

RESEARCH

Open Access



# Soil quality estimation using environmental covariates and predictive models: an example from tropical soils of Nigeria

Isong Abraham Isong<sup>1\*</sup> , Kingsley John<sup>2</sup>, Paul Bassey Okon<sup>1</sup>, Peter Ikor Ogban<sup>1</sup> and Sunday Marcus Afu<sup>1</sup>

## Abstract

**Background:** Information addressing soil quality in developing countries often depends on results from small experimental plots, which are later extrapolated to vast areas of agricultural land. This approach often results in misinformation to end-users of land for sustainable soil nutrient management. The objective of this study was to estimate the spatial variability of soil quality index (SQI) at regional scale with predictive models using soil–environmental covariates.

**Methods:** A total of 110 composite soil samples (0–30 cm depth) were collected by stratified random sampling schemes at 2–5 km intervals across the Cross River State, Nigeria, and selected soil physical and chemical properties were determined. We employed environmental covariates derived from a digital elevation model (DEM) and Sentinel-2 imageries for our modelling regime. We measured soil quality using two approaches [total data set (TDS) and minimum data set (MDS)]. Two scoring functions were also applied, linear (L) and non-linear (NL), yielding four indices (MDS\_L, MDS\_NL, TDS\_L, and TDS\_NL). Eleven soil quality indicators were used as TDS and were further screened for MDS using principal component analysis (PCA). Random forest (RF), support vector regression (SVR), regression kriging (RK), Cubist regression, and geographically weighted regression (GWR) were applied to predict SQI in unsampled locations.

**Results:** The computed SQI via MDS\_L was classified into five classes:  $\leq 0.38$ , 0.38–0.48, 0.48–0.58, 0.58–0.68, and  $\geq 0.68$ , representing very low (class V), low (class IV), moderate (class III), high (class II) and very high (class I) soil quality, respectively. GWR model was robust in predicting soil quality ( $R^2 = 0.21$ , CCC = 0.39, RMSE = 0.15), while RF was a model with inferior performance ( $R^2 = 0.02$ , CCC = 0.32, RMSE = 0.15). Soil quality was high in the southern region and low in the northern region. High soil quality class (> 49%) and moderate soil quality class (> 14%) dominate the study area in all predicted models used.

**Conclusions:** Structural stability index, sand content, soil organic carbon content, and mean weight diameter of aggregates were the parameters used in establishing regional soil quality indices, while land surface water index, Sentinel-2 near-infrared band, plane curvature, and clay index were the most important variables affecting soil quality variability. The MDS\_L and GWR are effective and useful models to identify the key soil properties for assessing soil quality, which can provide guidance for site-specific management of soils developed on diverse parent materials.

**Keywords:** Decision support system, Geospatial technology, Predictive model, Remote sensing, Soil quality

\*Correspondence: aisong247@gmail.com

<sup>1</sup> Department of Soil Science, Faculty of Agriculture, University of Calabar, Calabar PMB 1115, Nigeria  
Full list of author information is available at the end of the article

## Introduction

Knowing soil quality is essential when considering land degradation assessment, soil management, crop production, and food security. Soil quality is a prerequisite for better planning and utilization of land resources (Amalu and Isong 2017b; Okon et al. 2019; Kalambukattu et al. 2018). A decline in soil quality interrupts primary soil functions and may hamper crop production and food security. However, soil quality evaluation is necessary for identifying areas with corresponding high and low soil quality, and their suitability for agricultural land use in general and cultivated crops in particular. This would provide valuable information on the possibility of soil degradation and nutrient mining to farmers, land managers, and policymakers to make sustainable land management decisions. Several models, including fuzzy set techniques (Rezaee et al. 2020), Nemoro soil quality index (Nabiollahi et al. 2017), simple additive soil quality index (Mukherjee and Lal 2014), weighted additive soil quality index (Vasu et al. 2016), among others, have been developed and applied for the estimation of soil quality index.

Inadequate land use planning and soil suitability assessment in most developing countries have become a constraining factor for crop production. For example, both plantation and arable lands in Nigeria were established primarily to forestall land encroachment by land speculators and landlords of communities without resorting to adequate land use planning (Amalu and Isong 2015). As such, no feasibility studies regarding soil quality were usually carried out to determine the suitability of soils for crop cultivation at the time of establishment. However, wider crop yield differences observed under similar management practices across these farms call for a concerted effort to understand the soil quality in the region, with the primary aim of managing such soils responsibly. This will guarantee their ever-indisputable usefulness and services to humanity.

It is worth noting that soil is not directly consumed compared to air and water; hence, estimating its quality is quite difficult (Debi et al., 2019). In addition, soil quality does not depend on a single factor but the integration of physical, chemical, and biological factors for its quantification through soil quality index (Shekhovtseva and Mal'tseva 2015). Therefore, the initial approach to understanding and evaluating the soil quality of a given land under cultivation is to estimate the soil quality index using soil quality indicators sensitive to changes in soil management practices. The soil quality indices are models that provide numerical data concerning the capacity of soil to carry out one or more functions (Asensio et al. 2013).

The choice of soil quality indicators (i.e., physical, chemical, and biological properties) in soil quality estimation depends on their sensitivity to cause a change in soil function and financial budget. However, this study applied soil physical parameters and some selected chemical and biological properties, as they could serve as a proxy to reveal to a greater extent, soil quality in sub-Saharan Africa. Soil physical properties have been reported in several studies as having an overarching influence in controlling chemical and microbial soil properties (Dexter 2004; Igwe et al. 2013; Pulido-Moncada et al. 2015; Phogat et al. 2015; Jat et al. 2018). Hence, the physical properties of soils require careful monitoring as they strongly affect soil water movement, nutrient absorption, nutrient mining, and solute and pollutant movement (Jat et al. 2018). They also control plant nutrient storage, soil aggregation, structural development, leaching and erosion potential, the energy balance of the soil–plant system and pedogenesis, soil organic matter stabilization, and optimum plant development (Dexter 2004; Carrizo et al. 2015).

Evaluating soil quality requires a large number sampling points, and obtaining a large number of samples to quantify soil quality over large areas through conventional soil survey methods is often tedious, expensive, and time-consuming. The adventure relies on the experience of the pedologist and would not provide sufficiently detailed information about the soil variation required for many environmental applications. Thus, an alternative approach, the digital soil mapping (DSM) technique, can overcome this problem by predicting soil quality index or soil quality classes utilizing soil properties, remote-sensing data, digital elevation model (DEM), micro-climatic data, land use and cover data, and geological data (Nabiollahi et al. 2018a, b; John et al. 2021b, c, d; Zeraatpisheh et al. 2020) as covariates or ancillary variables aided by geostatistics (e.g. John et al. 2022a, 2022b) and machine learning (e.g. John et al. 2020, 2021a) for unsampled areas.

The machine learning and geostatistical models for predicting soil properties in sub-Saharan Africa, particularly in Nigeria, have already been used (Ogunwole et al. 2014). In addition, in West Africa as a whole, Hengl et al. (2015) expanded the idea and scope, and in the Africa Soil Information Services project (AfsIS) (Hengl et al. 2017), these tools were employed to modeled the spatial distribution of selected soil nutrient indicators but at a coarser scale. Since then, many sub-Saharan African countries have applied similar methods to produce detailed maps of soil nutrients at different scales. Despite the acceptability and utilization of machine learning and geostatistics in soil nutrient mapping, only a few global studies utilize a similar approach to model soil quality

(Nabiollahi et al. 2018a; Paul et al. 2020; Zeraatpisheh et al. 2020). However, no feasible study has been carried out elucidating this approach in sub-Saharan Africa despite the region's active engagement in food crop production.

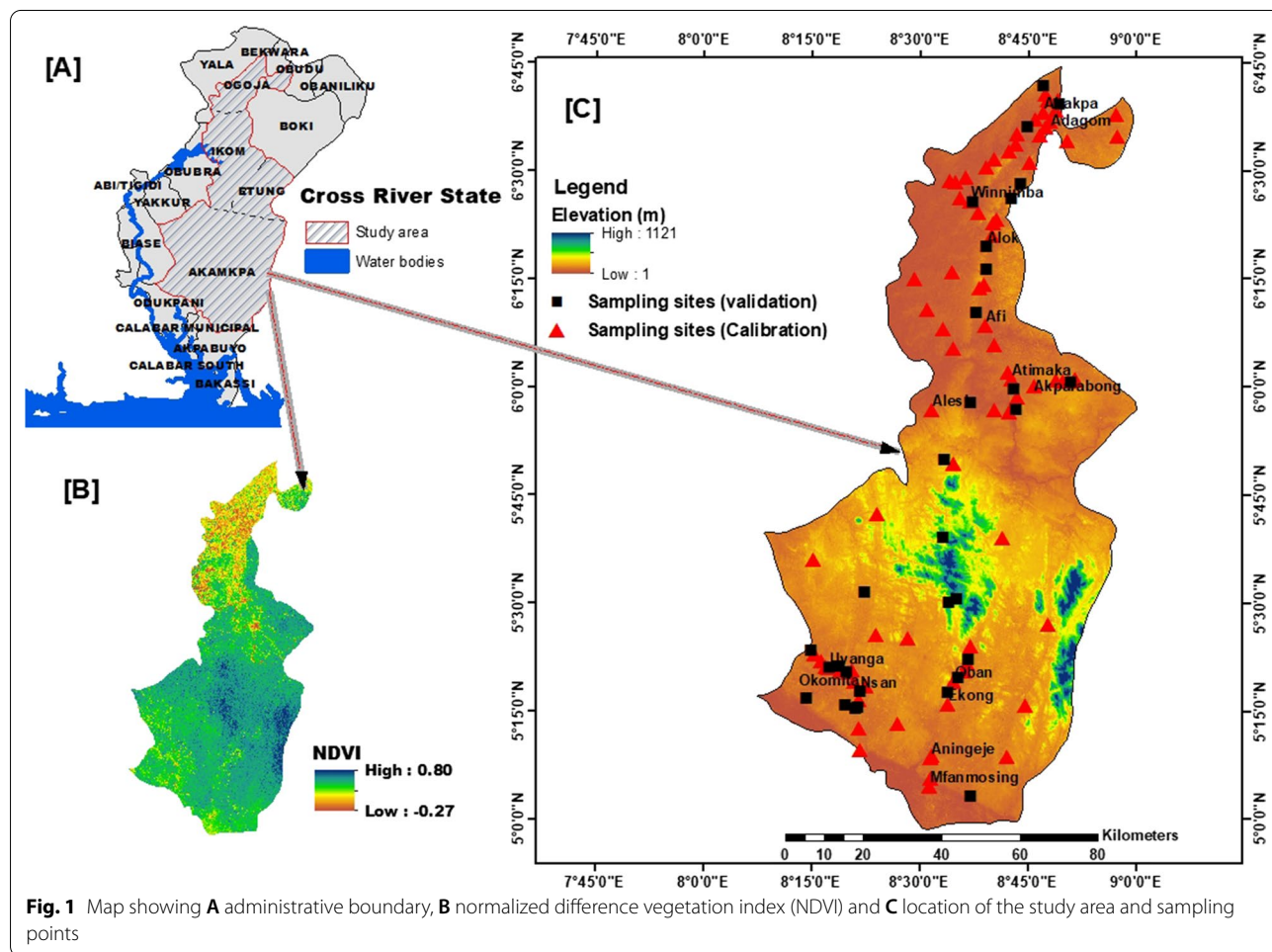
Nevertheless, there is now a growing interest in applying machine learning and geostatistical techniques to produce detailed maps of soil quality for the zone to monitor soil resources and support crop production, given the threat posed by land degradation in African farming systems. Therefore, information and knowledge on soil quality are keys to guiding management and restorative measures and decisions by farmers and other land managers. The information obtained could help curb inappropriate land use and soil management, as they could lead to the deterioration of soil quality, threatening crop production, food security, economic growth, and a healthy environment. This study is designed to assist farmers, land managers, and policymakers in supporting decision-making about sustainable cropland management in Cross River State, Nigeria. We hypothesize that parent

materials and environmental factors strongly influence soil quality in tropical soils. The major drivers of spatial variability in soil quality index are remote-sensing-based variables and topographic data. At the same time, the study aimed to estimate soil quality index and map their spatial variability using machine learning and geostatistical models.

### Materials and methods

#### Research location

The present study was conducted in Cross River State, situated in southeastern Nigeria. The study location (Fig. 1C) has an area of 9456.09 km<sup>2</sup>, geographically bounded between latitudes 5° 00'N–6° 40'N and longitudes 7° 55'E–9° 10'E. We divided the area of investigation into the northern, central, and southern zones corresponding to the Ogoja, Ikom, and Akamkpa Local Government Areas. The site is characterized by a diverse geological and physiographic setup with rugged topography (ranging from 1 m to nearly 1121 m) and wider variability in climate and normalized difference vegetation



index (NDVI) (-0.27 to 0.80) (Fig. 1B) and is characterized by udic moisture and isohyperthermic soil temperature regimes (Soil Survey Staff 2014). The main geology/parent materials underlying soils in the study area are basement complex rocks known as Oban Massif, limestone, basalt, shale, and sandstone (Ofomata 1975; Eshett et al. 1990; Aki et al. 2014). The predominant mineralogy consists of kaolinite, quartz, Fe and Al oxides, montmorillonite, microcline, and hematite (Njar 2018). In general, the study area is characterized by a tropical wet or monsoon climate according to the Koppen–Geiger climate classification (Beck et al. 2018). The site has two distinct seasons: the rainy season, which starts in April and ends in early November with a double peak, usually in July and September, and the dry season, which spans from November to March. The mean annual rainfall is usually higher in the southern part of the state and lowest in the northern part, while temperature increases from the southern zone to the northern zone. The average annual rainfall across the study area exceeds 2000 mm, while the average minimum and maximum temperatures are about 22 °C and 30 °C, respectively, with a mean relative humidity of 83%. The predominant agricultural land use in the region is plain cropland, shrubland, grassland, plantation, and wetland. The principal crops grown in the area include maize, sugar cane, rice, cassava, yam, groundnut, oil palm, cocoa, rubber, and vegetable crops (i.e., okra, *Telfairia occidentalis*, pepper, water leaf, *Amarathus cruentus*, etc.). In terms of natural vegetation, the southern region is characterized by tropical rainforests. In the central, the vegetation is the transition between tropical rainforest and guinea savannah, while derived savannah is in the northern region.

