Earth Systems and Environment, Volume 5, Issue 4, December 2021, pp. 899-908
DOI:10.1007/s41748-021-00263-0

Estimating Soil Organic Matter: A Case Study of Soil Physical Properties for Environment-Related Issues in Southeast Nigeria

Kokei Ikpi Ofem[1] · Kingsley John[2] · Mark Pawlett[3] · Michael Otu Eyong[1] · Chukwuebuka Edwin Awaogu[4] · Pascal Umeugokwe[4] · Gare Ambrose-Igho[5] · Peter Ikemefuna Ezeaku[4] · Charles Livinus Anija Asadu[4]
[1]Department of Soil Science, University of Calabar, Calabar, Nigeria
[2]Department of Soil Science and Soil Protection, Czech University of Life Sciences, Prague, Czechia
[3]School of Agrifood and Environment, Cranfield University, Bedfordshire, UK
[4]Department of Soil Science, University of Nigeria, Nsukka, Nigeria
[5]Department of Geoscience, University of Nevada-Las Vegas, Las Vegas, NV, USA
*Corresponding author: johnk@af.czu.cz

Received: 15 August 2021 / Revised: 2 October 2021 / Accepted: 4 October 2021

Abstract

The different deposition periods in sedimentary geological environment have made the build-up and estimation of soil organic matter ambiguous to study. Soil organic matter has received global attention in the ambience of international policy regarding environmental health and safety. This research was to understand the inter-relationship between soil organic matter and bulk density, saturated hydraulic conductivity (Ksat), total, air-filled and capillary porosities for organic matter estimation, via different multiple linear regression functions (i.e., leap backward, leap forward, leapseq and lmStepAIC), in soils developed over the sedimentary geological environment. Eight mapping units were obtained in Ishibori, Agoi Ibami and Mfamosing *via* digital elevation model. Two pits were sited within each mapping unit, and 53 soil samples were used for the study. In soils over shale–limestone–sandstone, two pits were sited, six in alluvium, four in sandstone–limestone and four in limestone. Overall correlation between SOM with Ksat ($r = 0.626$) and BD ($r = -0.588$) was significant ($p < 0.001$). The strongest correlation was obtained for SOM with BD ($r = -0.783$) and Ksat ($r = 0.790$) in soils over limestone. In contrast, soils over shale–limestone and sandstone geological environment gave the weakest relationship ($r < 0.6$). Linear regression gave a similar prediction output. The best performing was leapbackward (RMSE = 11.50%, $R^2 = 0.58$, MAE = 8.48 %), which produced a smaller error when compared with leap forward, leapseq and lmStepAIC functions in organic matter estimation. Therefore, we recommend applying leapback linear regression when estimating soil organic variation with physical soil properties for solving soil–environmental issues towards sustainable crop production in southeast Nigeria.

Keywords Agriculture · Environment · Multivariate statistics · Soil health · Humid tropics

1. Introduction

Soil organic matter (SOM) is an essential component of the soil. It is pivotal for maintaining multiple soil-derived ecosystem services, such as the production of food and materials for shelter, fuel and clothing, the maintenance of biodiversity, and critically mitigating effects of global climate change (Li et al. 2017). In addition, it positively impacts soil fertility. It contains an unknown number of compounds derived from living and non-living organic substances, varying from easily decomposable simple organic materials to complex recalcitrant compounds and organisms (Kogel-Knabner 2002).

Besides sequestering or acting as a source or sink of atmospheric carbon, SOM storage in arable soils influences soil physical, chemical and biological properties (Saint-Laurent et al. 2017; Blanco-Canqui et al. 2013). These properties are exposed to more risks in cultivated soils. Land degradation occurs globally due to poor land management strategies, such as inappropriate land uses like bush burning, continuous cultivation and tillage (Blanco-Canqui et al. 2013). This results in a decline in SOM and concurrent impacts on soil physical parameterssuch as porosity, increased bulk density (BD) (Tisdall and Oades 1982) and reduced infiltration (Li et al. 2007) as they are functions of SOM (Jiao et al. 2020). Organic matter reduces soil BD and increases the porosity of compacted soil layers (Boni et al. 1994; Bonini and Alves, 2010), while its mineralization may lead to increased BD (Oliveira et al. 2018). In addition, researchers in Northern Karakoram (Ali et al. 2017) and Nepal (Ghimire et al. 2018) identified negative correlations between organic C and BD. In contrast and surprisingly, Lei et al. (2019) reported positive correlations between soil organic carbon (SOC) and BD in subsurface soils, while Masri and Ryan (2006) reported decreasing hydraulic conductivity with reduced SOM.

Several factors have been reported to affect the build-up of SOM. They include topography (Cardinael et al. 2017), climate (Munoz-Rojas et al. 2017), soil type (Zhao et al. 2016), soil depth, land use (Kafle, 2019), texture (Lei et al. 2019), soil microorganisms (Komarov et al. 2017) and soil pH (Zhou et al. 2020). When wholly considered, these factors make studies related to SOC complex and make its measurement and inter-relationship with other soil properties difficult. However, SOM is important in soil studies and maybe a sole indicator of fertile and healthy soil.

There have been several studies on the horizontal spatial distribution of SOM using various mathematical models as influenced by topography, vegetation and land use (Takata et al. 2007; Liu et al. 2015). Applying different machine learning (ML) in predicting soil properties is recent in soil science and precision agriculture (for example, random forest, support vector machine, artificial neural network and others) (John et al. 2020). Multiple linear regression (MLR) has been applied in modeling and predicting SOC via environmental variables and soil nutrient indicators (John et al. 2020), arsenic estimation via XRF and ICP-OES data (Kebonye et al. 2020), and the mapping of soils of Minas Gerais, Brazil via XRF data using the stepwise multiple linear regression techniques (Silva et al. 2017). However, the stepwise variable selection is automatic and has many statistical problems that could worsen if the covariates are collinear. Therefore, this study attempts to reduce covariates collinearity. Currently, no published studies compare the different stepwise linear regression functions in the modeling of SOM under diverse sedimentary geological environments; hence, this research introduces a new approach in explaining the variability of SOM in soils over the different sedimentary geological environments. We hypothesize that SOM will vary in its inter-relationship with soil physical properties in different sedimentary geological environments, and subsequently, SOM can be predicted by soil physical properties. Consequently, this research studied the inter-relationships between SOM and BD, saturated hydraulic conductivity (Ksat) and porosity, and applied various multiple linear regression functions to predict SOM accumulation via some selected soil physical properties dominating the different geological environments.

