

SPATIAL VARIABILITY OF SOIL pH AND CEC USING PREDICTIVE MODELS IN THE MOUNTAINOUS REGION OF OBANLIKU, CROSS RIVER STATE, NIGERIA

Mary A. Akpan, Sunday M. Afu, *Denis M. Olim and Fidelis V. Aberagi

Department of Soil Science, University of Calabar, Calabar.

*Corresponding author's email: olimdenis001@gmail.com

ABSTRACT

This study predicted the distribution pH and CEC in Obudu Cattle Ranch, Obanliku Local Government Area of Cross River State. Sixty (60) composite soil samples were collected (0-30 cm) between 200 to 500 m apart for the study. Normalized difference moisture index (NDMI), land surface temperature (LST), normalized difference vegetation index (NDVI), soil adjusted vegetation index (SAVI), clay, soil organic carbon and pH were used as covariates. Multiple linear regression (MLR), random forest (RF), ordinary kriging (OK), cubist regression (CR) and regression kriging (RK) were used to predict soil pH and CEC and were evaluated using bias, coefficient of determination (R^2), and root mean square error (RMSE), mean square error (MSE) and Lin's concordance correlation coefficient (CCC). The soils had pH varying from strongly acidic (pH=5.2) to slightly acid (pH=6.6) with high OC (>2 %). CEC of the soil was high (25 to 40 cmol/kg). CEC model estimated using OK had strong degree of spatial dependence while pH model had a moderate degree of spatial dependence. CEC predicted models showed almost similar spatial pattern with high CEC found in western part of the study. The highest predicted pH values dominated western part while the least predominated eastern parts of the study area in all the models. Cubist model better predicted CEC (CCC=0.605; MSE=28.79; RMSE=5.366) followed by MLR ($R^2=0.552$) than other models whereas pH was better modeled by OK (CCC=0.636; MSE=0.103; RMSE=0.322) followed by RF (RF=0.664). Cubist and OK should be used for assessment of spatial variation of CEC and pH in mountainous areas.

Keywords: Soil pH, CEC, spatial variation, prediction

Introduction

Mountainous areas account for about 20 percent of the world's terrestrial land area and, are home to about 10 percent of the world's population (WMV, 2011). The African continent alone occupies approximately 11 % of the world's mountains, and 20 % of Africa's total land area is made up of mountains (Nsengiyumva, 2019). Within the African continent, the population of people who live and farm in the mountainous areas is approximately 288

million (Cuni-Sanchez *et al.*, 2025), yet the soils are poorly studied placing them at risk of degradation due to poor management which reduces their agricultural potential. Presently, the global food supply is threatened because of rapid decline in the fertility of mountain soils as they host 25 percent of terrestrial biodiversity including agro-biodiversity playing a crucial role in ensuring security of nutritious food to 900 million people (FAO, 2015a).

Sub-Saharan Africa including Nigeria is known for its rich biodiversity and agricultural potential yet, the soils of its mountainous terrain present challenges for sustainable land use. Intensification of soil studies in mountainous areas is therefore imperative for informed land use and management. Studies on the mountainous soils presented contrasting views about their agricultural potential. WMV (2001) posited that Africa's mountainous soils are more fertile and tend to have intensive use for agriculture due to their favorable environmental conditions and greater resource potential than the surrounding soils. The soils are often centers of food and shelter materials production due to higher rainfall and high-quality agricultural land (FAO, 2015b; Cuni-Sanchez *et al.* 2016). Contrarily, FAO (2015a) opined that about 45 percent of mountain soils are either not suitable, or only marginally suitable for crop production due to shallowness caused by erosion, landslide and steep slopes. The soils are poorly developed, skeletal, shallow, acidic and relatively infertile (FAO, 2015a). This implies that cultivating these marginal soils without good management practices can exacerbate their deplorable condition leading to rapid decline in their fertility potential (FAO and ITPS, 2015).

Soil pH and CEC are key soil fertility indicators that determine soil productivity due to their influence on nutrient availability, plant growth and overall soil health. Soil pH determines the soil reactions and processes (Javed *et al.*, 2022), controlling cation and anion exchange reactions, nutrient and pollutant availability and organic matter decomposition (Chalchissa and Kuris, 2024;

Jones and Jacobsen, 2005). On the other hand, CEC is the sum of total exchangeable cations that the soil can adsorb; it is a measure of soil fertility, nutrient retention capacity and the capacity to protect groundwater from cation contamination and is capable of buffering fluctuations in the nutrient availability and soil pH (Javed *et al.*, 2022). Due to the importance pH and CEC in soil fertility and productivity, extreme cases of these soil properties severely affect soil productivity. For instance, extremely low pH condition can facilitate toxic solubility and availability of micronutrients such as iron, copper, boron, zinc, manganese etc. while extremely high pH causes toxic availability of nutrient like molybdenum (Foth, 2006) which usually results in a complete change in soil chemistry.

Despite the potential and limitations of mountainous soils as have been reported, the spatial distribution of soil pH and CEC in mountain soils in Nigeria remains poorly studied and understood. While several machine learning techniques have been used in prediction of soil properties (Mohammad *et al.*, 2010; John *et al.*, 2020), little or none of such studies have been conducted on mountain soils and, most of these studies focused on organic carbon, salinity and plant cover mapping using support vector machine, artificial neural network, kriging, cokriging and inverse distance weighing. Nonetheless, Ballabio (2009) argued that due to the overwhelming influence of topography on soil forming factors including climate and vegetation, digital soil mapping in mountainous areas can be best carried out using support vector regression. Consequently, there is need to compare the

performance of different machine learning techniques such as multiple linear regression, random forest, cubist regression, ordinary kriging and regression kriging in mapping of pH and CEC in mountainous terrain.

Farmers and inhabitants of mountainous areas of Obudu cattle ranch manage their soils as homogeneous entity regardless of the influence of elevation, slope gradient, climate and erosion in modifying soil properties. This practice often results in wrong application of agricultural inputs and poor soil management. Given the need for sustainable land use and crop production in mountainous areas, spatial mapping of soil properties has become increasingly important. Spatial mapping of soil properties enhances optimization of agricultural production as it guarantees precise application of inputs and adoption of soil management techniques based on field need. The aims of this study were to assess the spatial variability of soil pH and CEC, and evaluate the performance of multiple linear

regression, random forest, cubist regression, ordinary kriging and regression kriging in mapping of pH and CEC in mountainous terrain of Obudu cattle ranch of Cross River State, Nigeria.

Materials and methods

Location of the study area

The study was conducted in Obudu Cattle Ranch, Obanliku Local Government Area of Cross River State. The Obudu Cattle Ranch lies geographically between latitude 6° 21' N – 6° 24' N and longitude 9° 22' E – 9° 25' E in the tropical rainforest belt of Nigeria (Figure 1). This area has diverse soil type, land use and cover and topographic conditions with elevation ranging from 689 to 1654 m asl. Basement complex is the major parent material underlying soils in the area. The climate is typical moist tropical humid climate with rainfall and temperature ranges of 2000 to 3200 mm/annum and 15 to 31.80 °C.