The farming systems in the study area are fallow, shifting cultivation, intercropping, multistorey cropping system, relay cropping, and sequential cropping, where crops were cultivated on three basic types of agricultural land (compound land, family land, and community land). This is similar to what is obtained in other sub-Saharan African countries (Callo-Concha et al. 2012). As the need to sustainably meet the increasing demand for food intensified, the emphasis shifted to large-scale production on privatized estates, government-owned plantations/farms, and commercial farms. The soils are managed via organic and inorganic fertilization, with minimal use of herbicides and insecticides to control weeds and pests. However, the increasing and continuous utilization of land resources in sub-Saharan Africa without prior knowledge of their soil quality may pose a serious challenge to achieving food security and sustainable development goals by 2030. In a bid to provide a solution to the pressing issues, this study was undertaken

in response to the region's urgent need to produce a soil quality map, given the threat posed by land degradation in farming systems. This paper's idea will help achieve this objective by providing more insights on environmental factors exerting significant influence on soil quality in the study area through a satisfactory, rapid, sustainable, and low-cost approach.

### Soil sampling

A total of 110 soil samples were collected in stratified random schemes to cover the entire study area. On each sampling point, a composite of three sub-samples (0–30 cm soil depth) was collected randomly within the grid area, hand-mixed, and placed in labeled plastic bags. The exact locations of the sampling points were determined using Global Positioning System (GPS Model eTrex Legend H) receiver.

### Laboratory methods

All collected soil samples were transported to the laboratory. The soil samples were air-dried, crushed to pass through a 2 mm sieve, bagged, labeled, and subjected to laboratory analysis, except for samples for aggregate stability analysis. Samples for the determination of aggregate stability were collected using a spade, while those for the determination of hydraulic conductivity, bulk density, and total porosity were collected using a core sampler. The particle size analysis was determined using the Bouyocous hydrometer method after the mixture was dispersed with distilled water and Calgon (Gee and Or 2002). Bulk density ( $\rho_b$ ) was determined using the core method as described by Grossman and Reinsch (2002). Particle density was calculated by the pycnometer method following the procedure outlined by Blake (1965). The total porosity was calculated from the particle and bulk densities using the relationship established by Vomocil (1965). The gravimetric method of determining moisture content was used by Gardner (1986). The saturated hydraulic conductivity of the soil was determined using the constant head method according to the procedure described by Klute and Dirksen (1986). The wet sieving method of aggregate analysis was used following the procedure of Angers and Mehuys (1993) using a stack of sieves with 2.0, 1.0, 0.5, and 0.25 mm openings. Soil pH was determined following the procedure described by Udo et al. (2009) using a pH meter. Soil organic carbon was determined by the Walkley and Black wet oxidation method (Nelson and Sommers 1996). Soil organic matter was then calculated by a factor of 1.724 (Van Bemmelen's Correction Factor). The structural stability index (SSI) was calculated from soil organic carbon (SOC) and fine soil texture components following Castellini et al. (2016) procedure:



**Table 1** Environmental covariates initially considered for soil quality prediction

Environmental covariates	Variable	Description
DEM	AHShad	Analytical hillshade
	Aspect	Aspect
	CNBL	Cannel network base level
	CIndex	Convergence index
	Elv (m)	Elevation
	LS (m)	Slope length factor
	Plcurv	Plane curvature
	Prcurv	Profile curvature
	Slope (degree)	Land surface slope
	TCA	Total catchment area
	TWI	Topographic wetness index
	VDepth	Valley depth: The relative height difference to the immediate adjacent channel lines
	Sentinel-2	B2
B3		0.54–0.58 μm
B4		0.65–0.68 μm
B5		0.70–0.71 μm
B6		0.73 – 0.7 μm
B7		0.77 – 0.79 μm
B8		0.78 – 0.90 μm
B8A		0.85 – 0.87 μm
B11		1.56 – 1.65 μm
B12		2.10 – 2.28 μm
Carbonate index (CBI)		$CBI = \frac{RED}{GREEN}$
Iron index (Fe)		$Fe = \frac{RED}{BLUE}$
Clay index (CI)		$CI = \frac{SWIR1}{SWIR2}$
Normalized Difference Build-up Index (NDBI)		$NDBI = \frac{(SWIR - NIR)}{(SWIR + NIR)}$
Ratio Vegetation Index (RVI)		$RVI = \frac{NIR}{RED}$
Modified Soil Adjusted Vegetation Index (MSAVI)		$MSAVI = \frac{2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - RED)}}{2}$
Normalized Difference Vegetation Index (NDVI)		$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{(NIR - GREEN)}{(NIR + GREEN)}$	
Land Surface Water Index (LSWI)	$LSWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$	

Sentinel-2 (B2: blue, B3: green, B4: red, B5: red edge, B8: NIR, B8A: NIR, SWIR1: 11, SWIR2: 12); DEM digital elevation model

$$SSI = 1.724 \times \frac{SOC (\%)}{(silt\% + clay\%)} \times 100 \tag{1}$$

**Environmental data**

Table 1 presents the digital elevation model (DEM) and its derivatives. The DEM was obtained from ASTER data (<https://earthexplorer.usgs.gov>) at the spatial resolution of 30 m and processed with the System for Automated Geoscientific Analysis and Geographical Information System (SAGA–GIS).

Remote sensing data (Cloud-free Sentinel-2 imageries) were acquired at the European Space Agency’s Copernicus Open Access Hub (<https://scihub.copernicus.eu/>) in level 1c tiles which produce the type S2MSI1c imageries used for the study. The images of Sentinel-2, level 1c, are at TOA (top of atmosphere reflectance). Therefore, in the pre-processing step, the images were subject to atmospheric corrections to obtain Sentinel-2 level 2a at BOA (bottom-of-atmosphere reflectance). The processing was performed using the Sen2Cor module in the SNAP tool. These soil–environmental covariates were selected due

to their proven correlation with soil properties (Campos et al. 2018) and are reported in Table 1.

**Soil quality computation procedure**

In this study, the soil quality index (SQI) was computed using three methods (Karlen et al. 2003; Andrews et al. 2004). The methods involved the selection of soil quality indicators using both total data set (TDS) and minimum data set (MDS), indicator transformation, and indicator integration into the overall index.

Minimum data sets (MDS) selection: principal components analysis (PCA) was used as a method of MDS selection. In using PCA, only the principal components (PCs) with eigenvalues  $\geq 1$ , which explained at least 5% variation of the data, were retained for interpretation. Andrews and Carroll (2001) suggested that indicators with weighted absolute values within 10% of the highest indicator value for each PC should be selected as the MDS. However, this criterion only accounted for the loading of the variable to a single PC and did not provide information for the variable on a multi-dimensional space; hence norm values, as suggested by Yemefack et al. (2006), were used for grouping and selection of variables as MDS.

The norm value was calculated using the following equation:

$$N_{ik} = \sqrt{\sum_{i=1}^k (u_{ik}^2 \lambda_{ik})} \tag{2}$$

where  $N_{ik}$  = load for  $i$ th soil property on PCs with eigenvalues  $\geq 1$ ;  $u_{ik}$  = load for the  $i$ th soil property on the principal component of  $k$ ;  $\lambda_{ik}$  = the eigenvalue of the  $i$ th soil property on the principal component of  $k$ .

When more than one soil property within a PC fulfills the selection criteria, the multivariate correlation matrix was used to determine the correlation between them, and the non-correlated parameters ( $r < 0.60$ ) under a particular PC were considered important and retained (Andrews et al. 2002). Conversely, among correlated variables within a PC, the variable with the highest correlation sum was selected for the MDS.

Indicator transformation: since each soil indicator has different units, they were transformed and normalized into a unitless score between 0 and 1 using both linear and non-linear scoring approaches before final integration into the overall soil quality index (SQI). Three established soil scoring functions (SSFs) based on if it has a "negative" or "positive" relationship with soil quality or if it is positively or negatively related within an "optimum range" (Li et al. 2018) were used. This research presents the linear and non-linear functions (Eqs. 3–6) used

below. Equations (3), (4), and (5) are linear functions and correspond to "more is better" function ( $M$ ), "less is better" function ( $L$ ), and "optimal range" function ( $O$ ), respectively:

$$M = \begin{cases} 0.1 & x \leq L \\ 0.1 + 0.9 \times \left(\frac{x-L}{U-L}\right) & L \leq x \leq U \\ 1 & x \geq U \end{cases} \tag{3}$$

$$L = \begin{cases} 1 & x \leq L \\ 1 - 0.9 \times \left(\frac{x-L}{U-L}\right) & L \leq x \leq U \\ 0.1 & x \geq U \end{cases} \tag{4}$$

$$O = \begin{cases} 0.1 & x < L_1 \text{ or } x \geq U_2 \\ 0.1 + 0.9 \times \left(\frac{x-L_1}{L_2-L_1}\right) & L_1 \leq x \leq L_2 \\ 1 & L_2 \leq x \leq U_1 \\ 0.1 + 0.9 \times \left(\frac{x-U_1}{U_2-U_1}\right) & U_1 \leq x \leq U_2 \end{cases} \tag{5}$$

where  $x$  = measured value of the soil quality indicator.  $M$ ,  $L$ , and  $O$  = values of the variables after transformation for "more is better", "less is better" and "optimal range" scoring functions.  $L$  and  $U$  = the lower and the upper threshold values of soil indicators, respectively.

Similarly, Eq. (6) depicts a non-linear function, where  $x$  is the soil property value,  $b$  is the slope assumed to be  $-2.5$  for "more is better" and  $+2.5$  for "less is better", and  $x_o$  is the mean value of the soil variable:

$$S_{NL} = \frac{1}{\left[1 + (x/x_o)^b\right]} \tag{6}$$

Indicator integration into the overall index: the soil quality index (SQI) of each sampling site was calculated using the integrated quality index (IQI) (Eq. 7):

$$IQI = \sum_{i=1}^n W_i S_i \tag{7}$$

where IQI is the weighted additive soil quality index,  $n$  is the number of selected soil properties,  $S_i$  is the score of soil properties  $i$ ,  $W_i$  is the assigned weight of each soil property for the TDS and MDS based on the communality of principal component analysis (PCA). The weights were computed as

$$W_i = \frac{C_i}{\sum_{i=1}^n C_i} \tag{8}$$

where  $i$ th is the soil variable, and  $C_i$  is the communality value of a soil variable  $i$ th.

Soil quality classification scheme: according to the classification criteria in Guo et al. (2017) "an ideal soil would have SQI value of 1 for the highest quality soil and 0 for the severely degraded soil, soil quality was divided into five grades: very high, high, moderate, low, and very low". Chen et al. (2013) suggested using "values 10% more than or less than the average as Grade III in classifying soil quality, and other grades can be derived as the increments or decrements of 20% from the adjacent grades". According to the authors, "the quality of soil decreased as the grade increases". Thus, Grade I soil should be considered suitable for plant growth; Grade V soil is characterized by having the most severe limitations for plant growth. According to Li et al. (2018), "Grade I is described as having a very high value, which is the most suitable for crop growth; Grade II showed a high IQI value and is suitable for crop growth; Grade III had a moderate value and had some limitation; Grade IV showed a lower IQI value and had more limitations than Grade III; grade V was characterized as having a very low IQI value and the most severe limitations". Soil quality grade was mapped with the aid of ArcGIS software.

Soil quality index validation: the proposed models' performance was validated using the sensitivity index (SI), as presented in the following equation:

$$SI = \frac{SQI_{max}}{SQI_{min}} \tag{9}$$

where  $SQI_{max}$  = maximum soil quality index value, and  $SQI_{min}$  = minimum soil quality index value

In addition, the efficiency ratio (ER) was computed to assess the efficiency of each SQI in evaluating the soil quality index (Eq. 10).

$$ER = (K/N) \times 100 \tag{10}$$

where  $K$  is the number of significant paired correlations between the specific SQI and all indicators;  $N$  is the number of all feasible paired correlations in the data set, which equals 4 and 11 for MDS and TDS in this research, respectively.

**Predictive models**

I. Multiple linear regression (MLR) and geographically weighted regression (GWR): MLR and GWR were utilized to model the relationship between SQI and selected predictors (i.e., soil–environmental covariates).