2. Materials and methods

2.1 Location and land use, geology, and climate of the study area

The study sites were located in Ishibori area (679 ha) of Ogoja (06o39′17" N, 08o47′51" E), Agoi Ibami (280 ha) in Yakurr (05o43′27"N, 08o10′37.2" E) and Mfamosing (2202 ha) in Akamkpa (05°04′41.8"N, 08° 27′49.8"E), all in the Cross River State of Nigeria. The Ogoja area is covered by the southern guinea savannah and cultivated to oil palm, teak and paddy rice, while the Yakurr and Akamkpa areas are covered by tropical rainforest. Common crops in the Yakurr and Akamkpa areas are oil palm, cassava and plantain.

Basement Complexes and Sedimentary Basins dominate the geology of Cross River State (Ekwueme 1987). The Sedimentary Basins, containing sediment fill of Cretaceous to Tertiary ages, dominate the Niger Delta region (Fatoye and Gideon 2013), with alluvium found in the low lying coastal areas. The limestone of the Cretaceous and Tertiary ages is often intercalated with shale, siltstone, and fine-grained sandstone (Ofem et al. 2020a).

Cross River State has a humid tropical climate, which varies from the southern guinea savannah in the Ogoja area to the tropical rainforest of Yakurr and Akamkpa. Consequently, rainfall fluctuates from 1251–3348 mm/year in the Ogoja area to 1760–2684 mm/year and 2109–3771 mm/year in Yakurr and Akamkpa, respectively (Sambo et al. 2016). Temperature varies from 23 to 34 °C in the Ogoja area and 23 to 32 °C in Yakurr and Akamkpa areas (Sambo et al. 2016).

Yakurr and Akamkpa have similar climates and vegetation and often experience slight temperature variation.

2.2 Field and laboratory procedures

Digital elevation models (DEM) of the study locations were acquired from USGS Explorer SRTM 1 arc-second Global at a resolution of 30 m. The DEM was employed to generate slope maps in ArcGIS (ESRI, US) environment. The elevation ranges created in the slope maps were used to delineate slope transition (Ofem et al. 2020a). Each of the eight slope transitions (IH1, IH2, AI1, AI2, AI3, MF1, MF2, MF3) represented a soil mapping unit (MU). Two soil pits were randomly sited in each MU and dug to represent the soils (2 m by 1.5 m by X m). Where X m represents variable depth to the water table or consolidated rock layer, this gave rise to sixteen pits in total, two in shale–limestone–sandstone (SLM) (IH1P1, IH1P2), six in alluvium (IH2P1, IH2P2, AI3P1, AI3P2, MF3P1, MF3P2), four in sandstone–limestone (AI1P1, AI1P2, MF1P2, MF2P2) and four in limestone (AI2P1, AI2P2, MF1P1, MF2P1). Thereby, a total of 53 soil samples were collected from pedogenic horizons and subjected to laboratory analyses. In addition, undisturbed core soil samples were vertically collected from pedogenic horizons for the determination of saturated hydraulic conductivity (Ksat), total porosity (Total_P), airfilled porosity (Air_P) and capillary porosity (CAP_P). Ksat was determined by the direct application of Darcy's equation to a saturated soil column of uniform cross-sectional area (SSS, 2014), such that:

$$Ksat = \frac{VL}{At(H_2 - H_1)} \qquad \text{(Equation 1)}$$

where V = volume of water that flows through the sample of cross-sectional area (A) in time (t); (H2-H1) = Hydraulic head difference; L = Length of sample.

Core soil samples were then drained at 60 cm of tension to determine Total_P, Air_P, and CAP_P. Total porosity, Air_P, and CAP_P were determined by dividing the volume of water in the soil at saturation, the volume of water drained at 60 cm of tension, and the volume of water retained at 60 cm of tension by the volume of the cylinder (Obi 2000).

Soil for organic carbon determination was air dried under room temperature in the laboratory at 29–30 °C for three days, ground with a wooden pestle to break peds and passed through a 2 mm sieve. Soil organic carbon (Walkley–Black modified acid-dichromate) was determined using standard procedures outlined in Soil Survey Staff (SSS 2014). SOM was calculated from SOC by multiplying by a factor of 1.72 to obtain SOM. Soil samples were analyzed in the Department of Soil Science, University of Nigeria, Nsukka. The field study was carried out between December 2018 and February 2019.

2.3 Correlation matrix

A simple correlation analysis was performed with categorical data (e.g., geological environment). This analysis explained the intra- and inter-relationships between the SOM and the selected physical properties and how the individual geological environment contribute to the relationship between SOM and the physical properties. The output of the correlation was reconfirmed through the application of a Principal Component Analysis (PCA).

2.4 Principal components analysis

PCA enabled the grouping of the selected soil properties into the different geological environments. It enabled the extraction of principal factors accounting for the sources of variation in the data (Belkhiri and Narany 2015) and to identify the geological material influencing SOM and other properties. Such litho-material would require further assessment as they may help explain certain SOM variability relating to the selected soil properties within the area.

2.5 Modeling Approach of SOM

Four (n = 4) stepwise multiple linear regression (MLR) functions were applied in this study. The forward, backward, both direction, and the regsubsets are available in the leap function. This study presented four functions available in R software for stepwise linear regression in estimating SOM using six predictors (BD, Ksat, Total_P, Air_P, CAP, geological material). The stepwise regression applied leaps and stepAIC functions available in R's leaps and MASS packages. The leaps package in R is composed of "leapBackward", which fits a linear regression with backward selection, and "leapForward", with fittings for linear regression with forward selection. The "leapSeq" fits a linear regression with stepwise selection, while in stepAIC (also referred to as direction), we applied the "lmStepAIC" (James et al. 2014). The approach was adopted to exhaustively establish that the intended selected model is suitable for SOM prediction in the soils overlying sedimentary geological environment in the region. The simple linear model used to predict SOM (%) via the selected soil properties is expressed as, thus:

$$SOM\ (\%) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j + \epsilon_j \qquad (2)$$

where β_0 is the y-intercept and or bias in the field of machine learning (Hastie et al. 2008). The X_j represents the predictor variable, while β_j is the slope coefficient of the predictor. An error term is also included and is denoted by ϵ_j.