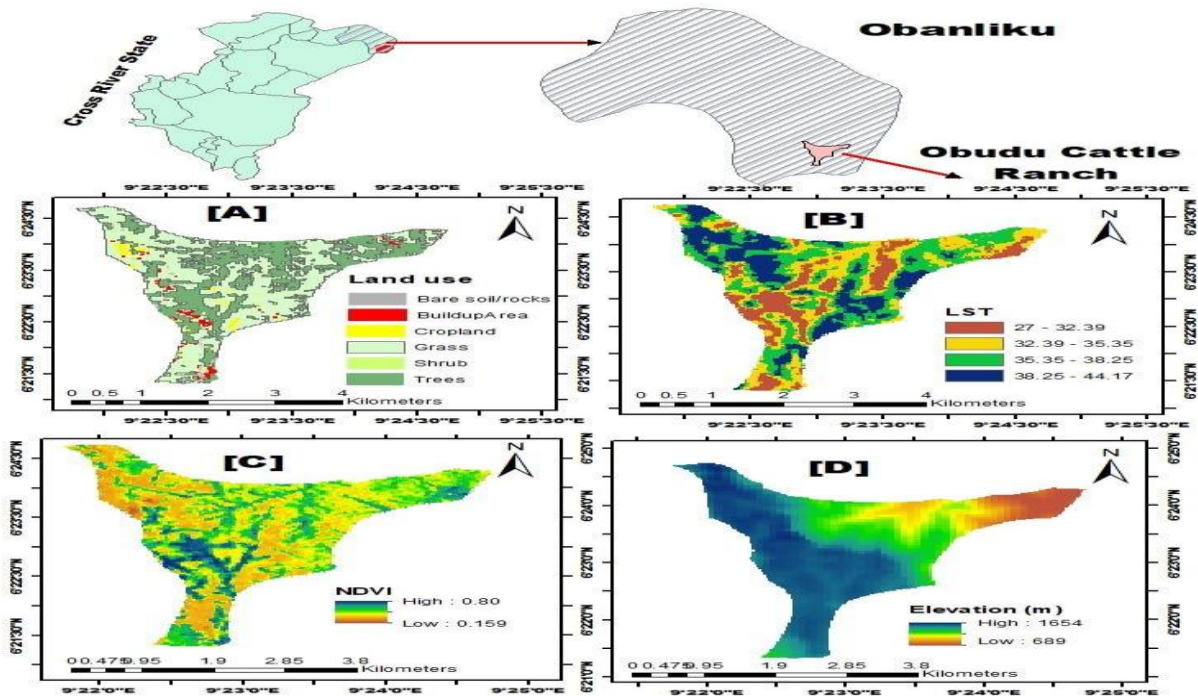


Figure 1: Map showing location of the study area and sampling points (LST=land surface temperature, NDVI= normalized difference vegetation index)

Soil sample collection and laboratory analysis

Sixty (60) georeferenced composite soil samples were obtained from Obudu Cattle Ranch in Obanliku Local Government Area of Cross River State for the study. Soil samples were collected with the use of soil auger at the depth of 0-30 cm using stratified random sampling technique. In each sampling point, four (4) sub-samples were collected randomly, hand mixed into a composite sample and placed in a well labeled polythene bag. The 60 composite soil

samples collected were transported to the laboratory and processed using standard procedures. The particle size distribution (sand, silt and clay) was determined using Bouyocous hydrometer method (Gee and Or, 2002). pH was obtained potentiometrically in soil:water suspension (1:2.5) using pH meter, while CEC was obtained using the methods outlined by Udo *et al.* (2009). Organic carbon was determined by Walkley-Black wet oxidation method using acid dichromate($K_2Cr_2O_7$) method (Nelson and Sommers 1996).

Environmental covariates

Environmental covariates used to derive environmental data were digital elevation

model (DEM) and Sentinel-2. Elevation, slope and aspect were gotten from DEM obtained at the space resolution of 30 m from

ASTER data (<https://earthexplorer.usgs.gov>) and were processed with the aid SAGA-GIS software terrain analysis toolbox. The European Space Agency's Copernicus Open Access was used to acquire Cloud-free Sentinel-2 imageries, processed using Google Earth Engine (GEE) to estimate land surface temperature (LST), normalized difference moisture index (NDMI), normalized difference vegetation index (NDVI), soil adjusted vegetation index (SAVI), canal network base level. Clay, soil organic carbon (SOC) and pH were obtained through interpolation techniques using the interpolated soil database at a resolution of 30 m. Climatic parameters were gotten from WorldClim database version 2 (Fick and Hijmans, 2017) and processed using ArcGIS software. The area of interest (AOI) for the soil and environmental data were demarcated using polygon feature of the study areas with the aid of ArcGIS 10.8 software (ESRI, Redlands, USA) environment.

Statistical analysis

Data generated were statistically analyzed for range, mean, standard deviation, coefficient of variability, skewness and kurtosis using SPSS v25. The statistical analyses for preprocessing environmental variables were done using SAGA-GIS

software terrain analysis toolbox. R and Rstudio software were used to carryout geostatistics and machine learning modeling and prediction. ArcGIS 10.8 was used for preparing maps.

Geostatistical interpolation method

Ordinary kriging (OK) was the geostatistical technique used in this study. The OK makes use of an assessed average of a given characteristic of soil in a location that is known to forecast the value of that sample characteristic in locations not sampled (Grunwald *et al.*, 2009) (Eqn. 1).

$$Z'(x_0) = \sum_{i=1}^n \lambda_i \cdot Z(x_i)$$

(1)

where:

$Z'(x_0)$ is the forecasted value for point x_0 , $Z(x_i)$ is the value known, and λ_i stands for kriging weight for the $Z(x_i)$ values. For each soil indicators, range, nugget and nugget ratio values were determined using semivariograms. Semivariance (h) is computed as half the average squared difference between the components of data pairs (Eqn. 2).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^n [Z(X_i) - Z(X_i + h)]^2$$

(2)

Where:

$\gamma(h)$ is the semivariance, h the lag distance, Z stands for soil property, $N(h)$ the number of pairs of areas disconnected by a lag distance h , $Z(x_i)$, and $Z(X_i+h)$ are values of

Z at positions x_i and $x_i + h$ according to Wang *et al.* (2013).

Machine Learning algorithms

Multiple linear regression (MLR), random forest (RF), cubist regression (CR) and regression kriging (RK) were used in this study to predict soil pH and CEC. In MLR, pH and CEC were forecasted as straight line combination of soil-environmental factors. The soil characteristic of interest was estimated using;

$$\hat{y}_{(i)} = \hat{\beta}_0 + \sum_{k=1}^k \hat{\beta}_k X_{k(i)} \quad (3)$$

Where, $\hat{y}_{(i)}$ is the estimated soil characteristic at point i , $\hat{\beta}_0$ the predicted intercept, $\hat{\beta}_k$ the predicted regression coefficient for predictor k and $X_{k(i)}$ the value for the k^{th} predictor at a given point i .