In MLR, the target soil property is modelled as a linear combination of predictors. It is a global model and assumes variables' independence, stationarity, and isotropy as the precondition but overlooks the spatial heterogeneity of the response variable and its auxiliary

information. Thus, the SQI of interest is predicted by the simple formula:

$$\hat{y}(i) = \hat{\beta}_0 + \sum_{k=1}^k \hat{\beta}_k X_{k(i)} \tag{11}$$

where  $\hat{y}(i)$  is the predicted SQI at location  $i$ ,  $\hat{\beta}_0$  is the estimated intercept,  $\hat{\beta}_k$  is the estimated regression coefficient for predictor  $k$ , and  $X_k(i)$  is the value for the  $k$ th predictor at location  $i$ . The regression coefficients are estimated by ordinary least squares (OLS).

The GWR model is a local form of linear regression that overcomes the limitation of the MLR model. GWR is a spatial regression model applied in DSM. The model is based on the local smoothing idea that considers the spatial locations of samples and uses the locally weighted least square method to model the observations of soil parameters. It can be written as

$$\hat{y}_i = \beta_0(u_i, v_i) + \sum_{K=1}^P \beta_k(u_i, v_i) x^*_{ik} + \varepsilon_i \tag{12}$$

where  $(u_i, v_i)$  are the coordinates of point  $i$ ,  $\beta_0(u_i, v_i)$  is the coefficient of different explanatory variables,  $x^*_{ik}$  is the value of explanatory variable  $k$  at point  $i$ ,  $p$  is the total number of explanatory variables,  $\varepsilon_i$  is the error term that is generally assumed to be explanatory and normally distributed with zero mean and constant variance, and the values of the above parameters vary with the location. The parameters can be estimated using a weighting function as in the following equation:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \tag{13}$$

where  $X$  is the matrix formed by  $x^*_{ik}$ ,  $Y$  is the vector formed by values of the response variable, and  $W(u_i, v_i)$  is a weight matrix to ensure that those observations near the point  $i$  have more influence on the results than those that are farther away. The parameters of the GWR are Kernel type and bandwidth selection criteria (AICC).

ii. Regression Kriging (RK): RK combines OK and a linear regression model. It combines a regression of dependent variables on predictors with kriging of the prediction residuals. RK has been proposed in several studies as a way to account for spatial autocorrelation in regression modelling (Fayad et al. 2016). For RK, a prediction at an unvisited site is given by summing the regression prediction and the kriging prediction of the regression residual.

iii. Random Forest (RF) model: Random Forest (RF) modeling has become a popular technique for regression and classification with complex environmental data sets (Freeman et al. 2015; Fox et al. 2020). In contrast to multiple regression, RF is an algorithmic procedure that makes no a priori assumptions about the relationship

between the predictor variables and the response. RF has a reputation for good predictive performance when the data contain many predictor variables, complex non-linearities, and interaction effects in the relationship between the predictors and response variables (Biau 2012; John et al. 2020). In addition, RF provides several measures of variable importance that allow the interpretation of the fitted model (Hastie et al. 2009).

iv. Support vector regression (SVR): support vector regression (SVR) is a supervised learning method that has recently gained popularity for predicting soil properties (Lamichhane et al. 2019). SVR is an extension of SVM and is used as a regression technique. This technique generates an optimal separating hyperplane to differentiate classes that overlap and are not separable in a linear way. In this case, a large, transformed feature space is created to map the data with the help of kernel functions to separate it along a linear boundary. More detailed explanations about SVR can be found in (Breiman 2001). The optimal function developed by SVR can be expressed as

$$f(x) = \sum_{k=1}^P (\alpha_k - \alpha_k^*) K(x_k, x_j) + b \tag{14}$$

where  $x$  is a vector of the input predictors (environmental variables),  $f(x)$  is an optimal function developed by SVR,  $b$  is a constant threshold,  $K(x_k, x_j)$  is the Gaussian radial basis kernel function with the best bandwidth parameter  $\sigma$  and  $\alpha_k$  and  $\alpha_k^*$  are the weights (Lagrange multipliers) with the constraints given in the following equation:

$$\begin{cases} \sum_{k=1}^P (\alpha_k - \alpha_k^*) = 0 \\ 0 \leq \alpha_k, \alpha_k^* \leq C \end{cases} \tag{15}$$

v. Cubist regression: the Cubist model was developed by Quinlan (1992) as a rule-based model, which is an extension of the M5 tree model. According to Kuhn and Johnson (2013), the model structure consists of a conditional component or piecewise function acting as a decision tree, coupled with multiple linear regression models. The trees are reduced to a set of rules eliminated via pruning or combined for simplification. The Cubist method's main benefit is adding multiple training committees and boosting to make the weights more balanced (Quinlan 1992; Wang 1997; Kuhn and Johnson 2013). The Cubist model adds boosting with training committees (usually greater than one) which is similar to the method of "boosting" by sequentially developing a series of trees with adjusted weights. The number of neighbours in the Cubist model is applied to amend the rule-based prediction (Kuhn and Johnson 2013). This model will

be implemented in R by tuning two hyper-parameters: neighbours (Instances) and committees (Committees). These two parameters are the most likely to have the largest effect on the final performance of the Cubist model.

**Modelling approach and evaluation**

Covariates used in building a more efficient soil quality model were selected via stepwise regression with forwarding selection and backward elimination of predictors following the procedure of Zounemat-Kermani et al. (2020). The data set was randomly calibrated into two with a 75% and 25% ratio, respectively. 75% of the data set was used for training the model, while 25% was used for validation. The following metrics were used, coefficient of determination ( $R^2$ ), root mean square error (RMSE), and Lin's concordance correlation coefficient (CCC) was used to compare and select the best model. The formulas are given as follows:

$$R^2 = 1 - \frac{\sum_i^n (Z_{oi} - Z_{pi})^2}{\sum_i^n (Z_{oi} - \bar{Z}_{pi})^2} \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_{pi} - Z_{oi})^2} \tag{17}$$

$$CCC = \frac{2r\sigma_o\sigma_p}{\sigma_o^2 + \sigma_p^2 + (\bar{Z}_p - \bar{Z}_o)} \tag{18}$$

where  $Z_{pi}$ =predicted values,  $Z_{oi}$ =observed values,  $n$ =the size of the observations, for the  $i$ th term observation,  $\bar{Z}_p$ =average of the predicted values,  $\bar{Z}_o$ =average

**Table 2** Summary statistics for studied soil quality indicators

Soil indicators	Min	Max	Mean	SD	CV (%)
pH (H <sub>2</sub> O)	4.90	6.55	5.70	0.48	8.49
SOC (%)	0.78	3.39	1.99	0.58	29.13
Ks (cm/h)	0.12	175.56	19.91	24.50	123.04
$\rho_b$ (Mg/m <sup>3</sup> )	0.82	1.88	1.33	0.24	18.15
MC (cm <sup>3</sup> /cm <sup>3</sup> )	0.00	0.61	0.18	0.11	62.39
TP (%)	15.38	66.67	45.02	10.24	22.74
MWD (mm)	0.23	1.62	0.52	0.23	45.14
Sand (%)	41.60	86.00	66.86	9.77	14.62
Silt (%)	9.00	47.00	24.53	7.48	30.48
Clay (%)	3.00	28.70	8.04	4.78	59.47
SSI	4.42	32.78	11.22	4.48	39.93

SOC soil organic carbon, Ks saturated hydraulic conductivity,  $\rho_b$  bulk density, MC moisture content, TP total porosity, MWD mean weight diameter, SSI structural stability index



of the predicted values,  $CCC = \text{Lin's concordance correlation coefficient}$ ,  $\sigma_o^2$  and  $\sigma_p^2$  are the variances of the predicted and observed values; and  $r$  is the Pearson correlation coefficient between the predicted and observed values.

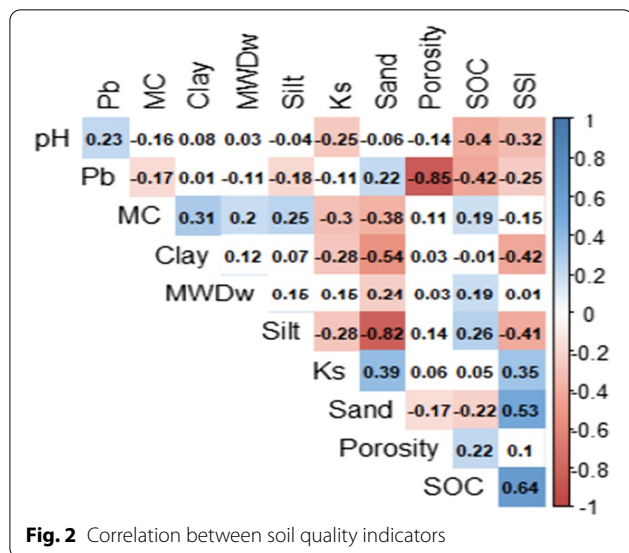


Fig. 2 Correlation between soil quality indicators

## Results

### Basic statistics of soil quality indicators and indicator selection methods

Summary statistics of physical and chemical properties are presented in Table 2. The soil pH of the studied locations ranged from 4.90 to 6.55, with a mean of 5.70. SOC had a minimum value of 0.78% and a maximum value of 3.39%.  $K_s$  value ranged from 0.12 to 175.56 cm/h. Bulk density ranged from 0.82 to 1.88 Mg/m<sup>3</sup>, sand content in all the regions was high (86.0%), accompanied by a low silt content of 47%, and lower clay content of 28.7%. The characteristics of other measured variables are shown in Table 2. The 11 soil properties reported in Table 2 were used as the total data set based on their sensitivity to cause a change in soil functions. The indicators were screened for the MDS using correlation and principal component analysis (PCA) with varimax rotation.

Positive and significant correlations were observed between SOC and SSI ( $r = 0.64$ ), SOC and silt ( $r = 0.26$ ),  $K_s$  and SSI ( $r = 0.35$ ),  $K_s$  and sand ( $r = 0.37$ ), MC and clay ( $r = 0.31$ ), MC and silt ( $r = 0.25$ ), and SSI and sand ( $r = 0.56$ ) at 1% significant level (Fig. 2). Similarly, among the negative correlations, those that were highly significant were between pH and SOC ( $r = -0.40$ ), pH and  $K_s$  ( $r = -0.25$ ), pH and SSI ( $r = -0.32$ ),  $\rho_b$  and SOC ( $r = -0.42$ ), MC and  $K_s$  ( $r = -0.30$ ), silt and  $K_s$  ( $r = -0.28$ ),

Table 3 Results of the principal component analysis

Indicators	Principal component (PC)				Norm values	Grouping	TDS		MDS	
	PC1	PC2	PC3	PC4			Com	Weight	Com	Weight
pH	-0.511	-0.228	-0.093	0.215	0.9896	1	0.510	0.058		
SOC	<b><u>0.904</u></b>	0.125	0.331	-0.048	1.6260	1	0.826	0.094	0.877	0.279
$K_s$	0.059	<b>0.608</b>	-0.189	-0.263	1.0432	2	0.929	0.105		
$\rho_b$	<b>-0.709</b>	-0.014	0.631	-0.154	1.4333	1	0.4960	.056		
Moisture	0.331	<b>-0.472</b>	0.105	0.366	1.0291	2	0.941	0.107		
Porosity	<b>0.534</b>	-0.031	-0.796	0.132	1.3035	1	0.498	0.056		
MWD	0.242	-0.247	0.304	<b>0.490</b>	0.8382	4	0.890	0.101	0.447	0.142
SSI	0.502	<b>0.779</b>	0.223	0.142	1.5553	2	0.957	0.108	0.980	0.312
Sand	-0.315	<b>0.888</b>	-0.029	0.272	1.5562	2	0.893	0.101	0.835	0.266
Silt	0.370	<b>-0.684</b>	0.024	-0.544	1.3899	2	0.503	0.057		
Clay	0.024	<b>-0.628</b>	0.018	0.331	1.0664	2	0.504	0.057		
eigenvalue	3.008	2.594	1.294	1.032						
Variance (%)	27.32	23.61	11.75	9.10						
Cumulative variance (%)	27.32	50.93	62.68	71.78						
KMO	0.502									
BTS	785.085									
df	55									
Sig	<0.01									

PC Principal component, Bold PC loadings are considered highly weighted. Bold and underlined factor loadings are selected as minimum data set (MDS); Com Communality, TDS Total data set, MDS Minimum data set. Grouping was based on norm values; Kaiser-Meyer-Olkin Measure of Sampling Adequacy; Bartlett's Test of Sphericity; MC moisture content, df degree of freedom; KMO Kaiser-Meyer-Olkin, BTS Bartlett's test of Sphericity,  $\rho_b$  bulk density, MWD mean weight diameter,  $K_s$  saturated hydraulic conductivity, SOC soil organic carbon, SSI structural stability index

$K_s$  and clay ( $r = -0.28$ ),  $\rho_b$  and TP ( $r = -0.85$ ),  $\rho_b$  and SSI ( $r = -0.25$ ), MC and sand ( $r = -0.36$ ), SSI and silt ( $r = -0.41$ ), clay and SSI ( $r = -0.42$ ), sand and silt ( $r = -0.85$ ), and sand and clay ( $r = -0.58$ ) at 1% significant level. The observed relationships from the correlation analysis indicated the intricate connections among the various soil properties, which can hardly be observed when using raw data obtained directly from laboratory analysis.