2.6 Model Accuracy and Assessment

The entire data were subjected to modeling. Mean absolute error (MAE), and root mean square error (RMSE), and coefficient of determination (R^2) were adopted as criteria in evaluating the models' performance. In the case of MAE and RMSE, a lower value is preferred. For R^2, values closer to 1 (Li et al. 2016).

2.7 Statistical Analysis

The R software performed all statistical analyses and model computations (R Core Team 2019).

3 Results and discussion

The summary of descriptive statistics for the soils, grouped by the geological environment, is presented in Table 1. At the same time, the results of the interaction between SOM and physical properties are shown in Fig. 1.

Table 1: Summary of descriptive statistics for the soils studied

Statistics	SOC	SOM	BD	Ksat	Total P	Air_P	CAP_P
	g/kg		g/cm^3	cm/h		%	
Shale-limestone and sandstone intercalation (SLM) (IH1P1, IH1P2)							
Mean	9.86	16.94	1.57	34.99	51.41	8.97	42.41
Std	13.59	23.39	0.08	42.12	7.23	6.7	4.15
SE	5.14	8.84	0.031	15.92	2.73	2.53	1.57
Min	1.37	2.36	1.45	0.61	44.4	2.3	35
Max	40.3	69.32	1.66	106.28	65	21	47.8
CV	1.38	1.38	0.05	1.2	0.14	0.75	0.1
Alluvium (IH2P1, IH2P2, AI3P1, AI3P2, MF3P1, MF3P2)							
Mean	18.12	31.16	1.19	49.55	57.76	6.63	51.13
Std	23.24	39.98	0.32	85.44	10.01	5.67	10.49
SE	5.81	9.99	0.08	21.36	2.5	1.42	2.62
Min	1.03	1.77	0.53	0.49	45.1	2	38.1
Max	86.64	149.02	1.63	256.54	80.2	23.9	76.5
CV	1.28	1.28	0.27	1.72	0.17	0.86	0.21
Sandstone-limestone (SS) (AI1P1, AI1P2, MF1P2, MF2P2)							
Mean	7.01	12.06	1.43	43.67	48.8	4.71	45.07
Std	6.17	10.61	0.17	39.27	6.32	2.56	8.23
SE	1.54	2.65	0.043	9.82	1.58	0.64	2.06
Min	1.72	2.96	0.99	1.22	40.5	2.3	34.6
Max	21.96	37.77	1.66	126.67	63.4	11.3	60.7
CV	0.88	0.88	0.12	0.9	0.13	0.54	0.18
Limestone (LS) (AI2P1, AI2P2, MF1P1, MF2P1)							
Mean	7.72	13.28	1.35	16.67	49.56	5.3	44.13
Std	11.73	20.17	0.19	20.19	7.2	5.54	7.06
SE	3.13	5.39	0.051	5.4	1.92	1.48	1.89
Min	0.69	1.19	0.9	0.49	37.1	0.9	31.6
Max	46.34	79.7	1.6	75.52	62.2	23.5	59
CV	1.52	1.52	0.14	1.21	0.15	1.05	0.16

SOC: soil organic carbon, SOM: soil organic matter, BD: bulk density, Ksat: Saturated hydraulic conductivity, total P: total porosity, Air_P: Air-filled porosity, Cap_P: capillary porosity, IH1, IH2, AI1, AI2, AI3, MF1, MF2, MF3: soil mapping units

3.1 Inter-relationships Between SOM and BD, Ksat and Porosity in the Sedimentary Geological Environment

3.1.1 Soil Organic Matter Versus Bulk Density

SOM correlated moderately and negatively with BD ($r = -0.588$, $p < 0.01$) (Fig. 1) in the studied soils and indicated an increase in SOM with decreasing BD values. The highest mean value of SOM and the lowest mean value of BD were obtained in soils over alluvium (Table 1) and further revealed that poorly drained alluvial soils are better accumulators of SOM. High SOM values are most likely to result in low BD values. Similar positive relationships were reported by Tisdall and Oades (1982) and Rawls et al. (2005) and contradict findings by Oliveira et al. (2018) that BD is unaffected by green manure. Others argue that organic matter does affect BD (Heuscher et al. 2005). An increase in organic matter oxidation rate is most likely to increase soil BD; for instance, poorly drained soils rich in accumulated organic matter have low BD compared to well-drained soils located in the upland. Conversely, an increase in green manure or SOM reduces BD (Boni et al. 1994; Parihar et al. 2016). However, this negative relationship was strongest in soils over limestone with higher r values (> 0.70); especially those with Vertic properties as reported in Ofem et al. (2020a), and alluvium ($r = 0.578$), which had Loamic and Humic properties in the WRB system (Ofem et al. 2020a), and weakest in soils over SLM lithology.