Random forest is a popular model used for regression and classification with complex environmental data sets (Freeman *et al.*, 2015; Fox *et al.*, 2020). RF provides several measures of variable importance that allow for interpretation of the fitted model (Hastie *et al.*, 2009). A collection of regression trees $\{T_k: k = 1, \dots, B\}$ each built from a bootstrap sample of the data set $\{Y, X\}$ defined RF. When growing each tree T_k , at each parent node a subset of m of the p predictor variables are randomly selected, and the best split-point is found among those m variables

to form two daughter nodes. A RF prediction at unvisited site with predictor values $x = (x_1, \dots, x_p)$ is found by averaging the predictions made by each tree in the ensemble:

$$f(x) = \left[\sum_{k=1}^K T_k(x) \right] / K \quad (4)$$

The cubist model was developed by Quinlan (1992) as a rule-based model which is an extension of the M5 tree model. CR is similar in approach to RF and its model design is made of parts or piecewise function which acts like choice making tree, joined together with MLR models (Kuhn, 2013). RK method is similar to MLR. It is soil prediction method that adds a regression of dependent variables on predictor quantities with kriging of the prediction residuals. RK prediction of $Y(s_0)$, at site not visited s_0 , is presented thus:

$$Z^*(x_0) = \hat{m}(x_0) + \hat{e}(x_0) = \sum_{k=0}^p \hat{\beta}_k * q_k(x_0) + \sum_{i=1}^n \lambda_i * e(x_i) \quad (5)$$

where $\hat{m}(x_0)$ is the fitted deterministic part, $\hat{e}(x_0)$ is the interpolated residual, $\hat{\beta}_k$ are the estimated deterministic model coefficients, λ_i are the kriging weights determined by the spatial dependence structure of the residual and $e(x_i)$ is the residual at position x_i .

Evaluation of model performance

Bias, coefficient of determination (R^2), and root mean square error (RMSE), mean square error (MSE) and Lin's concordance correlation

coefficient (CCC) were used to evaluate the performances of the models used in estimating the CEC and pH. The formulas are given as:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (Z_{oi} - Z_{pi}) \quad (6)$$

$$R^2 = 1 - \frac{\sum_i^n (Z_{oi} - Z_{pi})^2}{\sum_i^n (Z_{oi} - \bar{Z}_p)^2} \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_{pi} - Z_{oi})^2} \quad (8)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Z_{pi} - Z_{oi})^2 \quad (9)$$

$$\text{CCC} = \frac{2r\sigma_o\sigma_p}{\sigma_o^2 + \sigma_p^2 + (\bar{Z}_p - \bar{Z}_o)^2} \quad (10)$$

where, Z_{pi} = predicted values, Z_{oi} = observed values, n =observations size for the i -th term observation, \bar{Z}_i = average of the measured variable, σ_o^2 and σ_p^2 are the variances of the estimated and measured values; and r is the coefficient of correlation between the estimated and measured values.

Results and discussion

The results of the soil properties studied are shown in Table 1. Soil pH varied from strongly acidic (pH=5.2) to slightly acid (pH=6.6) with a mean of 5.3. The soils were high in organic carbon (mean OC > 2 %) (Landon, 1991). The results of CEC varied

from 15 to 44 cmol/kg with a mean of 25.9 cmol/kg and was rated high (25 to 40 cmol/kg) according to Landon (1991) while sand, silt and clay ranged from 49.2 to 73.2 %, 16 to 37 % and 6.8 to 26.8 % with mean values of 58.9 %, 26.6% and 14.5 % respectively. These results align with those reported in a related study by Afu *et al.* (2024) in the area. The values of coefficient of determination obtained showed that pH and sand were relatively homogenous, CEC, silt and clay were moderately variable while OC had heterogenous distribution in the soil (Essington, 2005). Strongly acid nature of the soils may be due to leaching of basic cations due to frequent rainfall that characterizes the Obudu cattle ranch. Strongly acid to slightly acid pH obtained in the soils coupled with high rainfall in the area may be responsible for high organic carbon content of the soils. Corroborating this view, Liu *et al.* (2025) found that soil pH decline promoted organic carbon accumulation by regulating microbial activities. Similarly, high CEC of the soils may be linked to high organic carbon content of the soils since both humic and non humic organic substances in soil exhibit cation exchange capacity through their charged surfaces (Tan, 2011).

Table 1: Summary statistics of studied soil properties

	pH	OC (%)	CEC (cmol/kg)	Sand → % ←	Silt	Clay
Min.	5.2	0.26	15	49.2	16	6.8
Max.	6.6	4.47	44	73.2	37	26.8
Mean	5.3	2.526	25.9	58.9	26.6	14.5
Std Dev.	0.438	0.96	6.701	6.47	5.353	3.567
CV	8	38	25.9	11	20.1	24.6
Skewness	1.239	-0.155	0.66	0.492	0.101	0.623
Kurtosis	0.399	-0.441	-0.026	-0.512	-0.703	1.532

Correlation of CEC and pH with soil covariates

Figure 2 shows bivariate linear correlation between CEC and pH with soil covariates. There was strong positive and significant relationship between CEC and OC ($r=0.75$, $p<0.05$) and mean rainfall ($r=0.25$, $p<0.1$), and weak negative and significant correlation between CEC and slope ($r=-0.15$, $p<0.1$), SAVI ($r=-0.28$, $p<0.1$). This result implies that an increase in OC and rainfall will correspondingly result to an increase in CEC and vice versa while an increase in slope and SAVI will result to a decrease in CEC. Similarly, soil pH showed a strong positive and significant relationship with slope ($r=0.34$, $p<0.05$), SAVI ($r=0.51$, $p<0.05$), NDVI ($r=0.39$, $p<0.05$), NDMI ($r=0.41$, $p<0.05$), mean temperature ($r=0.48$, $p<0.05$), and maximum temperature ($r=0.49$, $p<0.05$). Further result indicated that pH was a negatively and significantly correlated with LST ($r=-0.16$, $p<0.1$), CNBL

($r=-0.51$, $p<0.05$), elevation ($r=-0.57$, $p<0.1$), mean rainfall ($r=-0.15$, $p<0.1$). The result obtained for pH implies that an increase in slope, SAVI, NDVI, NDMI will lead to a corresponding increase in soil pH, while an increase in LST, CNBL elevation and mean rainfall will lead to a decrease in soil pH.