**Computed soil quality indices and indices validation**

The indicators were screened for the MDS using correlation and principal component analysis (PCA) with varimax rotation (Table 3). The studied soil indicators were grouped into four PCs, as they had eigenvalues > 1, each explaining at least 5% of the data variation and accounting for 71.78% of the total variance in the data set. Communalities for the soil indicators showed that the four components explained more than 80% of the variance in SOC,  $K_s$ , MC, MWD, SSI, and sand: > 50% of the variance in pH, silt, and clay, and < 50% of the variance in total porosity (TP) and bulk density (BD). In Group 1, SOC had the highest norm value (1.6260), and no other soil indicators had a norm value falling within the scope of 90% of the highest value. Hence, SOC was selected as MDS for Group 1, and other indicators were eliminated. Similarly, in Group 2, sand had the highest norm value (1.5562), which exceeded that of SSI (1.5553), but the norm value of SSI fell within the scope of 90% of the highest value. The correlation between sand and SSI was < 0.60; hence, SSI and sand were retained in Group 2 as MDS, and other indicators in this group were eliminated. There was no indicator under Group 3. Finally, MWD was the only indicator in Group 4 and was selected as MDS. Therefore, this study selected SOC, sand SSI, and MWD as MDS.

The weight value for TDS showed that SSI (0.108) had the highest weight, while SOC (0.094) had the lowest value. Similarly, for MDS, SSI (0.312) also had the highest weight, while MWD (0.142) had the lowest value. The screened indicators for both TDS and MDS were scored using linear and non-linear scoring functions, and the summary results are presented in Table 4. The

**Table 4** Summary statistics soil quality indices

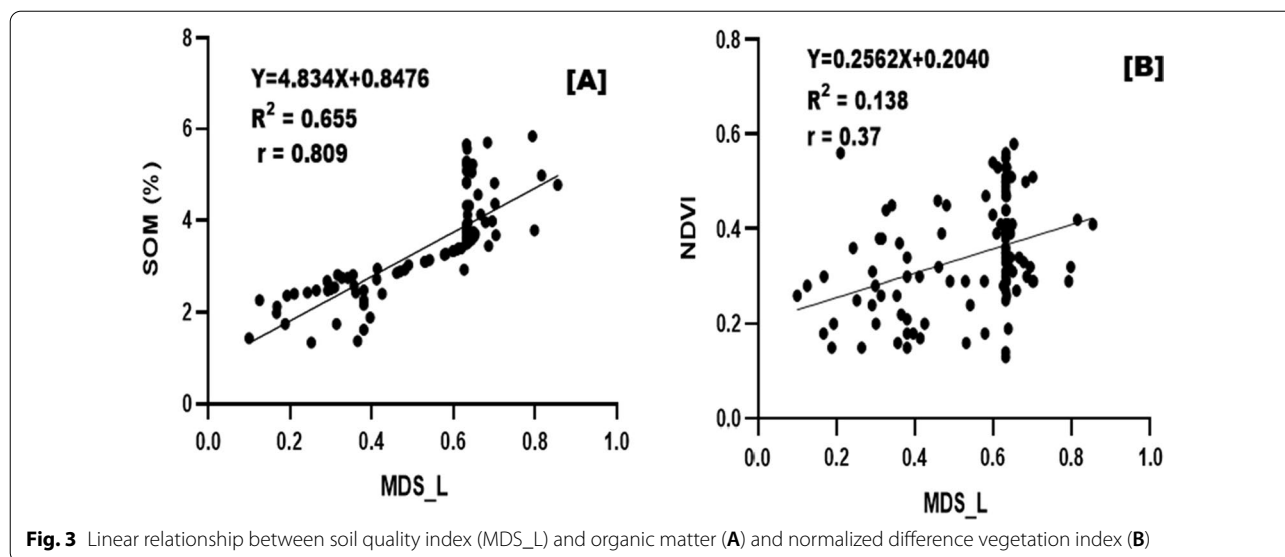
SQL	Min	Max	Mean	SD	CV (%)
MDS_L	0.10	0.86	0.53	0.17	17
MDS_NL	0.21	0.71	0.48	0.10	10
TDS_L	0.27	0.72	0.49	0.094	9.4
TDS_NL	0.23	0.57	0.41	0.069	6.9

**Table 5** Sensitivity index and efficiency ratio between each soil quality index value and soil indicators

Indicators/indices	MDS_L	MDS_NL	TDS_L	TDS_NL
Pearson correlation				
pH			-0.425**	-0.411**
SOC	0.808**	0.934**	0.683**	0.745**
$K_s$			0.153 <sup>NS</sup>	0.176 <sup>NS</sup>
$\rho_b$			-0.592**	-0.602**
MC			0.445**	0.480**
Porosity			0.473**	0.408**
MWD	0.230*	0.406**	0.140 <sup>NS</sup>	0.439**
SSI	0.485**	0.690**	0.378**	0.339**
Sand	-0.127 <sup>NS</sup>	-0.064 <sup>NS</sup>	-0.137	-0.337**
Silt			0.195*	0.289**
Clay			-0.044 <sup>NS</sup>	0.192*
ER (%) values	75	75	72.72	90.90
ER rank	2	2	3	1
SI values	8.60	3.38	2.67	2.48
SI rank	1	2	3	4
Rank scores	3	4	6	5
Final ranking	1	2	4	3

<sup>NS</sup> Not significant  
\* and \*\* mean correlation is significant at the 1% and 5% level respectively (2-tailed)

sensitivity index shows that MDS\_L is the most sensitive index, with a value of 8.60 for evaluating the soil quality index in Cross River State. At the same time, TDS\_NL had the least sensitivity index value of 2.48 (Table 5). Efficiency ratios (ER) were further calculated to specify the power of each SQI. The calculated efficiency ratios for the four developed soil quality indices are presented in Table 5. The TDS\_NL, having an ER ratio of 90.90%, was ranked first, followed by MDS\_L and MDS\_NL, with ER ratios of 75% and 75%, respectively. In addition, to make a decision, the final prioritizing of different indices was conducted by summation ranks of two criteria with an assumption that two selected criteria (SI and ER) have an equal quota on the final decision. The least ranked indicator was selected as the representative, and MDS\_L falls under this scope and hence was selected for further analysis and modelling. The MDS\_L selected through validation using ER and sensitivity analysis was further correlated with SOM and NDVI (Fig. 3) to check its scientific credibility. There is a strong positive correlation between MDS\_L and OM ( $r = 0.809$ ,  $p < 0.01$ ) and NDVI ( $r = 0.37$ ,  $p < 0.01$ ). This implies that improvement in soil organic matter content would subsequently lead to a corresponding increase in soil quality and, by extension, crop yield. This result indicates that MDS\_L could



**Table 6** Summary Statistics of MDS\_L soil quality index

MDS_L	N	Min	Max	Mean	SD	CV (%)	Skewness	Kurtosis
Calibration data set	79	0.10	0.80	0.53	0.17	32.07	−0.91	−0.28
Validation data set	31	0.17	0.86	0.54	0.17	31.48	−0.59	−0.29
Total data set	110	0.10	0.86	0.53	0.17	32.07	−0.81	−0.31

N number of sampling points

be applied to monitor soil quality and crop yield in the study area.

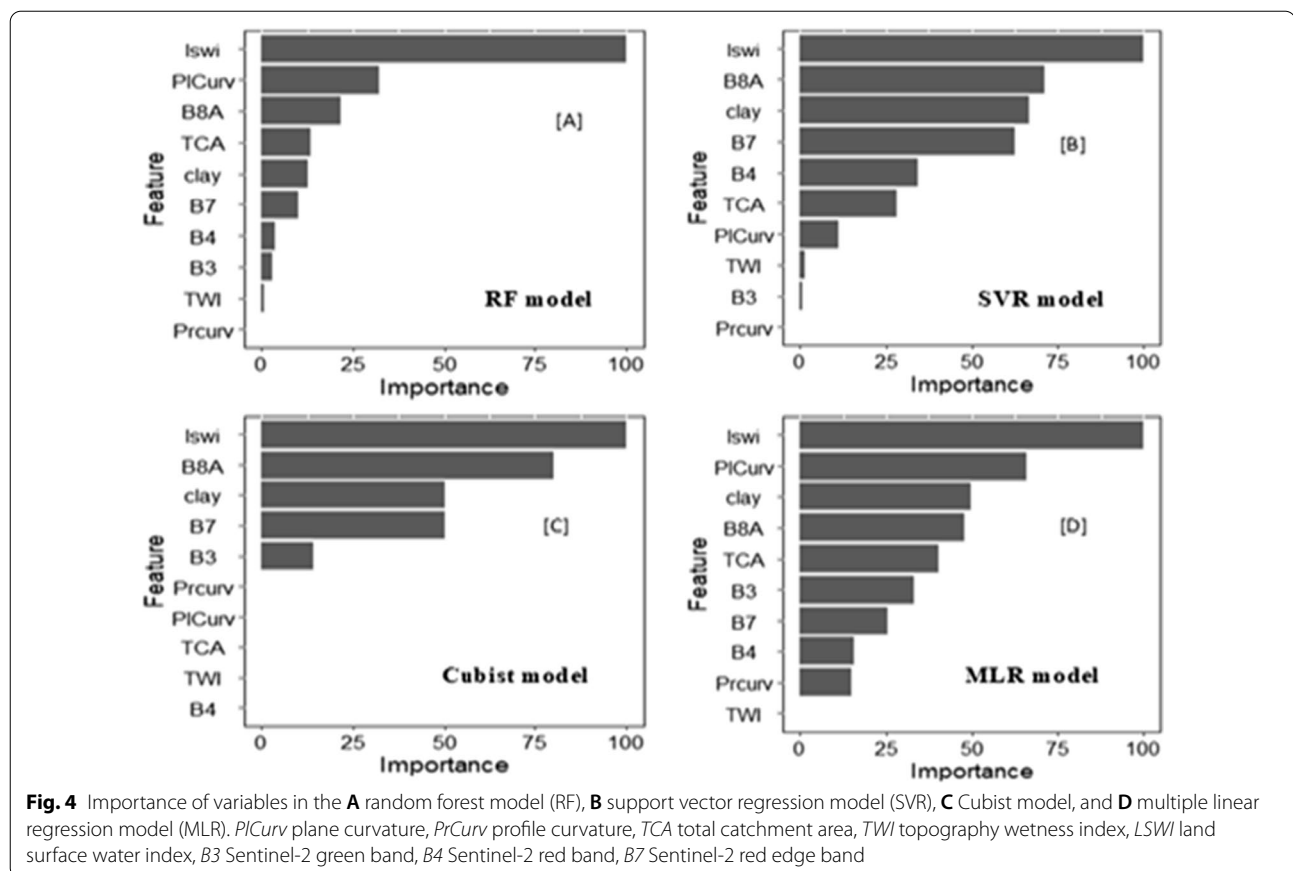
**Modeling approach and variables’ importance in the individual models**

The MDS\_L was selected through validation using ER, and sensitivity analysis was used for prediction. The MDS\_L was split into training (calibration) and testing (validation) data sets, and its summary statistics are presented in Table 6. The basic statistics of calibration and validation data sets were similar to those of the entire data set. The values for all the data sets had moderate variability (15% < CV < 35%), demonstrating the high variability of MDS\_L within the study area (Table 6), and were slightly negatively skewed.

Several models used in this study, including multiple linear regression (MLR), random forest (RF), support vector machine regression (SVR), and Cubist were first fitted with the covariates selected from stepwise regression to quantify their importance with the soil quality index. Predictors with at least 15% important to the soil quality index were finally selected and used for modeling. The relative importance of variables for the applied models is presented in Fig. 4. LSWI [ranked first (100%)],

Plcurv, B8A, clay index, and TCA were the most effective covariates in predicting SQI utilizing RF and MLR models. Similarly, LSWI, B4, B7, B8A, clay index, and TCA were the most effective covariates in predicting SQI utilizing the SVR model. The results also indicated that LSWI, B3, B7, B8A, and clay index showed high importance with the soil quality index using the Cubist model. However, the importance of variables to soil quality prediction through RK was from those already provided by the RF model, and those for prediction of SQI via GWR were from the MLR model. This implementation was appropriate because RK in this study utilizes the residual obtained from the RF model. Similarly, GWR is a localized form of MLR.

The importance of variables (Fig. 4) in the RE, Cubist, SVR, and MLR models was slightly different, revealing different dominating environmental features in these models. The result showed LSWI as the most important variable in all models, with a relative importance of about 100%. The spectral indicators derived from Sentinel-2 imagery, ranging from VIS to NIR reflectance bands (e.g., B8A, B7, B4, and B3), have strong relationships with SQI. However, the importance of different spectral indicators varied. Although the ranking order was different across



the five studied models, LSWI was considered the most relevant variable in all the five models, as it was consistently ranked 1st, and B8 was among the top four important variables across all the models.