3.1.2 Soil Organic Matter Versus Saturated Hydraulic Conductivity

Soil organic matter correlated moderately and positively with Ksat ($r = 0.626$) ($p < 0.001$) in the studied soils. Greater SOM results in higher Ksat because soil aggregate formation is linked to organic matter content (Beare et al. 1994). The presence of a considerable amount of organic matter ensures good aggregate and soil structural formation. This facilitates the movement of water through the soil. The highest value of SOM in soils over alluvium, which coincides with the highest value of Ksat, may further affirm their correlation. The soils over alluvium have either Aquic or Gleyic properties (Ofem et al. 2020a) expresses poorly drained soil conditions. Such conditions tend to encourage SOM deposition. Similar results have been reported, such that increased Ksat was obtained through an increase in dairy manure application (Jiao et al. 2006; Eghball, 2002), and SOM in the Mediterranean region (Masri and Ryan 2006). However, the relationship is not always a straight positive correlation for any soil (Nemes et al. 2005). This indicates that SOM is most likely to increase if soil conditions that favor increased Ksat are created. Masri and Ryan (2006) recommended a legume rotation for improved Ksat. Generally, significant amounts of readily decomposed organic matter and enhanced nutrient release from such materials may improve physical soil conditions (Sanchez et al. 1989). A high positive correlation ($r > 0.70$) was obtained between Ksat and SOM for soils over alluvium and LS, indicating greater certainty for the relationship than soils over SLM lithology. According to Saxton and Rawls (2006) and Yao et al. (2015), SOM is an important predictor of Ksat but strongly influenced by vegetations in the subtropics (Hao et al. 2019). For instance, irrespective of lithology, a higher mean value of 16.9 g/kg for SOM was obtained in the well-drained soils of the southern guinea savannah area compared to 12.06 and 13.28 g/kg obtained in the tropical rainforest. This variation may be connected to the huge accumulation of litter in the oil palm and teak plantations.

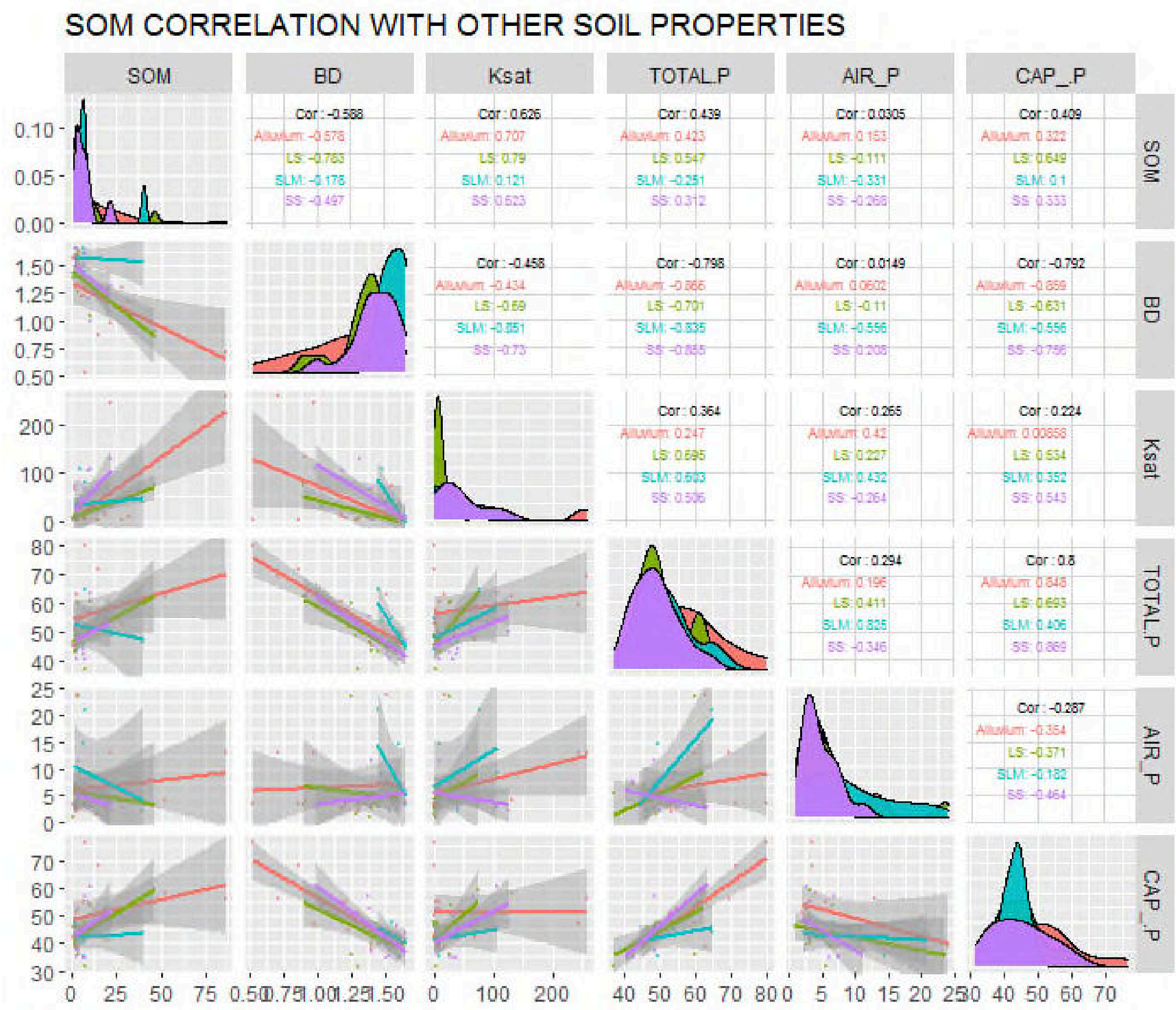

Fig. 1: Correlation of SOM with bulk density, saturated hydraulic conductivity and porosity

Gleyic properties (Ofem *et al.*, 2020a) which expresses poorly drained soil conditions. Such conditions tend to encourage SOM deposition. Similar results have been reported, such that increased Ksat was obtained through an increase in diary manure application (Jiao *et al.*, 2006, Eghball, 2002), and SOM in the Mediterranean region (Masri and Ryan, 2006). However, the relationship is not always a straight positive correlation for any soil (Nemes *et al.*, 2005). This indicates that SOM is most likely to increase if soil conditions that favour increased Ksat are created. Masri and Ryan (2006) recommended a legume rotation for improved Ksat. Generally, significant amounts of readily decomposed organic matter and enhanced nutrient release from such materials may improve soil physical conditions (Sanchez *et al*., 1989). High positive correlation ($r>0.70$) was obtained between Ksat and SOM for soils over alluvium and LS, indicating greater certainty for the relationship compared to soils over SLM lithology. According to Saxton and Rawls (2006), and Yao *et al*. (2015), SOM is an important predictor of Ksat, but strongly influenced by vegetations in the subtropics (Hao *et al*., 2019). For instance, irrespective of lithology, higher mean value of 16.9 g/kg for SOM was obtained in the well-drained soils of the southern guinea savannah area compared to 12.06 and 13.28 g/kg obtained in the tropical rainforest. This variation may be connected to the huge accumulation of litter in the oil palm and teak plantations.