Covariate selection for estimation of CEC and pH using regression analysis

The soil properties used for correlation analysis were further subjected to regression analysis. The result obtained showed that the coefficient of OC (4.65) was positive and statistically significant (Table 2). This means that 1 % increase in soil organic carbon content will correspondingly increase CEC by 4.65 cmol/kg. The result revealed that other soil environmental variables did not exert significant impact on soil CEC in the study area (Table 2). Seyedmohammadi *et al.* (2016) found that CEC was strongly correlated with clay and SOC contents

($R^2 = 0.77$) in the soils of Guilan Province, Northern Iran. Similarly, for the estimation of soil pH, the coefficient of clay (0.03) was positive and statistically significant. This suggests that 1 % increase in clay content will correspondingly increase soil pH by 0.03 unit.

The high R^2 values of 0.66 and 0.33 for CEC and soil pH indicate a strong correlation between predictors and CEC and pH. However, due to multicollinearity problem between the target variable and predictors, variance inflation factor (VIF) was used to select the most stable variable in predicting

pH and CEC. Variable whose VIF was greater than 5 was eliminated. However, variable with $VIF < 5$ was selected for prediction of CEC and pH via machine learning models and geostatistics. For CEC model, minimum temperature (VIF = 5.79), mean rainfall (VIF = 6.06), NDVI (VIF = 9.75), NDMI (VIF = 1.84), clay (VIF = 1.35), OC (VIF = 1.81), and pH (VIF = 1.84) met this criterion. Similarly, for pH model minimum temperature (VIF = 5), mean rainfall (VIF = 5), elevation (VIF = 7.75), NDVI (VIF = 7.75), NDMI (VIF = 7.30), and clay (VIF = 1.19) met this criterion.

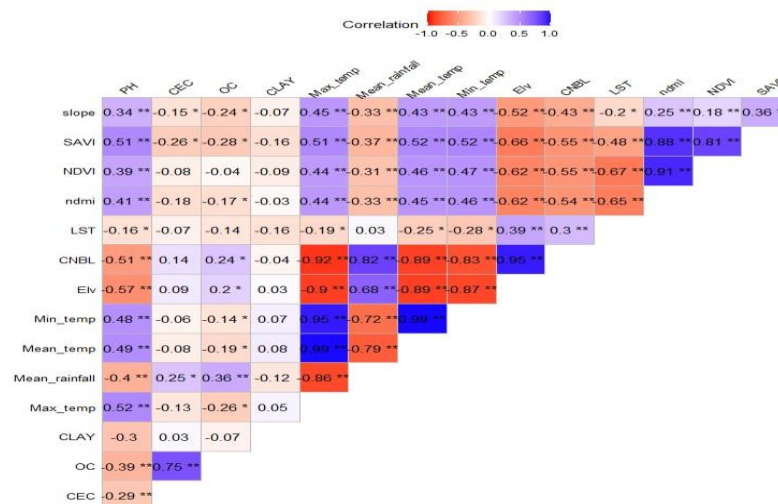


Figure 2: Correlation between CEC/pH and predictors variables

Table 2: Result of multiple linear regression

Variables	Coefficients	Std_Error	t value	Pr(> t)	VIF
CEC model					
Constant	-37.63	246.24	-0.15	0.88	
Min temp	-2.06	2.87	-0.72	0.48	5.79
Mean rainfall	0.01	0.14	0.08	0.94	6.06
Elevation	-0.02	0.02	-1.29	0.21	29.13
NDVI	1.15	17.79	0.06	0.95	9.75
NDMI	-15.21	22.86	-0.68	0.50	1.84
Clay	0.22	0.20	1.10	0.28	1.35
OC	4.65	0.88	5.29	0.00	1.81
pH	-1.46	2.16	-0.68	0.50	1.84
CNBL	0.12	0.18	0.65	0.52	31.77
pH model					
Constant	17.92	20.13	0.89	0.38	
Min temp	-0.13	0.24	-0.55	0.58	5.0
Mean rainfall	-0.01	0.01	-0.54	0.59	5.0
Elevation	-0.001	0.001	-1.03	0.31	7.75
NDVI	0.05	1.42	0.03	0.97	7.75
NDMI	0.44	1.86	0.24	0.81	7.30
Clay	0.03	0.02	2.06	0.05	1.19
CNBL	0.01	0.02	0.45	0.66	31.56
CEC_R ²	0.66				
CEC_Adjusted R ²	0.56				
CEC_F(9,32)	6.79, p = 0.00				
pH_R ²	0.33				
pH_Adjusted R ²	0.20				
pH_F(7,34)	2.43, p = 0.04				

NDMI=normalized difference moisture index, NDVI=normalized difference vegetation index, CNBL=channel network base level, CEC=cation exchange capacity

Spatial despondency of measured CEC and pH

The results in Table 3 show the parameters of measured CEC and pH model through semivariogram (Figures 3 and 4). The results implied that soil CEC was best modelled by exponential model while soil pH was best modelled by spherical model. This finding confirms the report of John *et al.* (2019) whose report showed that pH and SOC are best modelled using spherical models. CEC

had a Nugget/Sill ratio of 12.3 whereas soil pH had a Nugget/Sill ratio of 39.52 (Table 3). CEC model estimated using ordinary kriging (OK) had strong degree of spatial dependence (Nugget/Sill ratio < 25 %) while pH model estimated using ordinary kriging (OK) had a moderate degree of spatial dependence (Nugget/Sill ratio lies between 25 and 75%). A lower Nugget/Sill ratio for CEC suggests that structural factors such as climate, parent material, topography, soil

properties and other natural factors, play a significant role in spatial variability of CEC in the study area. However, the nugget/Sill ratio for pH suggest that spatial variability in pH is mainly caused by stochastic factors, including fertilization, farming management practices and other human activities. According to Kerry and Oliver (2004) the

higher Nugget/Sill connotes that spatial variability is mainly caused by stochastic factors, including fertilization, farming management practices and other human activities. Like in this study, John et al. (2019) obtained moderate spatial autocorrelation for pH in their studies.

Table 3: The Semivariogram parameters of measure CEC and pH

Variable	Model	Nugget (C_0)	Partial Sill (C_1)	Sill (C_0+C_1)	Range (m)	Nugget/Sill	Spatial class
CEC (cmol/kg)	Exponential	5.71	40.89	40.6	673.71	12.3	Strong
pH	Spherical	0.0423	0.065	0.107	1040.22	39.52	Moderate