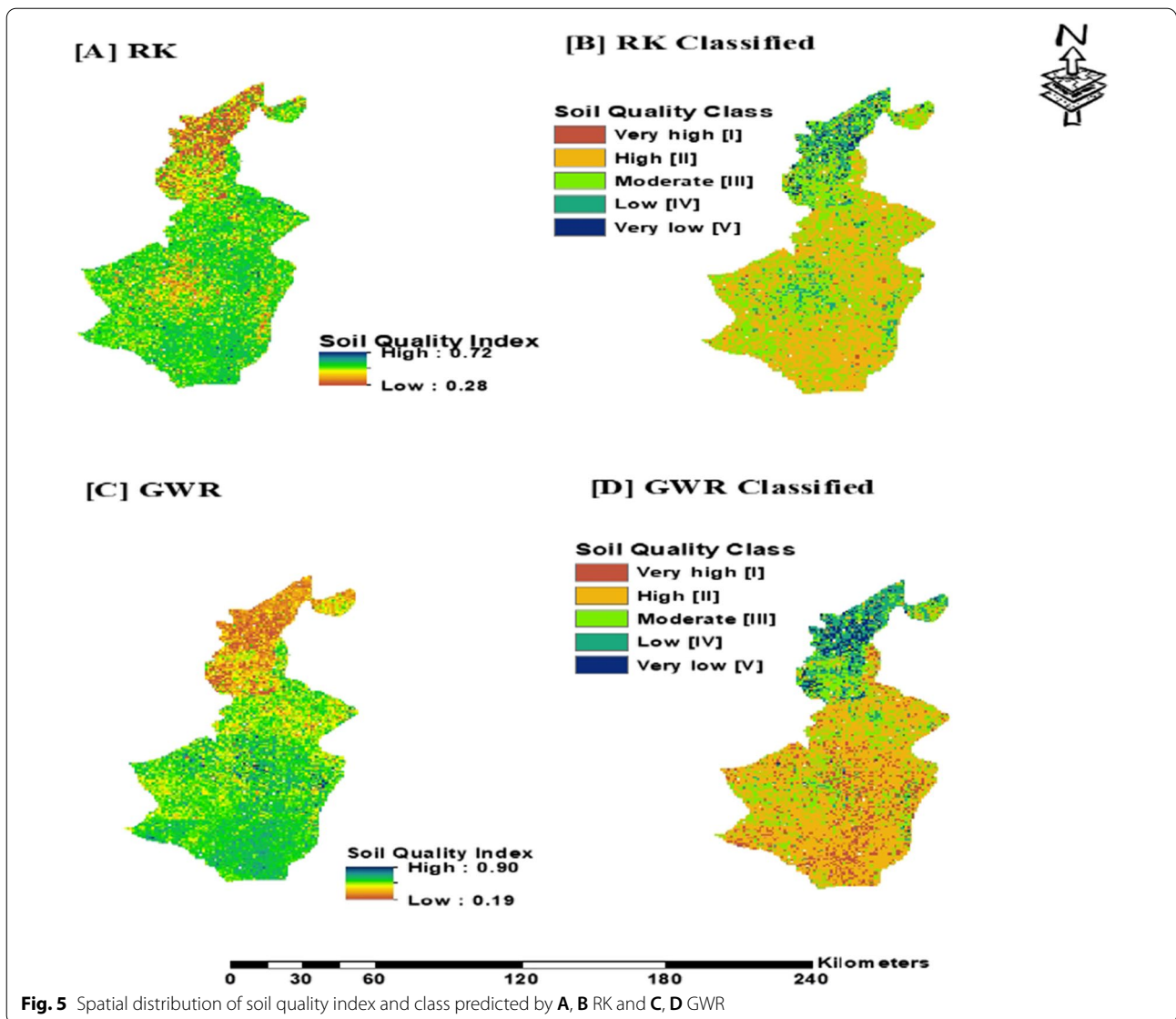
#### Spatial prediction and mapping of soil quality

The spatial distribution of soil quality index (SQI) and class predicted by RF, SVR, Cubist, RK, and GWR is illustrated in Figs. 5 and 6. The mean and standard deviation of the MDS\_L soil quality index for the entire study area based on the Cubist predicted map were 0.58 and 0.084, respectively. In RF predicted map, they were 0.56 and 0.066; in RK predicted map, they were 0.56 and 0.066, respectively; in SVR predicted map, they were 0.61 and 0.055, while in GWR predicted map, they were 0.58 and 0.086, respectively. Similarly, the predicted mean value for SQI by RF (0.56) and RK (0.56) was closer to the mean of the measured value (0.53) than Cubist (0.58), GWR (0.58), and SVR (0.61). This suggests RK and RF could perform well in predicting soil quality index over other proposed models, although this would later be validated. The spatial distribution maps of soil quality index and class obtained using the RF, SVR, RK, Cubist, and GWR models indicated that soil quality in

Cross River State varies with classes ranging from very low to very high quality. All models showed almost the same overall spatial pattern, with a high soil quality region mostly found in the central and southern parts of the study area with values ranging from 0.72 to 0.90. In contrast, low soil quality areas were located mostly in the northern part of the study area, with values ranging from 0.19 to 0.35.

The soil quality index in this study was classified into five classes, with values of  $SQI \leq 0.38$ , 0.38–0.48, 0.48–0.58, 0.58–0.68, and  $\geq 0.68$ , representing very low (class V), low (class IV), moderate (class III), high (class II) and very high (class I) soil quality, respectively. From the spatial distribution map, areas that required high input for optimized crop production are mostly in the northern area. As these maps indicate, soil quality decreased spatially from south to north. The maps reflect that most of these areas may be affected by some soil degradation. In Table 7, with the different predictive models, the dominant grade, called grade II (high class), covered more than 49% of the study area, and this was closely followed by moderate class (grade III), which also covered approximately 14% of the study site. Conversely, the very low class (grade V) was the least observed grade in the area,





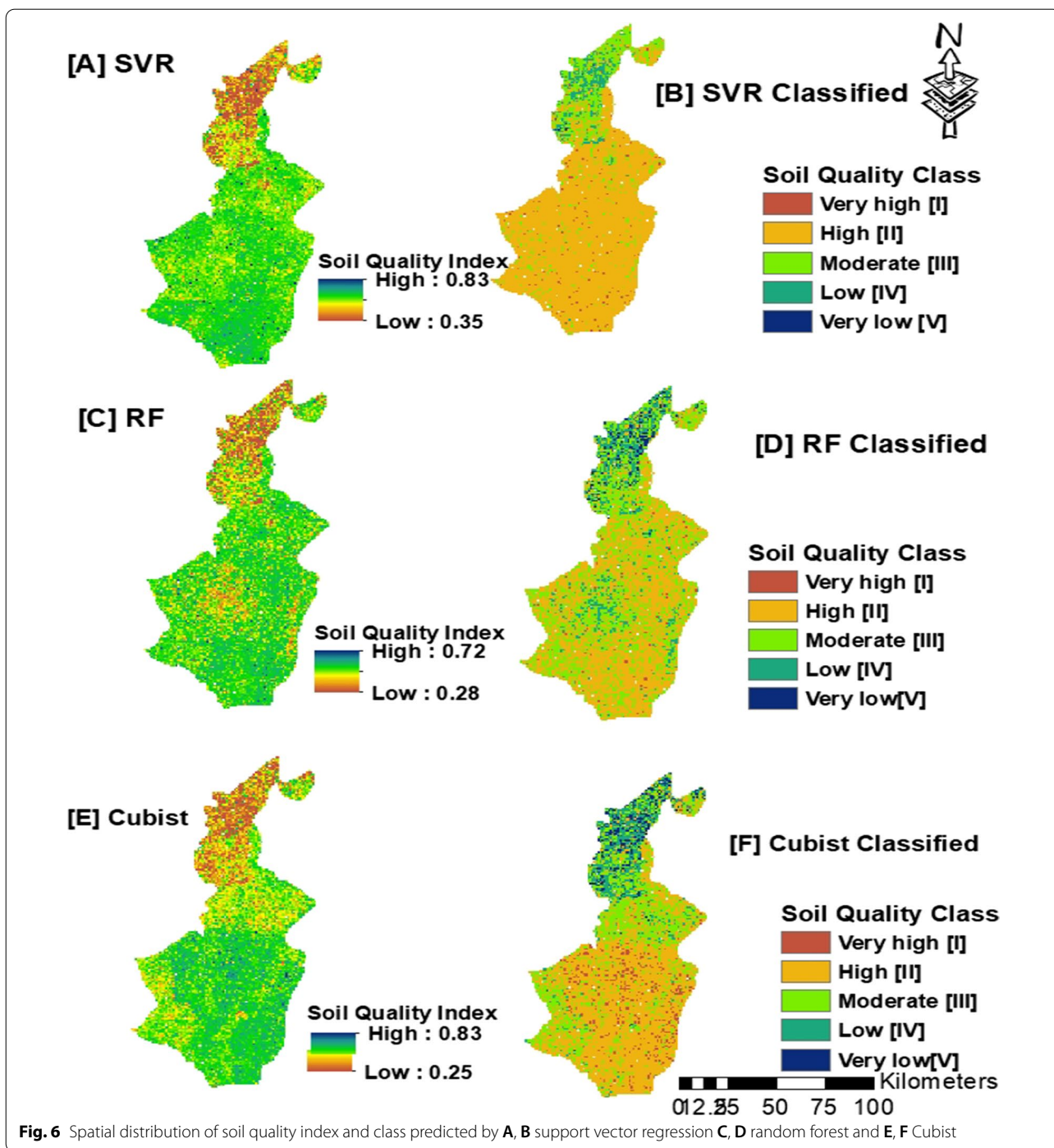
covering less than 4%, and the very high class also occupied a small portion (<8%).

**Evaluation of the performance of machine learning in predicting soil quality index**

The average RMSE,  $R^2$ , and CCC for SQI prediction by cross-validation are shown in Table 8. The proposed machine learning models showed different abilities to predict SQI at unsampled locations in the study area. This could be related to the various mathematical functions of each algorithm and covariates used for fitting. Prediction values of SQI using RF, SVR, Cubist, RK, and GWR were compared, and the results showed discrepancies between these models. SVR had the highest coefficient of determination ( $R^2=0.24$ ), indicating high

precision; RK and GWR had equal and the highest CCC (0.39), implying good agreement with the 45° line, while SVR, RK, and GWR had an equal and lowest root mean squared error (RMSE=0.15). The RK and GWR models predicted SQI better than other models; this is evident, because the regression lines observed against the predicted are closer to the 1:1 line than what is obtainable by SVR, RF, and Cubist. Lin’s concordance correlation coefficient (CCC) is used to compare the regressions to the 1:1 line. For the RK and GWR models, concordance was 0.39 and 0.39. Upon visual examination of Fig. 7, the RK and GWR models show more variability or scatter in the data than the other models.

The RF, Cubist, and SVR models showed a high tendency for overestimation or underestimation, while



RK and GWR showed a minimal tendency for over-estimation or underestimation. RF and Cubist models underestimate high values and overestimate low values of SQI, while the SVR model overestimates high values and underestimates low values of SQI as shown by the 1:1 regression line (45° line) (Fig. 7). Overall results considering all validation indices showed RK with

criterion ( $R^2=0.20$ ,  $CCC=0.39$ ,  $RMSE=0.15$ ) and GWR ( $R^2=0.21$ ,  $CCC=0.39$ ,  $RMSE=0.15$ ) to be the best performing models. This was followed by SVM with a slightly inferior performance for each error criterion ( $R^2=0.24$ ,  $CCC=0.32$ ,  $RMSE=0.15$ ). Cubist and RF showed a higher deviation of predicted to measured values. From the results, it can be concluded that the GWR

**Table 7** Location falling within different soil quality grades according to the utilized predictive models

Soil quality grade	RF		SVR		Cubist		RK		GWR	
	Area (ha)	%	Area (ha)	%	Area (ha)	%	Area (ha)	%	Area (ha)	%
Very high (class I)	7405	0.79	15,876	1.69	48,832	5.19	7405	0.79	70,978.05	7.54
High (class II)	477,602	50.74	749,193	79.59	525,479	55.82	477,602	50.74	469,098.09	49.83
Moderate (class III)	341,259	36.25	132,478	14.07	244,495	25.97	341,259	36.25	258,513.75	27.46
Low (class IV)	92,883	9.87	42,796	4.55	87,426	9.29	92,883	9.87	120,964.14	12.85
Very low (class V)	22,182	2.36	989	0.11	35,100	3.73	22,182	2.36	21,777.93	2.31
Total	941,332	100	941,332	100	941,332	100	941,332	100	941,332	100

**Table 8** Performance of predictive models in predicting soil quality index

Models	R <sup>2</sup>	CCC	MSE	RMSE	Bias
RF	0.02	0.16	0.027	0.17	0.002
SVR	0.24	0.32	0.022	0.15	0.039
Cubist	0.07	0.26	0.026	0.16	0.022
RK	0.20	0.39	0.022	0.15	0.031
GWR	0.21	0.39	0.022	0.15	0.034

model showed better performance in predicting SQI at new locations than other models, given its lowest RMSE and highest  $R^2$  and CCC. The GWR approach applied regressions locally, which accounted for both the spatial trends and local variations resulting in superior estimations of SQI.

