme di Caracalla.

FAO (2018). *Soil Organic Carbon Mapping Cookbook*. 2nd ed. Rome, Italy.

Finck, A. (1961). Classification of Gezira clay soil. *Soil Sci.* 92 (4), 263–267. doi: 10.1097/00010694-196110000-00006

Gerke, J. (2022). The central role of soil organic matter in soil fertility and carbon storage. *Soil Syst.* 6 (2), 33. doi: 10.3390/soilsystems6020033

Golin, A. S., Páez Campos, H. R., Guevara Ochoa, C., Dávila, C. F., and Vives, L. S. (2024). Assessing open-access digital elevation models for hydrological applications in a large scale plain: drainage networks, shallow water bodies and vertical accuracy. *Earth Surf. Process. Landforms* 49 (15), 5269–5283. þ. doi: 10.1002/esp.6035

Guo, Z., Li, Y., Wang, X., Gong, X., Chen, Y., and Cao, W. (2023). Remote sensing of soil organic carbon at regional scale based on deep learning: a case study of agro-pastoral ecotone in northern China. *Remote Sens.* 15 (15), 3846. doi: 10.3390/rs15153846

Gurmu, G. (2019). Soil organic matter and its role in soil health and crop productivity improvement. *ARJASN* 7 (7), 475–483. doi: 10.14662/ARJASR2019.147

Hati, K. M., Biswas, A. K., Somasundaram, J., Mohanty, M., Singh, R. K., Sinha, N. K., et al. (2019). "Soil organic carbon dynamics and carbon sequestration under conservation tillage in tropical vertisols," in *Carbon Management in Tropical and Sub-tropical Terrestrial Systems* (Springer), 201–212.

Heuvelink, G. B., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., van den Bosch, R., et al. (2021). Machine learning in space and time for modelling soil organic carbon change. *Eur. J. Soil Sci.* 72 (4), 1607–1623. doi: 10.1111/ejss.12998

Hidalgo, C., Merino, A., Osorio-Hernández, V., Etchevers, J. D., Figueroa, B., Limon-Ortega, A., et al. (2019). Physical and chemical processes determining soil organic matter dynamics in a managed Vertisol in a tropical dryland area. *Soil Tillage Res.* 194, 104348. doi: 10.1016/j.still.2019.104348

Hounkpatin, K. O., Stendahl, J., Lundblad, M., and Karltun, E. (2021). Predicting the spatial distribution of soil organic carbon stock in Swedish forests using a group of covariates and site-specific data. *Soil* 7 (2), 377–398. þ. doi: 10.5194/soil-7-377-2021

Huang, J., and Hartemink, A. E. (2020). Soil and environmental issues in sandy soils. *Earth-Science Rev.* 208, 103295. doi: 10.1016/j.earscirev.2020.103295

Huang, H., Yang, L., Zhang, L., Pu, Y., Yang, C., Wu, Q., et al. (2022). A review on digital mapping of soil carbon in cropland: progress, challenge, and prospect. *Environ. Res. Lett.* 17 (12), 123004. þ. doi: 10.1088/1748-9326/aca41e

Huang, X., Ibrahim, M., Luo, Y., Jiang, L., Chen, J., and Hou, E. (2024). Land use change alters soil organic carbon: constrained global patterns and predictors. *Earth's Future* 12, e2023EF004254. doi: 10.1029/2023ef004254

Khir-Eldien, K., and Zahran, S. A. (2017). "Nile basin climate changes impacts and variabilities," in *The Nile River* (þ: Springer International Publishing), 533–566.