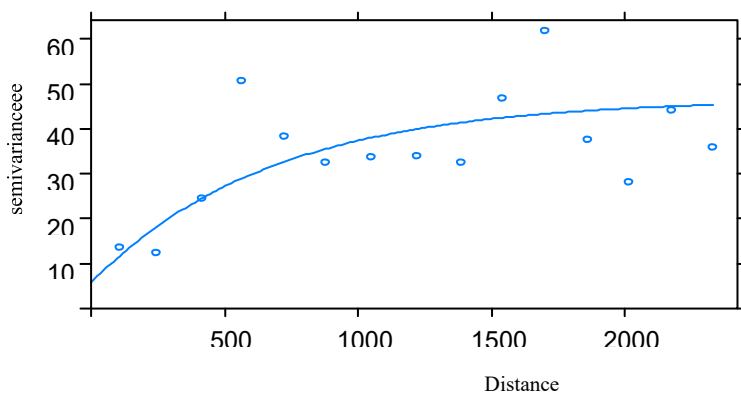


Figure 3: Semivariogram model for measured soil CEC

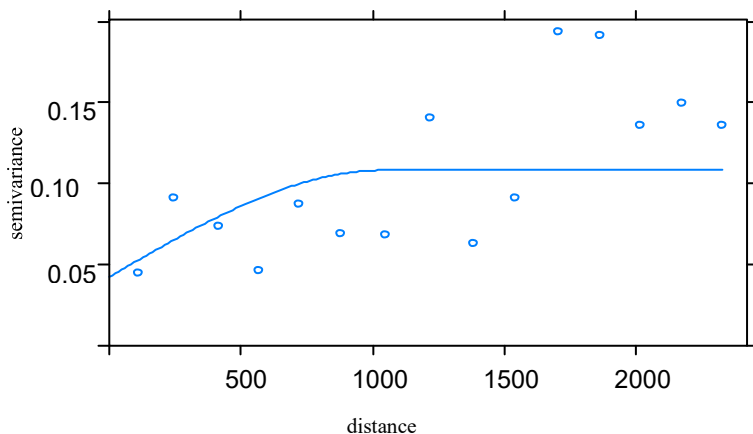


Figure 4: Semivariogram model for measured soil pH

Spatial prediction and mapping of soil CEC and pH

The spatial distribution of soil CEC and pH predicted by RF, cubist, RK, OK, and MLR are illustrated in figures 5 and 6 and their summary statistics are presented in Table 4. The mean laboratory determined values of CEC and pH are 25.9 cmol/kg and 5.44. The mean predicted CEC for the entire study area were 25.55 cmol/kg (RF), 25.38 cmol/kg (OK), 25.75 cmol/kg (RK), 25.79 cmol/kg (cubist) and 25.29 cmol/kg (MLR). Similarly, the mean predicted pH for the study area were 5.40 (RF), 5.47 (OK), 5.39 (RK), 5.40 (cubist) and 5.46 (MLR) (Table 4). Both laboratories determined or measured CEC and OC predicted values in all the models were within similar range. The maximum predicted CEC values were dominant in the northern and southern part of the study area in RK and cubist prediction map, whereas in RF and OK prediction maps it was almost the same distribution in the entire study area. In MLR, high values were mostly found in the northern part of the study area. Similarly, the minimum predicted CEC

values were prevalent in the central part of the study area in RK, MLR, RF and cubist prediction maps, whereas in OK prediction map it was patchy across the entire study area. Conversely, the highest predicted pH values were dominant western part while the least predominated eastern parts of the study area in all the models (Figures 5 and 6).

The spatial pattern of predicted CEC showed that moderate to high CEC (15 to >25) (Landon 1991) dominated majority of the study area. Further result showed that in pH prediction map, spatial pattern distribution of pH in all models revealed that strongly acidic pH (pH < 5.5) predominated the study area. The predicted results of CEC obtained in this present study are quite in disagreement with those of Kaya *et al.* (2023) whose predicted CEC values using RF were in the range of 5-15 cmol/kg in their study in Nagaland, India. However, observation similar to this study has been reported by Mishra *et al.* (2022) in their study on soils of different land uses in India, Sudan, Italy, Iran and Senegal.

Table 4: Determined and predicted soil CEC (cmol/kg) and pH

Laboratory determined CEC	16.25	44	25.9
CEC model			
RF predicted CEC	17.07	37.58	25.55
OK predicted CEC	16.66	35.99	25.38
RK predicted CEC	15.03	38.48	25.75
Cubist predicted CEC	13.93	36.32	25.79
MLR predicted CEC	15.12	34.31	25.29
Laboratory determine pH	5.0	6.6	5.44
pH model			
RF predicted pH	5.09	6.18	5.40
OK predicted pH	5.10	6.18	5.47
RK predicted pH	4.93	6.34	5.39
Cubist predicted pH	5.04	6.21	5.40
MLR predicted pH	5.01	6.02	5.46

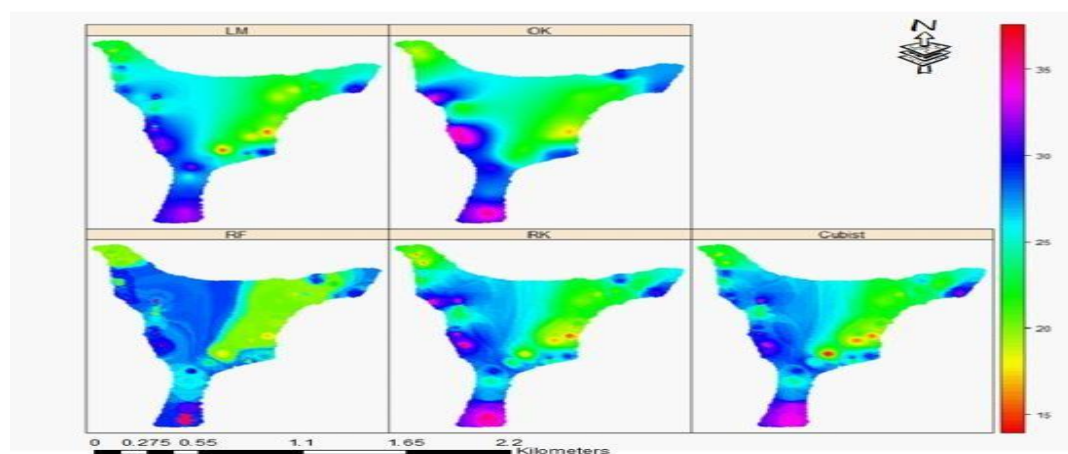


Figure 5: Spatial distribution of soil CEC by different machine learning models

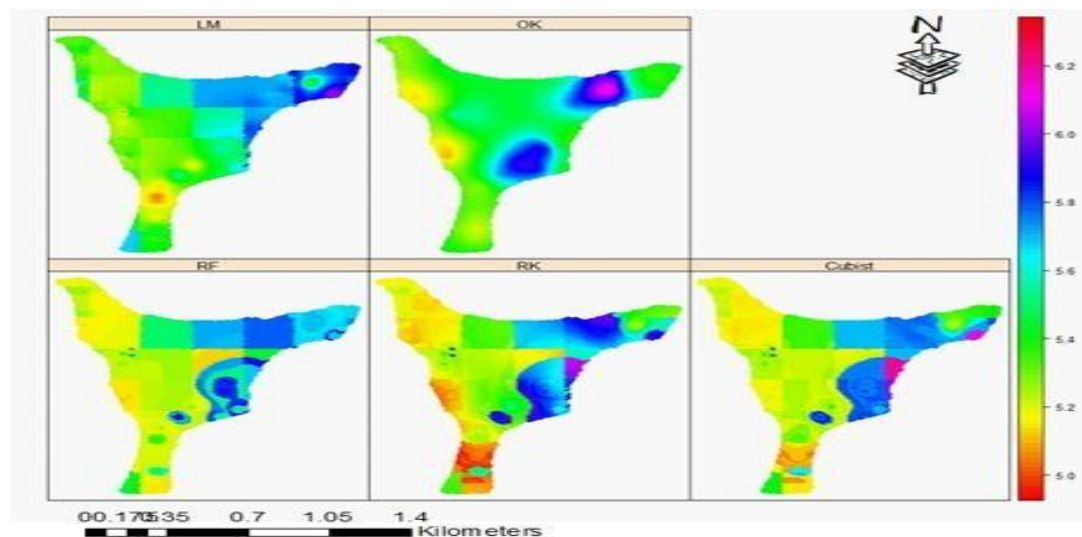


Figure 6: Spatial distribution of soil pH by different machine learning models

Evaluation of the competence of machine learning algorithms in mapping CEC and pH