#### Impact of geological materials on soil quality

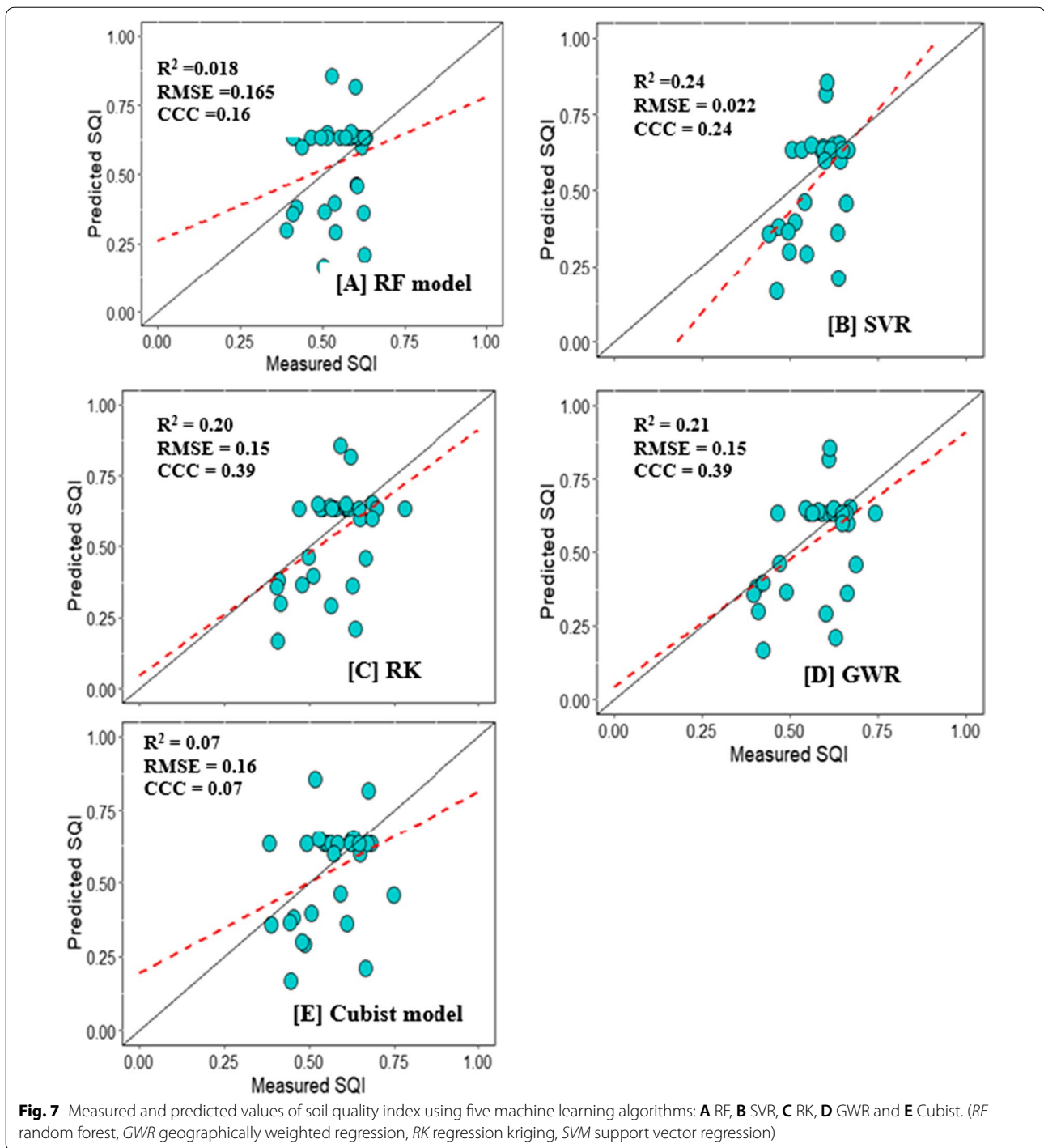
The study indicated that about 53,509.23 ha (75.95%) and 12, 851.01 ha (18.24%) of soil with the very high-quality class were found in soil developed on basement complex formation and sandstone, while only 2.59%, 3.14%, and 0.08% were found for soil developed on shale, limestone, and basalt, respectively (Table 9 and Fig. 8). Similarly, under the high soil quality class, 317,625.40 ha (68.19%) and 107, 533.10 ha (23.08%) were found in soil developed on basement complex formation and sandstone, whereas 4.86%, 2.17%, and 1.70% were associated with shale, basalt and limestone soils. In addition, in the moderate soil quality class, 46.51%, 37.54%, 10.25%, 5.58%, and 0.11% of soils developed on basement complex formation, sandstone, shale, basalt, and limestone were associated with moderate soil quality class.

However, low soil quality classes were associated with sandstone (42.84%) and shale (41.67%) as well as very low soil quality classes: sandstone (41.49%) and shale (52.93%). The results revealed that very high, high, and moderate soil quality was mostly associated with soil developed on basement complex formation and

sandstone, whereas low and very low soil qualities were associated with soil developed on basalt, limestone, and shale. The result of this study pointing out basement complex formation has been very high, high, and moderate in soil quality is not surprising. A relatively high soil quality index was observed in the Alesi in the central part of the study area and Oban, Aningeji, among other locations in the southern part, while a relatively low soil quality index was observed in Winnimba, Abakpa, and Alok, among others in the northern part of the study area (Fig. 8). These mentioned areas are the major crop production base in Cross River State.

#### Discussion

In this study, eleven soil quality indicators were assessed as TDS based on their sensitivity to cause a change in soil functions (Andrews et al. 2004), and later screened to four indicators (SOC, MWD, SSI, and sand) via PCA procedures. In the context of soil quality estimation, PCA is recognized as one of the most effective tools for reducing the number of variables by identifying those that are most significant in the field scale for estimating soil quality (Andrews and Carroll 2001; Fathizad et al. 2020). The MDS selected via PCA are indicators of soil texture, nutrient and soil aggregation, and structural development (Phogat et al. 2015; Jat et al. 2018), and could play an important role in assessing soil quality in the study area. These properties have also been utilized elsewhere to study soil quality. For instance, Chen et al. (2013), Fathizad et al. (2020), Choudhury and Mandal (2021), and Nabiollahi et al. (2018b) have also used soil quality indicators utilized in this study as a data set in their studies on soil quality assessment. Prior to the advent of precision agriculture and geospatial technologies, numerous soil variables were required for sustainable soil management. The introduction of soil quality aided by machine learning and geostatistics became robust in identifying key soil indicators for nutrient management, which reduce both the time and costs associated with



in-situ and laboratory analyses of numerous soil variables required for soil nutrient management.

In the studied soil, the mean bulk density of the plough layer ( $\rho_b = 1.33 \text{ Mg/m}^3$ ) was lower than the 'optimum' value of  $1.40 \text{ Mg/m}^3$  (USDA-NRCS 2001), stipulated for sandy loams, loams, sandy clay loams, loams, clay loams which

were the dominant observable texture in the study area. Nevertheless, a maximum bulk density of  $1.88 \text{ Mg/m}^3$  was obtained. The soil bulk density values exceeding the critical value in the studied soil can impede crop root growth and development, thereby reducing soil quality and crop yield. Similarly, saturated hydraulic conductivity ( $K_s$ ) in the soil



**Table 9** Impact of geological materials on soil quality

Parent materials	Very high class		High class		Moderate class		Low class		Very low class	
	Area (ha)	%	Area (ha)	%	Area (ha)	%	Area (ha)	%	Area (ha)	%
Sandstone	12,851.01	18.24	107,533.10	23.08	96,024.87	37.54	50,898.51	42.84	8982.54	41.49
Basalt	58.23	0.08	10,107.90	2.17	14,267.34	5.58	2131.83	1.79	199.26	0.92
Basement complex	53,509.23	75.95	317,625.40	68.19	118,961.6	46.51	16,266.06	13.69	1009.08	4.66
Limestone	1826.1	2.59	7901.64	1.70	292.05	0.11	0	0.00	0.00	0.00
Shale	2212.2	3.14	22,656.33	4.86	26,227.44	10.25	49,504.41	41.67	11,457.72	52.93
Total	70,456.77	100	465,824.30	100	255,773.3	100	118,800.8	100	21,648.6	100

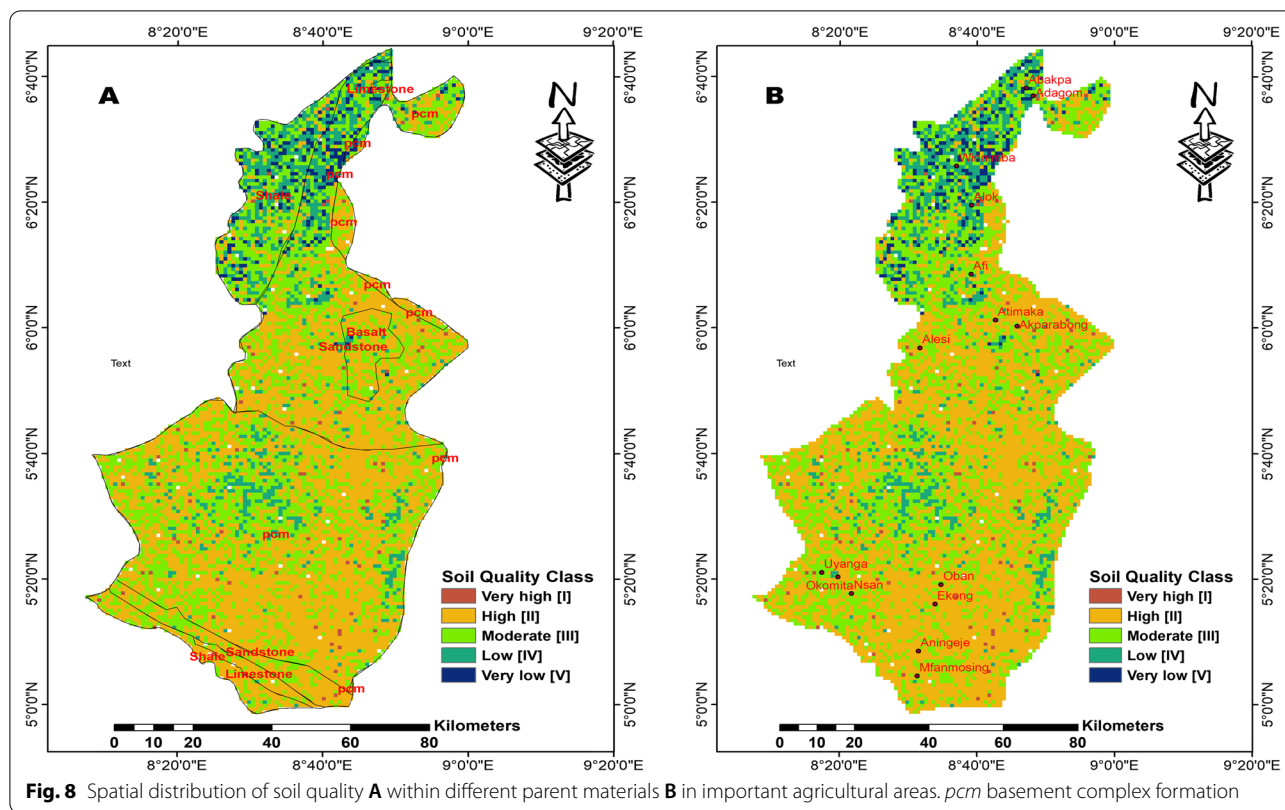
was found to be moderate.  $K_s$  is often used as a measure of soil physical quality (Andrews et al. 2004).  $K_s$  and texture are interrelated; areas with the higher  $K_s$  value are expected to have high sand content and low clay content. This effect can be illustrated by a study conducted by Lim et al. (2016), where  $K_s$  of 5.98 m/day for coarse sand decreased by 57%, 88%, and 96%, with the successively decreasing sand content in fine sand, loam, and clay textured soils. The pH value of 5.7 obtained in this study indicates that the region's soil is moderately acidic. Most of the nutrient elements are available at a pH range of 5.5–7.0 for optimal growth, hence, the soil of the area can be utilized with minimum application of lime in places where pH is less than 5.5 (Brady and Weil 2002). Soil organic carbon ranged from 0.78% to 3.39%. Low SOC in topsoil (0–30 cm) was expected in the northern region, where the temperature is hot, and vegetation is low compared to the southern part. Very low NDVI values predominate the northern part of the study area, indicating poor vegetative growth. This impacted soil organic carbon, and perhaps soil quality in these areas. However, areas dominated by moderate and high NDVI values were found in the central and southern parts of the study area. This reflects thick vegetative cover, and soil quality in these areas is expected to be very high. Most of the studied soil properties were correlated. Several researchers also observed such correlations in their studies. For instance, MacCarthy et al. (2013) in their research found a negative relationship between SOC and  $\rho_b$ , Evrendilek et al. (2004) had a similar relationship between SOC and  $\rho_b$ , and SOC with soil pH, while Adhikari and Bhattacharyya (2015) found such a relationship existing between SOC and sand.

The physical indicators investigated in this study provide information on root growth, ease of plant emergence, and water infiltration. In contrast, the chemical indicators provide information on the proliferation of soil organisms and nutrient availability. The major parameters used in establishing regional soil quality indices in the study area are SSI, sand, SOC, and MWD. Hence, understanding these parameters is essential for illustrating the potential steps

of proper soil management for sustainable agricultural production.

In many countries of the world, soil quality is declining rapidly. Therefore, the estimation and prediction of soil quality are considered the basis for monitoring and maintaining sustainable agricultural systems. In the literature, soil quality index ranged from 0.0 to 1.0 (Andrews et al. 2004). The soil quality index in this study was classified into five classes, with values of  $SQI \leq 0.38$ , 0.38–0.48, 0.48–0.58, 0.58–0.68, and  $\geq 0.68$ , representing very low (class V), low (class IV), moderate (class III), high (class II) and very high (class I) soil quality, respectively. The delineation and reclassification were guided by previous studies (Andrews et al. 2004; Chen et al. 2013; Guo et al. 2017; Fathizad et al. 2020). In addition, the delineation of soil quality into classes for homogeneous management can serve as a cost-effective approach for responsibly improving and managing soil resources. This present study revealed more accurately different soil classes (very high, high, moderate, low, and very low) with the combination of soil-covariates and predictive models.