Kome, G. K., Enang, R. K., Tabi, F. O., and Yerima, B. P. K. (2019). Influence of clay minerals on some soil fertility attributes: a review. *Open J. Soil Sci.* 9 (9), 155–188. þ. doi: 10.4236/ojss.2019.99010

Kowalska, A., Singh, B. R., and Grobelak, A. (2022). Carbon footprint for post-mining soils: the dynamic of net CO2 fluxes and SOC sequestration at different soil remediation stages under reforestation. *Energies* 15 (24), 9452. þ. doi: 10.3390/en15249452

Kumar, N., and Sinha, N. K. (2018). "Geostatistics: principles and applications in spatial mapping of soil properties," in *Geospatial Technologies in Land Resources Mapping*, 143–159.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M., and Saby, N. (2017). "Assessment of uncertainty on a digital soil map: a sensitivity analysis on the uncertainty indicators," in *Pedometrics 2017*, 298.

Leenaars, J. G., Claessens, L., Heuvelink, G. B., Hengl, T., González, M. R., van Bussel, L. G., et al. (2018). Mapping rootable depth and root zone plant-available water holding capacity of the soil of Sub-Saharan Africa. *Geoderma* 324, 18–36. doi: 10.1016/j.geoderma.2018.02.046

Legg, S. (2021). IPCC, 2021: climate change 2021-the physical science basis. *Interaction* 49 (4), 44–45. doi: 10.3316/informit.315096509383738

Liao, C., Chang, K., Wu, B., Zhang, D., Wang, C., and Cheng, X. (2024). Divergence in soil particulate and mineral-associated organic carbon reshapes carbon stabilization along an elevational gradient. *Catena* 235, 107682. doi: 10.1016/j.catena.2023.107682

Liu, Y., Meng, Q., Zhang, L., and Wu, C. (2022). NDBSI: a normalized difference bare soil index for remote sensing. *Catena* 214, 106265. doi: 10.1016/j.catena.2022.106265

Lorenz, K., and Lal, R. (2018). "Soil carbon stock," in *Carbon Sequestration in Agricultural Ecosystems*, 39–136.

Ma, Y., Minasny, B., Malone, B. P., and Mcbratney, A. B. (2019). Pedology and digital soil mapping (DSM). *Eur. J. Soil Sci.* 70 (2), 216–235. doi: 10.1111/ejss.12790

Martin, T., and Sprunger, C. D. (2022). Sensitive measures of soil health reveal carbon stability across a management intensity and plant biodiversity gradient. *Front. Soil Sci.* 2, 917885. doi: 10.3389/fsoil.2022.917885

McBratney, A. B., Santos, M. M., and Minasny, B. (2003). On digital soil mapping. *Geoderma* 117 (1-2), 3–52. doi: 10.1016/S0016-7061(03)00223-4

Mikha, M. M., Obour, A. K., and Holman, J. D. (2018). Soil nutrients status after fifty years of tillage and nitrogen fertilization. *Commun. Soil Sci. Plant Analysis* 49 (16), 1953–1975. þ. doi: 10.1080/00103624.2018.1492599

Milà, C., Mateu, J., Pebesma, E., and Meyer, DH. (2022). Nearest neighbour distance matching leave–one–out cross–validation for map validation. *Methods Ecol. Evol.* 13 (6), 1304–1316. doi: 10.1111/2041-210X.13851

Minasny, B., McBratney, A. B., Malone, B. P., and Wheeler, I. (2013). Digital mapping of soil carbon. *Adv. Agronomy*, 118, 1–47.þ doi: 10.1016/B978-0-12-405942-9.00001-3

Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S., and Mukhopadhyay, A. (2017). Spatial soil organic carbon prediction by regression kriging. *Egypt. J. Remote Sens. Space Sci.* 20 (1), 61–70. doi: 10.1016/j.ejrs.2016.06.004

Negassa, M. K., Haile, M., Feyisa, G. L., Wogi, L., and Liben, F. M. (2023). Soil organic carbon stock prediction under 2050 climate scenarios. *Sustainability* 15 (8), 6495. doi: 10.3390/su15086495

Negassa, M. K., Haile, M., Feyisa, G. L., Wogi, L., and Merga, F. (2025). Modeling spatial distribution of soil organic carbon stock. *Geol. Ecol. Landscapes* 9 (1), 106–121. doi: 10.1080/24749508.2023.2167632

Nelson, D. W., and Sommers, L. E. (1996). "Total carbon, organic carbon, and organic matter," in *Methods of Soil Analysis*, 961–1010.