3.1.3 Soil Organic Matter and Porosity

The correlation of SOM versus Total_P resulted in a moderate positive correlation ($r = 0.44$) ($p < 0.01$) in the studied soils and implies increased SOM with an increase in Total_P, but with moderate certainty in the positive relationship between SOM and Total_P within sedimentary formations. Tisdall and Oades (1982). Nemes et al. (2005) obtained similar relationships as reported in this study. Boni et al. (1994), Whalen and Chang (2002), and Alves and Suzuki (2004) reported an increase in Total_P by the use of green manure, dairy manure and successional cover crops. Similarly, Li et al. (2007) opinionated that a decrease in SOM will decrease porosity, reduced water and air storage. In soils over SS, the relationship was weak and positive. Soils over SS are high in the sand (Souza et al. 2019; Ofem et al. 2020b) and most likely to be well drained and more porous with a good supply of oxygen, and thus will most likely facilitate oxidation of organic matter (Bohn et al. 2001). This results in a high decomposition rate and low SOM accumulation.

Soil organic matter was very weakly correlated with Air_P. This may suggest the indirect involvement of Air_P in soil organic matter decay in the humid tropical region of southeast Nigeria. On the other hand, CAP_P was positively moderately correlated ($p < 0.01$) with SOM ($r = 0.41$) in the studied soils. Total_P and CAP_P are highly correlated ($r = 0.80$), with each also positively correlated with SOM and both having the highest values in soils over alluvium. Soils over alluvium, therefore, exert a similar influence on SOM, Total_P and CAP_P. This implies that SOM increases with an increase in soil wetness conditions.

3.2 Principal Component Analysis

Principal Component Analysis (PCA) (Tables 2 and 3; Fig. 2) revealed that PC1 explained 54% of the variability in the dataset, while PC2 explained 22% of the variance between soils of diverse geological environments. PC1 was presented by the contribution of SOM, BD, Kstat, Total P, and CAP_P, while PC2 was described by the contribution of Ksat, Air_P and CAP_P to their loadings (Table 3). The points outside the ellipses are outliers of each of the geological environments. All the soil properties were significantly influenced by SOM ($p < 0.01, 0.001$) under alluvial deposits except Air_P. Similarly, BD ($r = -0.783$), Ksat, Total_P and Cap_P ($r > 0.54$) were affected by SOM under LS. SOM was reportedly positively inter-related with Ksat and inversely with BD in soils formed over SS, while SOM had no influence on the properties for soils over SLM. The PCA result reconfirmed the correlation matrix output (Fig. 1).

Table 2. Principal component contributions

Importance of components	PC1	PC2
Standard deviation	1.802	1.155
Proportion of variance	0.541	0.223
Cumulative proportion	0.541	0.764

Table 3. Principal components correlation with variables

	PC1	PC2
SOM	**0.4091***	0.1613
BD	**-0.5119***	0.1312
Ksat	0.3486	**0.4455***
Total_P	**0.4872***	0.0134
Air_P	0.0485	**0.7559***
CAP_P	**0.4569***	**-0.4321***

NB: * Contribution to each PC

3.3 SOM Prediction

Presented in Table 4 is the result of the four stepwise linear regression models for SOM prediction. Leapforward yield (RMSE = 12.51%, R2 = 0.53, MAE = 8.68%), Leapbackward gave (RMSE = 11.50%, $R^2 = 0.58$, MAE = 8.48%), leapseq yielded (RMSE = 12.51%, $R^2 = 0.53$, MAE = 8.68%) and lmStepAIC function produced (RMSE = 13.24%, $R^2 = 0.54$, MAE = 9.56%). The results revealed that the best performing function for SOM prediction is the leapbackward function since it produced the lowest error with a high coefficient of determination value. However, all the model functions were within the acceptable prediction range ($R^2 \geq 0.50$–0.75) as proposed by Li et al. (2016). These results suggest that prediction of SOM may vary depending on the method/functions adopted. The backward elimination (leapbackward) likewise, the rest functions procedure identified the best model as having BD** and Ksat***, respectively.