The most important covariates for predicting SQI with RF and MLR models were LSWI, Plcurv, B8A, clay index, and TCA. In contrast, LSWI, B4, B7, B8A, clay index, and TCA were the most important covariates for predicting SQI using the SVR model. In previous studies, the normalized difference water index (NDWI) was the most important variable that detected SOC variability in a study conducted by Falahatkar et al. (2016). Plane curvature controls the flow of solutes, water, and sediments and can affect soil development and the spatial distribution of soil properties (Li et al. 2013), and it was also found among the most effective covariates that could predict SQI in this study. Similar to the results of this study, Paul et al. (2020) and Fathizad et al. (2020) found that topographic variables were the most important predictors for soil quality prediction. Zeraatpisheh et al. (2019), in a study for modelling SOC in surface soils of central Iran, also reported clay index (CI), ranking 4th in order of importance for



predicting SOC. The spectral indicators derived from Sentinel-2 imagery, ranging from VIS to NIR reflectance bands (e.g. B8A, B7, B4, and B3), have strong relationships with SQL. Thus, the result of this work collaborates with the findings of John et al. (2020) and John et al. (2021c). They confirmed the high suitability of remotely sensed data and terrain attributes for predicting soil attributes.

The spatial trend of soil quality decreases from the southern to the northern region of the study area, and this trend is consistent with current conditions of the studied soil, as low soil quality in the north may be due to a low percentage of SOC, low soil structural stability and a high bulk density, as well as intensive mismanagement practices in agriculture. Zhang et al. (2016) and Mukherjee and Lal (2014) have shown that SOC, bulk density, clay, and pH are the most influential factors in the determination of SQL, confirming the importance of using these parameters in the current study. Developing soil quality classes can minimize agricultural management costs and input wastage. In the study area, soil nutrient recommendations for agricultural soils are usually uniform, with the spatial heterogeneity of nutrient content in soils often neglected, and the major source of chemical fertilizers in Nigeria and sub-Saharan Africa at large is NPK, urea (46% N), superphosphate [triple superphosphate (46% P)], muriate of potash (60% K)

(Liverpool-Tasie et al. 2010). In addition, poultry, pig and cow manure, and other soil amendments are also applied in the garden and on small-scale farms (Uko et al. 2019). These similar nutrient recommendations at a regional scale sometimes lead to over-fertilization in areas with high nutrient levels and vice versa. Therefore, classifying soils based on their potential to support crop production could increase nutrient use efficiency in commercial farms and decrease the risk of nutrient pollution. Improving soil quality for a moderate soil quality class requires fewer management costs than a very low soil quality class with the most severe limitations. The study showed that areas that could provide optimum plant growth conditions and lower sensitivity to erosive processes were found in the southern and central regions of the study area and have been rated as having higher soil quality.

In the last two decades, more emphasis has been placed on food security and sustainable management of resources (Okon et al. 2019; Uko et al. 2019; Zeraatpisheh et al. 2020). Soils are a resource base where crop production takes place and are important in ensuring food security because over 97.5% of the human food supply comes from the soil while less than 2.5% comes from aquatic systems (Brevik 2013), thus making soils critical to food security. Santos-Francés et al. (2022) stated that

"there has been an increased global demand to establish criteria for determining soil quality for quantitative indices that can be used to classify and compare that quality in different places." In previous studies, soil quality models have been performed using different methodologies and scales of analysis. For instance, Nabiollahi et al. (2018a) and Paul et al. (2020) utilize RF in modelling soil quality. However, most models utilized in this study have not been employed previously in soil quality mapping, although they have been widely used in assessments of other environmental issues and soil properties variability (Afu et al. 2021; John et al. 2020, 2021a, 2021b). Besides, there is now a growing interest in and the application of machine learning and geostatistical models in the spatial prediction and production of maps to monitor soil resources, support crop production, and reveal attention to the threat posed by land degradation in African farming systems. Compared with the traditional soil studies (Asensio et al. 2013; Chen et al. 2013; Amalu and Isong 2017), the results from digital soil mapping could ensure greater efficiency and better representation of the horizontal distribution of soil quality across a large region. Planners and farmers can easily use the digital soil map output in this study to identify suitable soils best suited for crop cultivation in this region.

Similar to the agricultural soils elsewhere, the soils of the present study site are developed on diverse parent materials and support different arable crops. In addition, comparing the potential of soil developed on diverse parent materials for agricultural productivity had mixed results as reported by (Abam and Orji 2019; Donatus et al. 2018; Corbett 2006; Gonçalves et al. 2013; Graham et al. 2017). However, Afu et al. (2021) and Ofem et al. (2020) have reported limestone, shale, and basaltic-derived soils to be suitable for agricultural production. The geographical location may have influenced their results as the soils were located in residential and industrial areas. In addition, the soils may either be contaminated with heavy metals or unavailable for crop production. Eshett et al. (1990) reported the potential of soils developed on basement complex formations to support peasant food crops and commercial tree crop production in sub-Saharan Africa for many years. Other scholars (Floyd 1969; Ofomata 1975; Eshett et al. 1990) reported in their studies that certain areas within the study site with soils from basement complex formation are actively utilized for both peasant and commercial crop production. In general, basement complex formation and sandstone are high in soil quality. This finding is associated with the fact that parent materials play a significant role in soil physical characteristics and nutrient supply, especially the release of basic cations, including but not limited to Ca, Mg and K (Afu et al. 2021).

## Conclusion

Among several predictors considered in this study, land surface water index (LSWI), plane curvature (PICurv), clay, and near-infrared band (B8) significantly affected soil quality in the study area. At the same time, soil stability index (SSI), soil organic carbon (SOC), and mean weight diameter (MWD) of aggregates were parameters to establish regional soil quality indices. They are valuable indicators in soil quality prediction on soil developed on diverse parent materials. The minimum data set (MDS) linear method and geographically weighted regression (GWR) are effective and useful models to identify the key soil properties for assessing soil quality for soils developed on diverse parent materials. The joint use of soil quality indicators, environmental covariates, and machine learning algorithms allows for an accurate and effective assessment of soil quality index. The resulting soil quality map from the GWR prediction showed low soil quality in the northern region and high soil quality in the southern and central regions. This indicates that the study area is advancing towards poor soil quality in the northward direction. Soil quality was poorer in soils developed on shale, basalt, and limestone parent materials and richer in soils developed on basement complex and sandstone parent materials in the studied region. Typically, low and very low soil quality requires the application of organic manures/crop residues and fallow cropping systems, in addition to other soil management practices, to achieve high soil quality.

Based on the findings from this study, we recommend soil management approaches (e.g., biochar and/or biofertilizer application, integrated nutrient management, etc.) before using soils developed on basalt, shale, and limestone for crop production. The predictive soil quality maps derived from this study should serve as a guide in establishing regionalized soil nutrient management programmes.

## Abbreviations

B8: Sentinel-2 near-infrared band; CCC: Lin's concordance correlation coefficient; GWR: Geographically weighted regression; LSWI: Land surface water index; MDS: Minimum data set; MDS\_L: Minimum data set linear model; MDS\_NL: Minimum data set non-linear model; OK: Ordinary kriging; RF: Random forest; RK: Regression kriging; PICurv: Plane curvature; SOC: Soil organic carbon; SSI: Structural stability index; SQL: Soil quality index; SVR: Support vector regression; TDS: Total data set; TDS\_L: Total data set linear model; TDS\_NL: Total data set non-linear model.

## Acknowledgements

The authors are grateful to the laboratory staff of the Department of Soil Science, University of Calabar for guiding in the processing of the samples and for their expertise in the laboratory protocol/analysis.

## Author contributions

Conceptualization, IAI; methodology, IAI and JK; formal analysis, IAI; field survey, data collection and investigation, IAI, PBO, and SMA; writing—original

draft preparation, IA; writing—review and editing, IA, JK, PIO and PBO; supervision, PBO and PIO. All authors read and approved the final manuscript.

#### Funding

Not available.

#### Availability of data and materials

All data generated and analyzed during this study are included in this published article.

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Soil Science, Faculty of Agriculture, University of Calabar, Calabar PMB 1115, Nigeria. <sup>2</sup>Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences, Kamýcká, 129 16500 Prague, Czech Republic.

Received: 11 June 2022 Accepted: 7 November 2022

Published online: 25 November 2022

#### References

- Abam PO, Orji OA (2019) Morphological and physico-chemical properties of soils formed from diverse parent materials in Cross River State, Nigeria. *J Appl Geol Geophys* 7(1):1–7
- Adhikari G, Bhattacharyya KG (2015) Correlation of soil organic carbon and nutrients (NPK) to soil mineralogy, texture, aggregation, and land use pattern. *Environ Monit Assess* 187:735
- Afu SM, Isong IA, Akpan JF, Olim DM, Eziedo PC (2021) Spatial assessment of heavy metal contamination in agricultural soils developed on basaltic and sandstone parent materials. *J Environ Sci Technol* 14:21–34
- Aki EE, Esu IE, Akpan-Idiok AU (2014) Pedological study of soils developed on biotite-hornblende-gneiss in Akamkpa Local Government Area of Cross River State, Nigeria. *Int J Agri Res* 9:187–199
- Amalu UC, Isong IA (2015) Land capability and soil suitability of some acid sand soil supporting oil palm (*Elaeis guineensis* Jacq) trees in Calabar, Nigeria. *Nigerian J Soil Sci* 25:92–109
- Amalu UC, Isong IA (2017) Long-term impact of climate variables on agricultural lands in Calabar, Nigeria. II. Degradation of physical properties of soils. *Nigerian J Crop Sci* 4(2):95–102
- Andrews SS, Carroll CR (2001) Designing a soil quality assessment tool for sustainable agroecosystem management. *Ecol Appl* 11:1573–1585
- Andrews SS, Karlen DL, Mitchell JP (2002) A comparison of soil quality indexing methods for vegetable production systems in Northern California. *Agric Ecosyst Environ* 90:25–45
- Andrews SS, Karlen DL, Cambardella CA (2004) The soil management assessment framework. *Soil Sci Soc Am J* 68(6):1945–1962
- Angers DA, Mehuys G (1993) Aggregate stability to water. In: Carter MR (ed) *Soil sampling and methods of analysis*. Lewis Publisher, Boca Raton, pp 651–657
- Asensio V, Guala SD, Vega FL, Covelo EF (2013) A soil quality index for reclaimed mine soils. *Environ Toxicol Chem* 32:2240–2248
- Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future koopen-geiger climate classification maps at 1-km resolution. *Scientific Data* 5:180214. <https://doi.org/10.1038/sdata.2018.214>
- Biau G (2012) Analysis of a random forests model. *J Mach Learn Res* 13:1063–1095
- Blake GR (1965) Particle density. In: Black CA (ed) *Methods of soil analysis, Part I: agronomy*. American Society of Agronomy, Madison, pp 371–373
- Brady NC, Weil RR (2002) *The nature and properties of soil*, 13th edn. Prentice Hall, Upper Saddle River
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Brevik EC (2013) The potential impact of climate change on soil properties and processes and corresponding influence on food security. *Agriculture* 3:398–417
- Callo-Concha D, Gaiser T, Ewert F (2012) Farming and cropping systems in the West African Sudanian Savanna. WASCAL research area: Northern Ghana, Southwest Burkina Faso and Northern Benin. ZEF Working Paper 100. Bonn, Germany.
- Campos AR, Giasson E, Costa JF, Machado RM, da Silva EB, Bonfatti BR (2018) Selection of environmental covariates for classifier training applied in digital soil mapping. *Rev Bras Cienc Solo* 42:1–15
- Carrizo ME, Alesso CA, Cosentino D, Imhoff S (2015) Aggregation agents and structural stability in soils with different texture and organic carbon contents. *Sci Agric* 72(1):75–82
- Castellini M, Iovino M, Pirastru M, Niedda M, Bagarello V (2016) Use of BEST procedure to assess soil physical quality in the Baratz Lake catchment, Sardinia, Italy. *Soil Sci Soc Am J* 80:742–755
- Chen YD, Wang HY, Zhou JM, Xing L, Zhu BS, Zhao YC, Chen XQ (2013) Minimum data set for assessing soil quality in farmland of Northeast China. *Pedosphere* 23(5):564–576
- Choudhury BU, Mandal S (2021) Indexing soil properties through constructing minimum datasets for soil quality assessment of surface and profile soils of intermontane valley, Barak, North East India. *Ecol Indic* 123:107369
- Corbett JR (2006) The genesis of some basaltic soils in New South Wales. *Eur J Sci* 19(1):174–185
- Debi SR, Bhattacharjee S, Aka TD, Paul SC, Roy MC, Salam MS, Islam MS, Azady AR (2019) Soil quality of cultivated land in urban and rural area on the basis of both minimum data set and expert opinion. *Int J Hum Capital Urban Manag* 4(4):247–258
- Dexter AR (2004) Soil physical quality. Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. *Geoderma* 120:201–214
- Donatus AEO, Osodeke VE, Ukpong IM, Osisi AF (2018) Chemistry and mineralogy of soils derived from different parent materials in Southeastern Nigeria. *Int J Plant Soil Sci* 25(3):1–16
- Eshett ET, Omuetti JAI, Juo ASR (1990) Physicochemical, morphological, and clay mineralogical properties of soils overlying basement complex rocks in Ogoja, northern Cross River State of Nigeria. *Soil Sci Plant Nutr* 36(2):203–214
- Evrendilek F, Celik I, Kilic S (2004) Changes in soil organic carbon and other physical soil properties along adjacent Mediterranean forest, grassland, and cropland ecosystems in Turkey. *J Arid Environ* 59:743–752
- Falahatkar S, Hosseini MS, Ayoubi S, Mahiny AS (2016) Predicting soil organic carbon density using auxiliary environmental variables in northern Iran. *Arch Agron Soil Sci* 62:375–393
- Fathizad H, Ali M, Ardakani H, Heung B, Sodaiezhadeh H, Rahmani A, Fathabadi A, Scholten T, Taghizadeh-Mehrjardi R (2020) Spatio-temporal dynamic of soil quality in the central Iranian desert modeled with machine learning and digital soil assessment techniques. *Ecol Indic* 118:106736
- Fayad I, Baghdadi N, Bailly J, Barbier N, Gond V, Hérault B, El Hajj M, Fabre F, Perrin J (2016) Regional scale rainforest height mapping using regression-kriging of spaceborne and airborne LiDAR data: application on French Guiana. *Remote Sens* 8:240
- Floyd B (1969) *Eastern Nigeria: a geographical review*. MacMillan Co. Ltd, London, pp 67–110
- Fox EW, Ver Hoef JM, Olsen AR (2020) Comparing spatial regression to random forests for large environmental data sets. *PLoS ONE* 15(3):e0229509
- Freeman EA, Moisen GG, Coulston JW, Wilson BT (2015) Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Can J For Res* 45:1–17
- Gardner WH (1986) Water content. In: Klute A (ed) *Methods of soil analysis, Part 1: physical and mineralogical methods*, 2nd edn. American Society of Agronomy and Soil Science Society of America, Madison, pp 635–662
- Gee WG, Or D (2002) Particle-size analysis. In: Dane J, Topp GC (eds) *Methods of soil analysis*. Book series: 5. Part 4. Soil Science Society of America, USA, pp 255–293