Obi, L. U., Olisaka, F. N., Ene, C., and Aniakor, U. (2024). "Exploring biotechnological strategies in soil quality monitoring," in *The Marginal Soils of Africa* (Springer), 173–204.

Odebiri, O., Mutanga, O., Odindi, J., Slotow, R., Mafongoya, P., Lottering, R., et al. (2025). Mapping sub-surface distribution of soil organic carbon stocks in South Africa's arid and semi-arid landscapes: implications for land management and climate change mitigation. *Geoderma Reg.*, 43, e01022.þ doi: 10.1016/j.geodrs.2025.e01022

Owusu, S., Yigini, Y., Olmedo, G. F., and Omuto, C. T. (2020). Spatial prediction of soil organic carbon stocks in Ghana. *Geoderma* 360, 114008. doi: 10.1016/j.geoderma.2019.114008

Oyeleke, O. R. (2024). *Revolutionizing Nigerian Agriculture: Leveraging Digital Soil Mapping for Sustainable Productivity and Food Security*þ: Iowa State University.

Padarian, J., Minasny, B., and McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil* 5 (1), 79–89. doi: 10.5194/soil-5-79-2019

Pal, D. K., Wani, S. P., and Sahrawat, K. L. (2012). Vertisols of tropical Indian environments. *Geoderma* 189, 28–49. doi: 10.1016/j.geoderma.2012.04.021

Papadopoulou, I., Karampini, C., Mingou, L., Arroyo-Cerezo, A., Moreno-Cuenca, L., Kalogeras, A., et al. (2026). Sustainable and precision viticulture: systematic insights from soil and remote sensing studies. *Agriculture* 16 (13), 1370. doi: 10.3390/agriculture16131370

Pebesma, E., and Gräler, B. (2015). *gstat: Spatial and Spatio-Temporal Geostatistical Modelling. R package version 1.1-0. Vienna, Austria: R Foundation for Statistical Computing*. Available online at: https://CRAN.R-project.org/package=gstat.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40 (8), 913–929. doi: 10.1111/ecog.02881

Robinson, T. P., and Metternicht, G. (2006). Testing spatial interpolation techniques. *Comput. Electron. Agric.* 50 (2), 97–108. doi: 10.1016/j.compag.2005.07.003

Saiz, G., Bird, M. I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T. R., et al. (2012). Variation in soil carbon stocks and their determinants across a precipitation gradient in W est A frica. *Glob. Change Biology* 18 (5), 1670–1683. doi: 10.1111/j.1365-2486.2012.02657.x

Sanderman, J., Hengl, T., and Fiske, G. (2017). Soil carbon debt of 12,000 years of human land use. *Proc. Natl. Acad. Sci.* 114, 9575–9580. doi: 10.1073/pnas.1706103114

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappie, M., Märker, M., et al. (2017). Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: the role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Sci. Total Environment*, 601, 821–832.þ doi: 10.1016/j.scitotenv.2017.05.239

Schmidt, F., Wagner, N., Mulder, I., Emmerich, K., Bore, T., and Bumberger, J. (2025). Broadband dielectric analysis of clays: impact of cation exchange capacity, water content, and porosity. *J. Geophys. Res. Solid Earth* 131, e2025JB031945. doi: 10.1029/2025jb031945

Soil Survey Staff (2014a). "Keys to soil taxonomy," in *USDA. Soil Analysis: Chemical Methods*. 12th ed. Wiley.

Sparks, D. L., Page, A. L., Helmke, P. A., and Loeppert, R. H. (2020). *Methods of Soil Analysis, Part 3: Chemical Methods* (John Wiley & Sons).