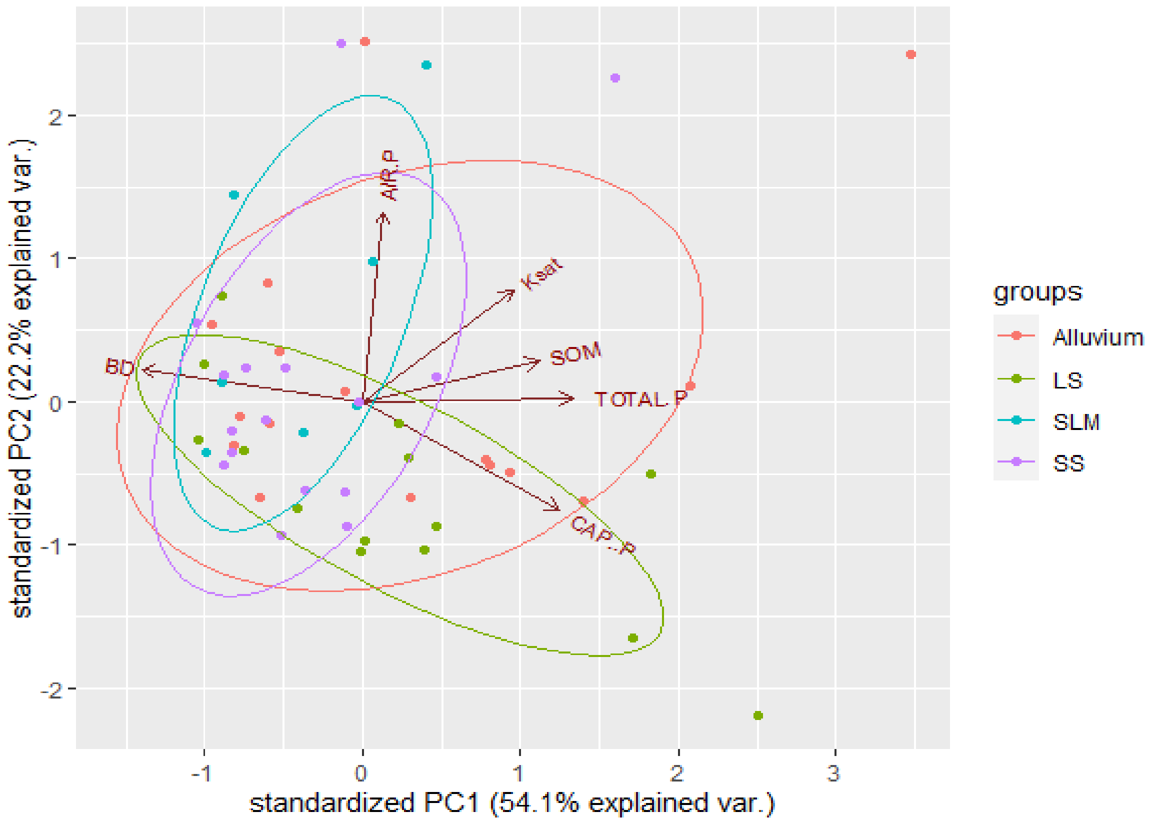

Figure 2. Principal component analysis of the variables grouped by lithologies

N/B: Bulk density (BD); saturated hydraulic conductivity (Ksat); total (Total_P); air-filled (Air_P); capillary porosities (Cap_P).

According to Sakin (2012), BD is closely related to SOC by storing large amounts of SOM. Compacted soil may contain more SOM, as it will occupy less space and more SOM per volume of soil, and the SOM in compacted soil is essentially "locked away". In contrast, soils that are not compacted have more contact with the air in the soil pores and so can be mineralized more efficiently and used as plant nutrients or leached. This relationship has been reported to aid the estimation of BD from SOM and vice versa (Perie and Ouimet, 2008). The study by Adams (1973) revealed that SOM had a dominant effect on both bulk and actual densities of soil in podzolic soil's organic and eluvial horizons.

Similarly, as shown in the correlation matrix, Ksat gave a higher correlation with SOM than BD; this was captured in all the four linear regression functions used in this study. The regression result confirms that an increase in SOM in the soil will result in a proportional increase in Ksat. This is because Ksat describes the capability of the bulk soil to transmit water when subjected to a hydraulic gradient. This is expressed by the volume of water flowing per unit area of bulk soil per unit time (Kosugi et al. 2002). Also, the result in this study is similar to the report of Ankenbauer and Loheide (2017). They reported an R2 = 0.625 in predicting SOM via volumetric water content at saturation in the meadow of the Sierra Nevada.

Generally, organic matter has been reported to significantly influence soil water retention and BD (Rawls et al. 2003; Olness and Archer, 2005; Saxton and Rawls, 2006). In contrast, other studies have reported that SOM is not necessary to estimate soil water retention properties accurately (Zhuang et al. 2001). However, in a dissimilar geological environment like this study, where the soils are predominantly similar in texture (Ofem et al. 2020a) and SOM content from 12.06 to 31.6 g/kg, SOM can easily be estimated via BD and Kstat. This is because SOM exerts a substantial control on surface water retention and BD variability.

Table 4 Prediction of soil organic matter (SOM) via various stepwise linear functions

LM Functions	RMSE	R^2	MAE %	Equations	Variable of importance
LeapForward	12.51	0.53	8.68	SOM=38.2–23.6xBD + 0.13xKsat	BD**, Ksat***
LeapBackward	**11.50**	**0.58**	**8.48**	**SOM=38.2–23.6xBD + 0.13xKsat**	**BD**, Ksat*****
LeapSeq	12.51	0.53	8.68	SOM=38.2–23.6xBD + 0.13xKsat	BD**, Ksat***
LmStepAIC	13.24	0.54	9.56	SOM=38.2–23.6xBD + 0.13xKsat	BD**, Ksat***

p = 0.001 '***'; 0.01 '**', Bold gave a good model fit

4 Conclusions

Soil organic matter is most likely to increase when favorable conditions for increased Ksat and porosity except Air_P, which did not influence SOM. Irrespective of geological material, BD decreases when SOM increases. The Ksat of soils over limestone (LS) and alluvium and BD of soils over LS had the strongest relationships with SOM with r > 0.70. However, air-filled (Air_P) porosity had no significant association with SOM and is most likely to have little effect on its decomposition in sedimentary geological environments. Farmers must put in place measures to regulate soil moisture (mulching and drainage), particularly in the sedimentary geological environment, which affects SOM. PC1 and PC2 contributed 74.38% of the total variance in the dataset of soils over diverse geological environments. The grouping pattern in the PCA explained that alluvial deposits influence most soil characteristics in this present study.

All the selected stepwise linear regression functions in the R environment performed the same as they fell within acceptable prediction criteria (R2 = 0.50–0.75). However, the best performing model function was leapbackward, which produced a smaller error when compared with others. The models selected BD and Ksat as the most important variables to explain the SOM variability in diverse sedimentary geology. The reason behind this result could not be presented at the time of this study; however, it could be interesting to access these functions with more variables and large sample densities. Therefore, we propose an increase in sample density per lithological make-up and the incorporation of soil properties known in works of literature to be influenced by SOM. This is to verify the performance of the leapbackward function over other functions, including the conventional lmStepAIC algorithm.

Data availability Not applicable.

Declarations

Conflict of interest The authors declare no conflict of interest regarding this work. Ethical approval Not applicable. Consent to participate All authors gave their consent. Concent for publication All authors gave their approval.