- Gonçalves MA, Filho JT, Vendrame PRS, Telles TS (2013) Toposequences of soils developed on basaltic rocks: physicochemical attributes. *Amazon J Agric Environ Sci* 56(4):359–370
- Graham RC, Schoeneberger PJ, Breiner JM (2017) Genesis and physical behavior of soils on sandstone and shale in Southern California. *Soil Sci* 182(6):216–226
- Grossman RB, Reinsch TG (2002) Bulk density and linear extensibility. In: Dane JM, Topp GC (eds) *Methods of soil analysis*. Part 4. Physical methods. Soil Science Society of America, Madison, pp 201–228
- Guo LL, Hao HJ, Liu YH, Ma HB, An JB, Sun Q, Yang Z (2017) The assessment of soil quality on the arable land in Yellow River delta combined with remote sensing technology. *World J Eng Technol* 5:18–26
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics, 2nd edn. Springer, New York
- Hengl T, Heuvelink GBM, Kempen B, Leenaars JGB, Walsh MG, Shepherd KD, Sila A, MacMillan RA, De Jesus JM, Tamene L, Tondoh JE (2015) Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS ONE* 10(6):e0125814. <https://doi.org/10.1371/journal.pone.0125814>
- Hengl T, Leenaars JGB, Shepherd KD, Walsh MG, Heuvelink GBM, Mamo T, Tilahun H, Berkhout E, Cooper M, Fegeus E, Wheeler I, Kwabena N (2017) Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr Cycl Agroecosyst* 109:77–102. <https://doi.org/10.1007/s10705-017-9870-x>
- Igwue CA, Zarei M, Stahr K (2013) Stability of aggregates of some weathered soils in southeastern Nigeria in relation to their geochemical properties. *J Earth Syst Sci* 122(5):1283–1294
- Jat ML, Stirling BCM, Jat HS, Tatarwal JP, Jat RK, Singh R, Lopez-Ridaura S, Shirsath PB (2018) Soil processes and wheat cropping under emerging climate change scenarios in South Asia. In: Sparks DL (ed) *Advances in agronomy*, vol 148. Academic Press, Cambridge, pp 111–171
- John K, Isong IA, Ayito KMN et al (2020) Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. *Land* 9:487
- John K, Afu SM, Isong IA, Aki EE, Kebonye NM, Ayito EO, Chapman PA, Eyong MO, Penizek V (2021a) Mapping soil properties with soil-environmental covariates using geostatistics and multivariate statistic. *Int J Environ Sci Technol* 18:3327–3342. <https://doi.org/10.1007/s13762-020-03089>
- John K, Agyeman PC, Kebonye NM, Isong IA, Ayito EO, Ofem KI, Qin C (2021b) Hybridization of cokriging and Gaussian process regression modelling techniques in mapping soil sulphur. *Catena* 206:105534
- John K, Bouslih Y, Ofem KI, Hssaini L, Razouk R, Okon PB, Isong IA, Agyeman PC, Kebonye NM, Qin C (2021c) Do model choice and sample ratios separately or simultaneously influence soil organic matter prediction? *Int Soil Water Conserv Res* 10:470–486. <https://doi.org/10.1016/j.iswcr.2021.11.003>
- John K, Isong IA, Kebonye MN, Agyeman CP, Ayito EO, Kudjo AS (2021d) Soil organic carbon prediction with terrain derivatives using geostatistics and sequential Gaussian simulation. *J Saudi Soc Agric Sci* 20:379–389. <https://doi.org/10.1016/j.jssas.2021.04.005>
- John K, Bouslih Y, Bouasria A, Razouk R, Hssaini L, Isong IA, M'barek AS, Ayito EO, Ambrose-Igho G (2022a) Assessing the impact of sampling strategy in random forest-based predicting of soil nutrients: a study case from Northern Morocco. *Geocarto Int*. <https://doi.org/10.1080/10106049.2022.2048091>
- John K, Bouslih Y, Isong IA, Hssaini L, Razouk R, Kebonye NM, Agyeman PC, Penizek V, Zádorová T (2022b) Mapping soil nutrients via different covariates combinations: theory and an example from Morocco. *Ecol Process* 11:23. <https://doi.org/10.1186/s13717-022-00368-y>
- Kalambukattu JG, Ghotekar KS, YS (2018) Spatial variability analysis of soil quality parameters in a watershed of Sub-Himalayan Landscape—a case study. *Eurasian J Soil Sci* 7(3):238–250
- Karlen DL, Andrews SS, Wienhold BJ, Doran JW (2003) Soil quality: human-kind's foundation for survival. *J Soil Water Conserv* 58:171–179
- Klute A, Dirksen C (1986) Hydraulic conductivity and diffusivity: laboratory methods. In: Klute A (ed) *Methods of soil analysis*, Part 1. Soil Science Society of America, Madison, pp 687–732
- Kuhn M, Johnson K (2013) *Applied predictive modeling*, vol 26. Springer, Berlin, Germany
- Lamichhane S, Kumar L, Wilson B (2019) Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review. *Geoderma* 352:395–413
- Li Y, Gao R, Yang R, Wei H, Li Y, Xiao H, Wu J (2013) Using a simple soil column method to evaluate soil phosphorus leaching risk. *Clean: Soil Air Water* 41:1100–1107
- Li X, Li H, Yang L, Ren Y (2018) Assessment of soil quality of croplands in the corn belt of Northeast China. *Sustainability* 10:248
- Lim TJ, Spokas KA, Feyereisen G, Novak JM (2016) Predicting the impact of bio-char additions on soil hydraulic properties. *Chemosphere* 142:136–144
- Liverpool SLO, Auchan AA, Banful AB (2010) An assessment of fertilizer quality regulation in Nigeria. The Nigeria strategy support program, Working papers.
- MacCarthy DS, Agyare WA, Vlek PL et al (2013) Spatial variability of some soil chemical and physical properties of an agricultural landscape. *West Afr J Appl Ecol* 21:47–61
- Mukherjee A, Lal R (2014) Comparison of soil quality index using three methods. *PLoS ONE* 9(8):e105981
- Nabiollahi K, Taghizadeh-Mehrjardi R, Kerry R, Moradian S (2017) Assessment of soil quality indices for salt-affected agricultural land in Kurdistan Province, Iran. *Ecol Indic* 83:482–494
- Nabiollahi K, Golmohamadi F, Taghizadeh-Mehrjardi R, Kerry R, Davari M (2018a) Assessing the effects of slope gradient and land use change on soil quality degradation through digital mapping of soil quality indices and soil loss rate. *Geoderma* 318:16–28
- Nabiollahi K, Taghizadeh-Mehrjardi R, Eskandari S (2018b) Assessing and monitoring the soil quality of forested and agricultural areas using soil-quality indices and digital soil-mapping in a semi-arid environment. *Arch Agron Soil Sci* 64:696–707
- Nelson DW, Sommers LE (1996) Total carbon, organic carbon and organic matter. In: Sparks DL (ed) *Methods of soil analysis*. Part 3. Chemical methods. SSSA Book Ser. 5. SSSA, Madison, pp 961–1010
- Njar GN (2018) Spatial pattern in solid minerals distribution in Cross River State, Nigeria. *J Appl Sci Environ Manag* 22(10):1661–1667
- Ofem KI, Asadu CLA, Ezeaku PI, John K, Eyon MO, Kateřina V, Karel N, Ondřej D, Vít P (2020) Genesis and classification of soils over limestone formations in a tropical humid Region. *Asian J Sci Res* 13:228–243
- Ofomata GEK (1975) *Nigeria in Maps*. Eastern States, Ethiopia Publication House, Benin City, Nigeria, pp 33–40
- Ogunwale JO, Obidike EO, Timm LC, Oduanze AC, Gabriels DM (2014) Assessment of spatial distribution of selected soil properties using geospatial statistical tools. *Commun Soil Sci Plant Anal* 45:2182–2200
- Okon PB, Nwosu NJ, Isong IA (2019) Response of soil sustainability indicators to the changing weather patterns in Calabar, Southern Nigeria. *Nigerian J Soil Sci* 29(1):52–61
- Paul GC, Saha S, Ghosh KG (2020) Assessing the soil quality of Bansloi river basin, eastern India using soil quality indices (SQIs) and Random Forest machine learning technique. *Ecol Indic* 118:106804
- Phogat VK, Tomar VS, Dahiya R (2015) Soil physical properties. In: *soil science: an introduction*, chapter 6. Indian Society of Soil Science, New Delhi, pp 135–171
- Pulido-Moncada M, Ball B, Gabriels D, Lobo D, Cornelis W (2015) Evaluation of soil physical quality index S for some tropical and temperate medium-textured soils. *Soil Sci Soc Am J* 79:9–19
- Quinlan JR (1992) Learning with continuous classes. In: *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*. Tasmania, Australia, 16–18 November 1992, pp 343–348
- Rezaee L, Akbar MA, Naser D, Sepaskhah AR (2020) Soil quality indices of paddy soils in Guilan province of northern Iran: spatial variability and their influential parameters. *Ecol Indic* 117:106566
- Santos-Francés F, Martínez-Graña A, Ávila-Zarza C, Criado M, Sánchez-Sánchez Y (2022) Soil quality and evaluation of spatial variability in a semi-arid ecosystem in a region of the Southeastern Iberian Peninsula (Spain). *Land* 11:5. <https://doi.org/10.3390/land11010005>
- Shekhovtseva OG, Mal'tseva IA (2015) Physical, chemical and biological properties of soils in the city of Mariupol, Ukraine. *Eurasian Soil Sci* 48:1393–1400
- Soil Survey Staff (2014) *Soil taxonomy: a basic systems of soil classification for making and interpreting soil surveys*, 12th edn. USDA, NRCS
- Udo EJ, Ibia TO, Ogunwale JO, Ano AO, Esu I (2009) *Manual of soil plant and water analysis*. Sibon Books Ltd, Lagos

- Uko AE, Effa EB, Isong I (2019) Performance of Mungbeans (*Vigna radiata* (L) Willczek) in soil amended with oil palm bunch ash and poultry manure in humid tropical environment of South Eastern Nigeria. *Int J Plant Soil Sci* 27(3):1–11. <https://doi.org/10.9734/IJPSS/2019/v27i330078>
- USDA-NRCS (Natural Resources Conservation Service) (2001) Soil quality test kit guide. [http://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/nrcs142p2\\_050956.pdf](http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_050956.pdf). Accessed 15 July 2020
- Vasu D, Singh SK, Ray SK, Duraisami VP, Tiwary P, Chandran P, Nimkar AM, Anantwar SG (2016) Soil quality index (SQI) as a tool to evaluate crop productivity in semi-arid Deccan Plateau, India. *Geoderma* 282:70–79
- Vomocil JA (1965) Porosity. In: Black CA (ed) *Method of soil analysis*, Part 1. American Society of Agronomy, USA, pp 659–662
- Wang YWI (1997) Inducing model trees for continuous classes. *Proceedings of the Ninth European conference on machine learning*. Springer, pp 128–137.
- Yemefack M, Jetten VG, Rossiter DG (2006) Developing a minimum data set for characterizing soil dynamics under shifting cultivation systems. *Soil till Res* 86:84–98
- Zeraatpisheh M, Ayoubi S, Sulieman M, Rodrigo-Comino J (2019) Determining the spatial distribution of soil properties using the environmental covariates and multivariate statistical analysis: a case study in semi-arid regions of Iran. *J Arid Land* 11(4):551–566
- Zeraatpisheh M, Bakhshandeh E, Hosseini M, Alavi SM (2020) Assessing the effects of deforestation and intensive agriculture on the soil quality through digital soil mapping. *Geoderma* 363:114139
- Zhang G, Bai J, Xi M, Zhao Q, Lu Q, Jia J (2016) Soil quality assessment of coastal wetlands in the Yellow River Delta of China based on the minimum data set. *Ecol Indic* 66:458–466
- Zounemat-Kermani M, Ramezani-Charmahineh A, Razavi R, Alizamir M, Ouarda TBMJ (2020) Machine learning and water economy: a new approach to predicting dams water sales revenue. *Water Resour Manag* 34:1893–1911

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---