The average MAE, RMSE, R^2 , and CCC for CEC prediction by cross-validation are shown in Table 5. The ML models showed different abilities to predict CEC and pH at unsampled locations in the study area probably due to various mathematical functions of each algorithm and covariates used for fitting. Prediction values of CEC using RF, OK, RK, cubist and MLR indicated that there are discrepancies in model performances. The MLR had the highest coefficient of determination ($R^2 = 0.552$) indicating very high precision while cubist consistently had the highest CCC (0.605), implying good agreement with the 45° line, as well as the lowest root mean squared error (RMSE = 5.366) and mean

square error (MSE = 28.796) signifying high accuracy, with random forest (RF) having the lowest bias (-0.458) (Table 5) which indicate no or little tendency for over- or underestimation.

The Cubist outperformed other models in prediction of CEC. Cubist regression lines of observed against predicted are closer to the 1:1 line compared to other models. RK predicted CEC well but there was over-prediction of low values and under-prediction of high values (Figure 7). Furthermore, RF, OK MLR all showed high tendency for overestimation or underestimation compared with other models. For example, OK model overestimate low values extremely and underestimate high values of CEC (Fig. 7) as

shown by 1:1 regression line (45° line). Overall results considering all validation indices showed OK model to be the least performed model while Cubist had the highest performance and was the best model for prediction of CEC. Observation similar to this present study was made by Kashi *et al.* (2014) who reported high performance of MLR ($R^2 = 0.77$; MAE = 1.85; RMSE= 1.92) in prediction of CEC using bulk density, soil texture, electrical conductivity, lime percentage, and sodium adsorption ratio in soils of the Ghoshe Region in Semnan Province, Iran. However, contrary to this study, a recent study by Kaya *et al.* (2023) reported better performance of RF for CEC mapping with highest R^2 (0.41) compared to other models. Reddy and Das (2023) also

reported higher prediction of CEC with RF in their studies in India.

From figure 8 and models evaluation results in Table 5, pH was better modeled by OK (CCC=0.636; MSE=0.103; RMSE=0.322) followed by RF (RF=0.664) than other models. In OK model the regression lines of observed against predicted are closer to the 1:1 line compared to other models (Figure 8). The least performed model for pH prediction was MLR as indicated by lower CCC (0.520) and rather high MSE (0.129) and RMSE (0.360). These findings are incongruent with those of Olorunfemi *et al.* (2016) concluded that MLR model was the best one for predicting pH in the forest soils of Nigeria ($R^2 = 0.71$; MAE = 1.16).

Table 5: Performance of machine learning algorithms in predicting soil CEC and pH

	R^2	CCC	MSE	RMSE	Bias
CEC model					
RF	0.467	0.603	31.574	5.619	-0.458
OK	0.046	0.169	62.946	7.933	-0.799
RK	0.442	0.561	33.279	5.789	-0.595
Cubist	0.532	0.605	28.796	5.366	-0.505
MLR	0.552	0.571	29.931	5.471	-0.881
pH model					
RF	0.664	0.553	0.129	0.359	-0.17
OK	0.592	0.636	0.103	0.322	-0.081
RK	0.548	0.625	0.117	0.342	-0.121
Cubist	0.533	0.604	0.121	0.347	-0.119
MLR	0.650	0.520	0.129	0.360	-0.141