Sufardi, S., Arabia, T., Khairullah, K., Zainabun, Z., Karnilawati, K., and Sahbudin, S. (2020). Charge characteristics and cation exchanges properties of hilly dryland soils Aceh Besar, Indonesia. *Aceh Int. J. Sci. & Technol.* 9 (2), 90–101. þ. doi: 10.13170/aijst.9.2.17565

Sulieman, M., Saeed, I., Hassaballa, A., and Rodrigo-Comino, J. (2018). Modeling cation exchange capacity. *Catena* 167, 327–339. doi: 10.1016/j.catena.2018.05.001

Takoutsing, B., and Heuvelink, G. B. (2022). Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. *Geoderma* 428, 116192. doi: 10.1016/j.geoderma.2022.116192

Tantarawongsa, P., Shen, C., Ketrot, D., Suttanukool, P., Thupwong, K., and Mekboonsonglarp, W. (2024). Chemical composition of organic carbon in tropical soils under agricultural cultivation and peat-swamp forest. *Commun. Soil Sci. Plant Analysis* 55 (4), 473–487. doi: 10.1080/00103624.2023.2274024

Tilahun, E., Haile, M., Gebresamuel, G., and Zeleke, G. (2022). SOC dynamics under land-use change. *Geoderma* 406, 115506. doi: 10.1016/j.geoderma.2021.115506

Tran, D. X., Dominati, E., Lowry, J., Mackay, A., Vibart, R., Pearson, D., et al. (2024). Effects of spatial data resolution on the modelling and mapping of soil organic carbon content in hill country grassland landscapes. *Soil Use Manag.* 40 (1), e12966. doi: 10.1111/sum.12966

Tsozué, D., Noubissie, N. M. M., Mamdem, E. L. T., Basga, S. D., and Oyono, D. L. B. (2021). Effects of environmental factors and soil properties on soil organic carbon stock in a natural dry tropical area of Cameroon. *Soil* 7 (2), 677–691. doi: 10.5194/soil-7-677-2021

Tziachris, P., Aschonitis, V., Chatzistathis, T., and Papadopoulou, M. (2019). Hybrid methods for predicting soil organic matter. *Catena* 174, 206–216. doi: 10.1016/j.catena.2018.11.010

Wendt, J. W., and Hauser, S. (2013). An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. *Eur. J. Soil Sci.* 64 (1), 58–65. doi: 10.1111/ejss.12002

Wiesmeier, M., Barthold, F., Blank, B., and Kögel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. *Plant Soil* 340 (1), 7–24. doi: 10.1007/s11104-010-0425-z

Williams, A., Dalal, R., and Hedlund, K. (2022). Decline in soil fertility and biological diversity, in soil constraints on crop production, Y. Dang, N. Menzies, and R. Dalal (eds) (Newcastle upon Tyne, United Kingdom: Cambridge Scholars Publishing). 267–292.

Yigini, Y., Olmedo, G. F., Reiter, S., Baritz, R., Viatkin, K., and Vargas, R. (2018). *Soil Organic Carbon Mapping Cookbook* 2nd ed. (Rome: Food and Agriculture Organization of the United Nations), 204.

Zeraatpisheh, M., Galford, G. L., White, A., Noel, A., Darby, H., and Adair, E. C. (2023). Soil organic carbon stock prediction. *Catena* 229, 107197. doi: 10.1016/j.catena.2023.107197

Zheng, J. Y., Zhao, J. S., Shi, Z. H., and Wang, L. (2021). Soil aggregates are key factors that regulate erosion-related carbon loss in citrus orchards of southern China: bare land vs. grass-covered land. *Agric. Ecosyst. & Environ.* 309, 107254. doi: 10.1016/j.agee.2020.107254

Zhong, Z., Chen, Z., Xu, Y., Ren, C., Yang, G., Han, X., et al. (2018). SOC stocks and clay content relationship. *Forests* 9 (10), 598. doi: 10.3390/f9100598

Zhu, C., Wei, Y., Zhu, F., Lu, W., Fang, Z., Li, Z., et al. (2022). Digital mapping of soil organic carbon. *Sensors* 22 (22), 8997. doi: 10.3390/s22228997