References

Adams WA (1973) The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. J Soil Sci 24(1):10–17

Ali S, Begum F, Hayat R, Bohannam BJM (2017) Variation in soil organic carbon stock in different land uses and altitudes in Bagrot Valley, Northern Karakoram. Acta Agric Scand B Soil Plant Sci 67:551–561

Alves MC, Suzuki LE (2004) Influência de diferentes sistemas de manejo do solo na recuperação de suas propriedades físicas. Acta Sci Agron 26:27–34

Ankenbauer KJ, Loheide SP (2017) The effects of soil organic matter on soil water retention and plant water use in a meadow of the Sierra Nevada, CA. Hydrol Process 31(4):891–901

Beare MH, Hendrix PF, Coleman DC (1994) Water-stable aggregates and organic matter fractions in conventional and no-tillage soils. Soil Sci Soc Am J 58:777–786

Belkhiri L, Narany TS (2015) Using multivariate statistical analysis, geostatistical techniques, and structural equation modeling to identify spatial variability of groundwater quality. Water Resour Manag 29:2073–2089

Blanco-Canqui H, Shapiro CA, Wortmann CS, Drijber RA, Mamo M, Shaver TM, Ferguson RB (2013) Soil organic carbon: The value to soil properties. J Soil Water Conserv 68(5):129A-134A. https:// doi. org/ 10. 2489/ jswc. 68.5. 129A

Bohn H, McNeal B, O'Connor G (2001) Soil Chemistry, 3rd edn. Wiley Interscience, New York

Boni, N. R., Espindola, C. R, Guimarães, E. C. (1994). Uso de leguminosas na recuperação de um solo decapitado. In: Simpósio nacional sobre recuperação de áreas Degradadas, Curitiba. pp. 563–568. (Fundação de Pesquisas Florestais do Paraná)

Bonini, C. S. B. and Alves, M. C. (2010). Relation between soil organic matter and physical properties of a degraded Oxisol in recovery with green manure, lime and pasture. 19th World Congress of Soil Science, Soil Solutions for a Changing World 1—6 August 2010, Brisbane, Australia.

Cardinael R, Chevallier T, Cambou A, Beral C, Barthès BG, Dupraz C, Durand C, Kouakoua E, Chenu C (2017) Increased soil organic carbon stocks under agroforestry: A survey of six different sites in France. Agric Ecosyst Environ 2017(236):243–255. https:// doi. org/ 10.

1016/j. agee. 2016. 12. 011hal- 01495 108
Eghball B (2002) Soil properties as influenced by phosphorus and nitrogen based manure and compost applications. Agron J 94:128–135
Ekwueme BN (1987) Structural orientations and Precambrian deformational episodes of Uwet Area, Oban Massif, SE Nigeria. Precambr Res 34:269–289
Fatoye FB, Gideon YB (2013) Geology and occurrences of limestone and marble in Nigeria. J Nat Sci Res 3(11):60–65
Ghimire P, Kafle G, Bhatta B (2018) Carbon stocks in Shorea robusta and Pinus roxburghii forests in Makawanpur district of Nepal. J AFU 2:241–248
Hao M, Zhang J, Meng M, Chen HYH, Guo X, Liu S, Ye L (2019) Impacts of changes in vegetation on saturated hydraulic conductivity of soil in subtropical forests. Sci Rep 9:8372. https:// doi. org/ 10. 1038/ s41598- 019- 44921-w
Hastie T, Tibshirani R, Friedman J (2008) The elements of statistical learning: data mining, inference, and prediction. Springer Science and Business Media, Berlin
Heuscher SA, Brandt CC, Jardine PM (2005) Using soil physical and chemical properties to estimate bulk density. Soil Sci Soc Am J 69:51–56. https:// doi. org/ 10. 2136/ sssaj 2005. 0051a
James G, Witten D, Hastie T, Tibshirani R (2014) An introduction to statistical learning: with applications in R. Springer Publishing Company Incorporated, Berlin
Jiao Y, Whalen JK, Hendershot WH (2006) No-tillage and manure applications increase aggregation and improve nutrient retention in a sandy-loam soil. Geoderma 134:24–33
Jiao S, Li J, Li Y, Xu Z, Kong B, Li Y, Shen Y (2020) Variation of soil organic carbon and physical properties in relation to land uses in the Yellow River Delta. China Scientific Reports 10:20317. https:// doi. org/ 10. 1038/ s41598- 020- 77303-8
John K, Abraham Isong I, Michael Kebonye N, Okon Ayito E, Chapman Agyeman P, Marcus Afu S (2020) Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. Land 9(12):487
Kafle G (2019) Vertical distribution of soil organic carbon and nitrogen in a tropical community forest of Nepal. Int J Forest Res. https:// doi. org/ 10. 1155/ 2019/ 30875 70
Kebonye NM, John K, Chakraborty S, Agyeman PC, Ahado SK, Eze PN, Borůvka L (2020) Comparison of multivariate methods for arsenic estimation and mapping in floodplain soil via portable X-ray fluorescence spectroscopy. Geoderma 384:114792
Kogel-Knabner I (2002) The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biol Biochem 34(2002):139–162
Komarov A, Chertov O, Bykhovetsa S, Shaw C, Nadporozhskaya M, Frolova P, Shashkova M, Shanina V, Grabarnika P, Priputinaa I (2017) Romul-Hum model of soil organic matter formation coupled with soilbiota activity. I. Problem formulation, model description, and testing. Ecol Model 345:113–124
Kosugi KI, Hopmans JW, Dane JH (2002) Parametric models. Methods Soil Analysis 5:739–757
Lei Z, Yu D, Zhou F, Zhang Y, Yu D, Zhou Y, Han Y (2019) changes in soil organic carbon and its influencing factors in the growth of Pinus sylvestris var. mongolica plantation in Horqin Sandy Land, northeast china. Sci Rep 9:16453. https:// doi. org/ 10. 1038/ s41598- 019- 52945-5
Li XG, Li FM, Zed R, Zhan ZY, Bhupinderpal S (2007) Soil physical properties and their relations to organic carbon pools as affected by land use in an alpine pasture land. Geoderma 139:98–105
Li L, Lu J, Wang S, Ma Y, Wei Q, Li X, Cong R, Ren T (2016) Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in situ leaf spectroscopy. Ind Crops Prod 91:194–204

Li Z, Liu C, Dong Y, Chang X, Nie X, Liu L, Zeng G (2017) Response of soil organic carbon and nitrogen stocks to soil erosion and land use types in the Loess hilly–gully region of China. Soil Tillage Res 166:1–9