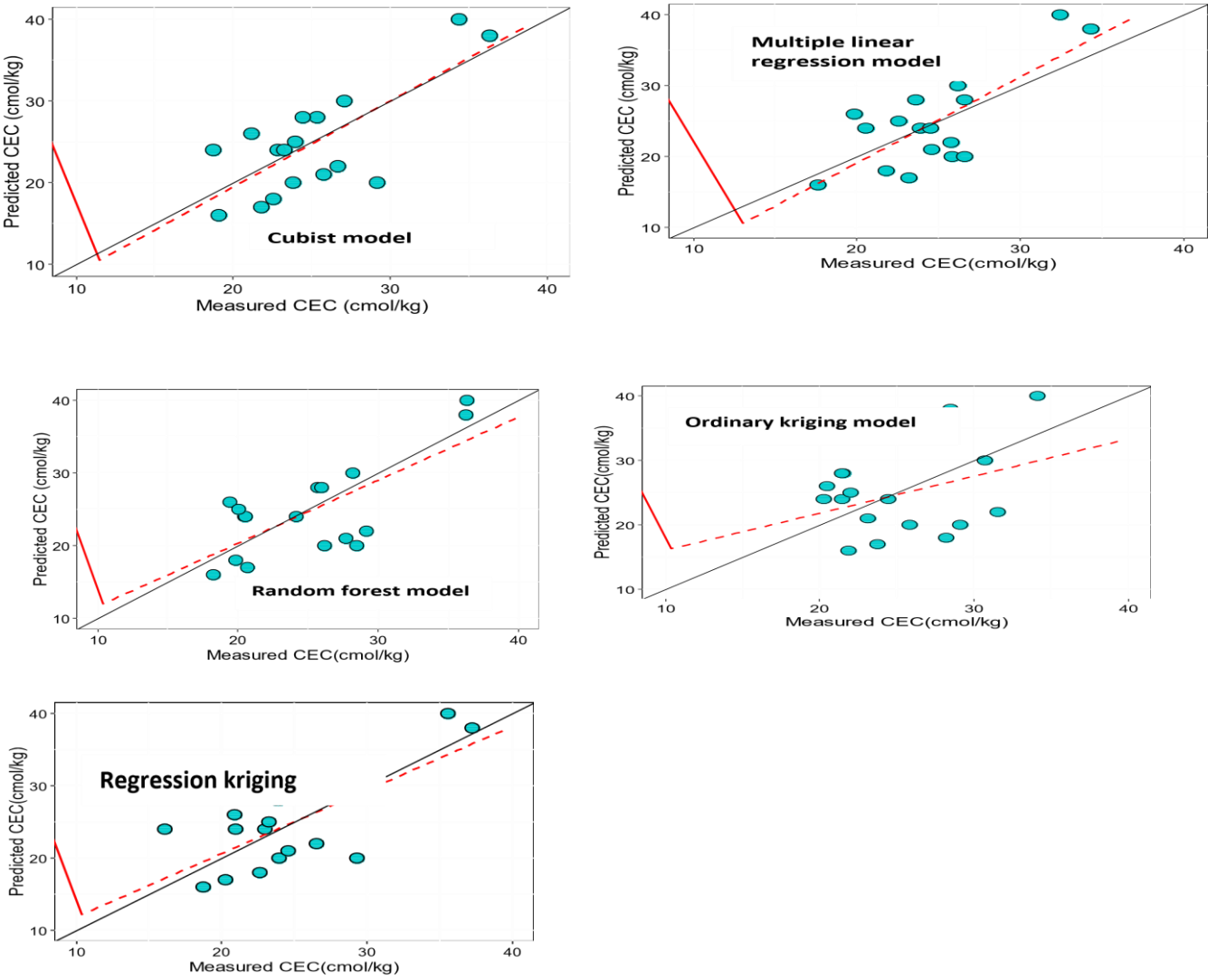


Figure 7: Measured and predicted values of CEC using five machine learning algorithms

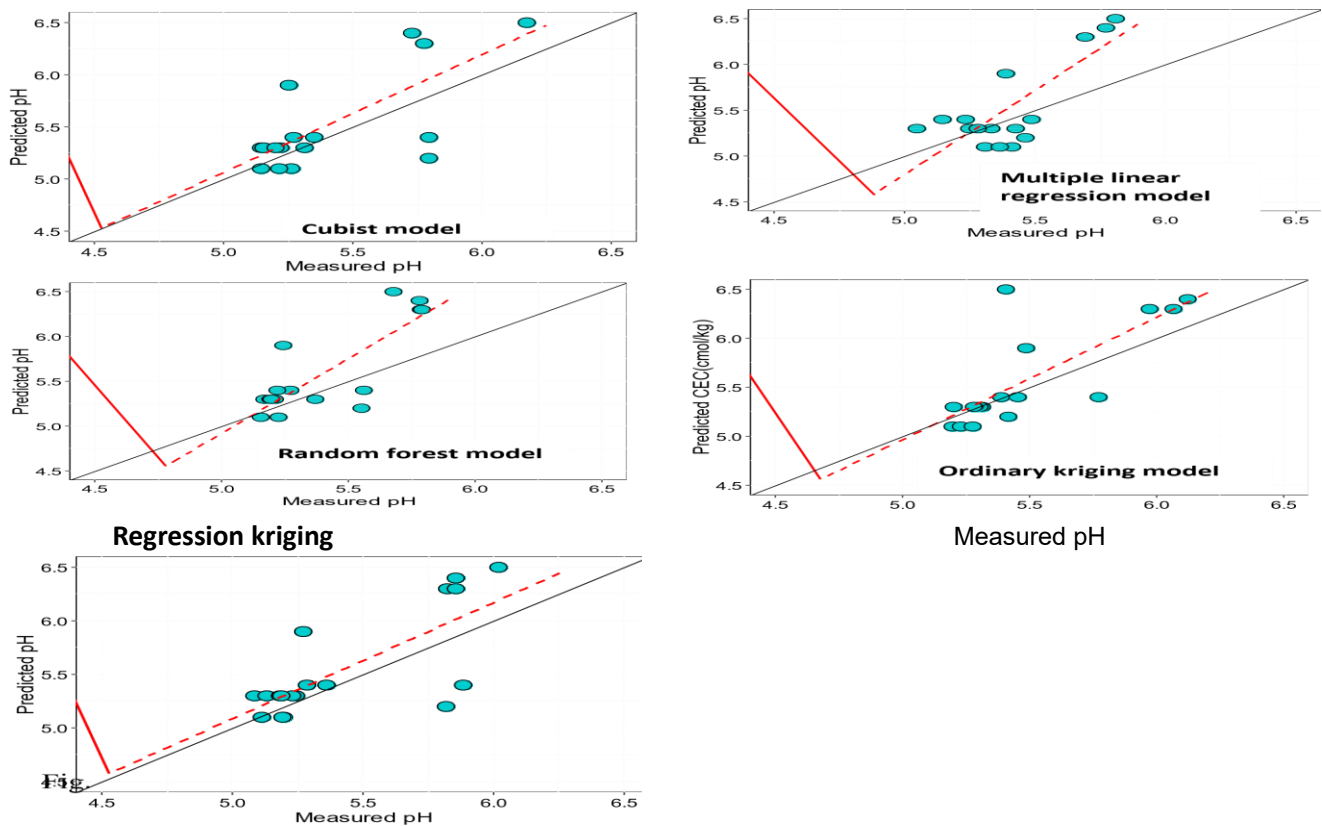


Figure 8: Measured and predicted values of soil pH using five machine learning algorithms

Summary and recommendation

The study was aimed to predict CEC and pH in the mountainous areas of Obudu cattle ranch. The results obtained revealed that the soils had strongly acid pH, high OC and CEC. Both determined or measured CEC and OC predicted values in all the models were within similar range. The CEC estimated through OK had strong degree of spatial dependence while pH had moderate spatial dependence. The maximum predicted CEC values were dominant in the northern and southern part of the study area in RK and CR models, whereas in RF and OK, CEC had

almost the same distribution in the entire study area. In MLR, high values were found in the northern part of the study area. Similarly, the minimum predicted CEC values were prevalent in the central part of the study area in RK, MLR, RF and CR maps, whereas in OK prediction map it was patchy across the entire study area. Conversely, the highest predicted pH values were dominant western part while the least predominated eastern parts of the study area in all the models. Cubist model better predicted CEC (CCC=0.605; MSE=28.79; RMSE=5.366) followed by MLR

($R^2=0.552$) than other models whereas pH was better modeled by OK (CCC=0.636; MSE=0.103; RMSE=0.322) followed by RF (RF=0.664). Therefore, cubist and OK should be used for assessment of spatial variation in CEC and pH in mountainous areas.

References

- Afu, S. M., Olim, D. M., Afangide, A. I., Ediene, V. F., Akpama, A. I. & Bisong, S. B. (2024). Assessment of soil quality for sustainable land management using machine learning and digital soil mapping techniques in Obudu Cattle Ranch, Nigeria. *Global Journal of Agricultural Sciences*, 23(1):71-83
- Ballabio, C. (2009). Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma*. 151(3-4):338-350
- Chalchissa, F. B. & Kuris, B. K. (2024). Modeling the impacts of extreme climate scenarios on soil acidity (pH and exchangeable aluminum) in Abbay River Basin, Ethiopia. *Heliyon*, 10(12). e33448.
- Cuni-Sanchez, A., Pfeifer, M., Marchant, R. & Burgess, N. D. (2016). Ethnic and locational differences in ecosystem service values: Insights from the communities in forest islands in the desert. *Ecosystem Services*, 19:42–50.
- Cuni-Sanchez, A., Aneseyee, A. B., Baderha, G. K. R., Batumike, R., Bitariho, R., Imani, G., Jha, N., Kaganzi, K. R., Kaplin, B. A., Klein, J. A., Leite, A., Marchant, R. A., Martin, E.H., Mcharazo, F., Mwangi, B., Ngute, A. S. K., Nkengurutse, J., Nkurunziza, A., Olaka, L., Soromessa, T., Tchoffo, R. O. K., Thorn, J. P. R., Twinomuhangi, I., Sullivan, M. J. P., & Zafra-Calvo, N. (2025). Perceived climate change impacts and adaptation responses in ten African mountain regions. *Nature Climate Change*.
- Essington, M.E. (2006). Soil and water chemistry: An integrative approach. CRC Press, New York.
- FAO (2015a). Understanding mountain soils. A contribution from mountain areas to the International Year of Soils
- FAO (2015b). Mapping the Vulnerability of Mountain Peoples to Food Insecurity. Authors: Romeo R, Vita A, Testolin R, Hofer T. Rome, Italy: FAO
- FAO & ITPS (2015). Status of the World's Soil Resources (SWSR)-Main Report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils. Rome, Italy.
- Fick, S. E. & Hijmans, R. J. (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. [International Journal of Climatology](https://doi.org/10.1002/gps.2753) 37 (12): 4302-4315.
- Foth, H.D. (2006). Fundamentals of soil science. 8Ed. John Wiley and Sons. New York.
- Fox, E. W., Ver Hoef, J. M. & Olsen, A. R. (2020). Comparing spatial regression to random forests for large environmental data sets. *PLoS ONE* 15(3): e0229509.
- Freeman, E. A, Moisen, G. G, Coulston, J. W. & Wilson, B. T. (2015). Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, 45,1–17.
- Gee, G.W. & Or, D. (2002). Particle Size Analysis. In: Methods of Soil

- Analysis, Part 4, Physical Methods, Dane, J.H. and G.C. Topp (Eds.). ASA and SSSA, Madison, WI., 255-293.
- Grundwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152 (3-4):195-207
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. *Springer Series in Statistics*. Springer, New York.
- Javed, A., Ali, E., Afzal, K.B., Osman, A. & Riaz, S. (2022). Soil Fertility: Factors Affecting Soil Fertility, and Biodiversity Responsible for Soil Fertility. *International Journal of Plant, Animal and Environmental Sciences*, 12, 021-033.
- John, K., Solomon, O. L., Ayito, O. E., Kebonye, M. N., Ogeh, J. S. & Vít, P. (2019). Predictive Mapping of Soil Properties for Precision Agriculture Using Geographic Information System (GIS) Based Geostatistics Models. *Modern Applied Science*, 13 (10), 60-77.
- Jones, C. & Jacobsen, J. (2005). Plant nutrition and soil fertility. *Environ. Sci.*, 2. 1-12
- Kashi, H., Emamgholizadeh S. & Ghorbani H. (2014). Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS models. *Communications in Soil Science and Plant Analysis*, 45(9), 1195–1213.
- Kaya, F., Mishra, G., Francaviglia, R. & Keshavarzi, A. (2023). Combining Digital Covariates and Machine Learning Models to Predict the Spatial Variation of Soil Cation Exchange Capacity. *Land*, 12, 819.
- Kerry, R. & Oliver, M. A. (2004). Average variograms to guide soil sampling for land management. *International Journal of Applied Earth Observation*, 5, 307-325
- Kuhn, N. I. (2013). Linking soil and sediment properties for research on biogeochemical cycles. *Geophysical Research Abstract*, 15
- Landon, J. R. (1991). Booker Tropical Soil Manual. A handbook for soil survey and agricultural land evaluation in the tropics and sub tropics. John Wiley and Sons, New York. 474.
- Li, C., Wang, X. & Qin, M. (2021) Spatial variability of soil nutrients in seasonal rivers: A case study from the Guo River Basin, China. *PLoS ONE*, 16, e0248655.
- Liu, J., Zheng, X., Luo, Q., Xiao, Q., Liu, S. & Oba, B. T. (2025). Soil pH decline promotes organic carbon accumulation by regulating the microbial carbon sequestration pathway. *Climate Smart Agriculture*. 2(3), 100068
- Mishra, G., Sulieman, M.M., Kaya, F., Francaviglia, R., Keshavarzi, A., Bakhshandeh, E., Loum, M., Jangir, A., Ahmed, I. & Elmobarak, A. (2022). Machine Learning for Cation Exchange Capacity Prediction in Different Land Uses. *Catena*, 216, 106404.
- Mohammad, Z. M., Taghizadeh-Mehrjardi, R. & Akbarzadeh, A. (2010). Evaluating geostatistical techniques for mapping spatial distribution of soil pH, salinity and plant cover affect by environmental factors in Southern Iran. *Notulae Scientiae Biologicae*, 2. 92-103
- Nelson, D. W. & Sommers, L. E. (1996). Total carbon, organic carbon and organic matter. In D.L. Sparks (Eds.). *Methods of soil analysis*. Part 3.

- Chemical methods. SSSA Book Ser. 5. SSSA, Madison, WI. 961-1010.
- Nsengiyumva, P. (2019). African mountains in a Changing Climate: Trends, Impacts, and Adaptation Solutions. *Mountain Research and Development*, 39(2):1-9.
- Olorunfemi, I., Fasinmirin, J. & Ojo, A. (2016). Modeling cation exchange capacity and soil water holding capacity from basic soil properties. *Eurasian Journal of Soil Science*, 5(4), 266– 274.
- Quinlan, J. R. (1992). Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Tasmania, Australia, 16–18 Nov. 1992; pp. 343–348.
- Reddy, N. N. & Das, B. S. (2023). Digital soil mapping of key secondary soil properties using pedotransfer functions and India Legacy Soil Data. *Geoderma*. 429.
- Seyedmohammadi, J., Esmacelnejad, L. & Ramezanpour, H. (2016). Determination of a suitable model for prediction of soil cation exchange capacity. *Modeling Earth Systems and Environment*, 2(3), 1–12.
- Tan, H. K. (2011). Principle of Soil Chemistry. 3rd Ed. Marcel Dekker Inc. New York.
- Udo, E. J., Ibia, T.O., Ogunwale, J.O., Ano, A.O. & Esu, I. (2009). *Manual of Soil, Plant and Water Analysis*. Sibon Books Ltd. Lagos.
- Wang, Y., Shao, M., Liu, Z. & Horton, R. (2013). Regional-scale variation and distribution patterns of soil saturated hydraulic conductivities in surface and subsurface layers in the loessial soils of China. *Journal of Hydrology*, 487, 13–23.
- WMC (World Mountain Conference) (2011). 20 years of sustainable mountain development in Africa. Lucerne, Switzerland. 11 and 12 October 2011