Liu S, An N, Yang J, Dong S, Wang C, Yin Y (2015) Prediction of soil organic matter variability associated with different land use types in mountainous landscape in southwestern Yunnan province, China. CATENA 133:137–144. https:// doi. org/ 10. 1016/j. catena. 2015. 05. 0100

Masri Z, Ryan J (2006) Soil organic matter and related physical properties in a mediterranean wheat-based rotation trial. Soil Tillage Res 87:146–154. https:// doi. org/ 10. 1016/j. still. 2005. 03. 003

Muñoz-Rojas M, Abd-Elmabod SK, Zavala LM, DelaRosa D, Jordán A (2017) Climate change impacts on soil organic carbon stocks of Mediterranean agricultural areas: a case study in Northern Egypt. Agr Ecosyst Environ 238:142–152. https:// doi. org/ 10. 1016/j. agee. 2016. 09. 001

Nemes A, Rawls WJ, Pachepsky YA (2005) Influence of organic matter on the estimation of saturated hydraulic conductivity. Soil Sci Soc Am J 69:1330–1337. https:// doi. org/ 10. 2136/ sssaj 2004. 0055

Obi ME (2000) Laboratory manual for soil physics. Department of Soil Science, University of Nigeria, Nsukka, p 34

Ofem KI, Abua SO, Umeugokwe CP, Ezeaku VI, Akpan-Idiok AU (2020a) Characterization and suitability evaluation of soils over sandstone for cashew (Anacardium occidentale L.) production in a Nigerian Southern Guinea Savanna. IJSRP 10(7):353–368

Ofem KI, Asadu CLA, Ezeaku PI, Kingsley J, Eyong MO, Katerina V, Václav T, Karel N, Ondrej D, Vít P (2020b) Genesis and classification of soils over limestone formations in a Tropical Humid Region. Asian J Sci Res 13:228–243

Oliveira FCC, Ferreira GWD, Souza JLS, Vieira MEO, Pedrotti A (2018) Soil physical properties and soil organic carbon content in northeast Brazil: long-term tillage systems effects. Soils Plant Nutr. https:// doi. org/ 10. 1590/ 1678- 992X- 2018- 0166

Olness A, Archer D (2005) Effect of organic carbon on available water in soil. Soil Sci 170(2):90–101

Parihar CM, Yadav MR, Jat SL, Singh AK, Kumar B, Pradhan S, Chakraborty D, Jat ML, Jat RK, Saharawat YS, Yadav OP (2016) Long term effect of conservation agriculture in maize rotations on total organic carbon, physical and biological properties of a sandy loam soil in north-western Indo-Gangetic Plains. Soil Tillage Res 161:116–128. https:// doi. org/ 10. 1016/j. still. 2016. 04. 001

Perie C, Ouimet R (2008) Organic carbon, organic matter and bulk density relationships in boreal forest soils. Can J Soil Sci 88(3):315–325

R Core Team. (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria [online]. https:// www.r- proje ct. org/. Accessed 13 May 2020.

Rawls WJ, Pachepsky YA, Ritchie JC, Sobecki TM, Bloodworth H (2003) Effect of soil organic carbon on soil water retention. Geoderma 116(1–2):61–76

Rawls WJ, Nemes A, Pachepsky YA (2005) Effect of soil organic matter on soil hydraulic properties. In: Pachepsky YaA, Rawls WJ (eds) Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam, pp 95–114

Saint-Laurent D, Gervais-Beaulac V, Paradis R, Arsenault-Boucher L, Demers S (2017) Distribution of soil organic carbon in riparian forest soils affected by frequent floods (Southern Québec, Canada). Forests 8:124. https:// doi. org/ 10. 3390/ f8040 124

Sakin E (2012) Organic carbon organic matter and bulk density relationships in arid-semi arid soils in Southeast Anatolia region. Afr J Biotech 11(6):1373–1377

Sambo, E. E., Ufoegbune, G. C., Eruola, A. O. & Ojekunle, O. Z. (2016). Impact of Rainfall Variability on Flooding of Rivers in Cross River Basin, Nigeria. Nigerian Metreological Society (NMETS), "Climate Variability And Change: Impact, Science, Innovation And Policy" at Federal College of Education Osiele, Abeokuta. 21—24 November, 2016

Sanchez PA, Palm CA, Szott LT, Cuevas E, Lal R (1989) Organic input management in tropical agrosystems. In: Coleman DC, Oades JM, Uehara G (eds) Dynamics of soil organic matter in tropical ecosystems. Honolulu, Capı' tulo 2, pp 125–152 (ISBN 0-8248-1251-4)

Saxton KE, Rawls WJ (2006) Soil water characteristics estimates by texture and organic matter for hydrologic solutions. Soil Sci Soc Am J 70:1569–1578. https:// doi. org/ 10. 2136/ sssaj 2005. 0117

Silva SHG, Teixeira AFS, Menezes MD, Guilherme LRG, Moreira FMS, Curi N (2017) Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). Ciênc Agrotec 41(6):648–664. https:// doi. org/ 10. 1590/ 1413- 70542 01741 60103 17

Soil Survey Staff (SSS) (2014). Kellogg soil survey laboratory methods manual. Soil Survey Investigations Report No. 42, Version 5.0. R. Burt and Soil Survey Staff (ed.). US Department of Agriculture, Natural Resources Conservation Service. P. 1001.

Souza DF, Barbosa RS, da Bezerra Silva YJ, de Soares Moura MC, de Porto Oliveira R, Martins V (2019) Genesis of sandstone- derived soils in the Cerrado of the Piauí State Brazil. Rev Ambient Águae. https:// doi. org/ 10. 4136/ 1980- 993X (ISSN 1980-993X)

Takata Y, Funakawa S, Akshalov K, Ishida N, Kosaki T (2007) Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. Soil Sci Plant Nutr 53(3):289–299. https:// doi. org/ 10. 1111/j. 1747- 0765. 2007. 00142.x

Tisdall JM, Oades JM (1982) Organic matter and water-stable aggregates in soils. J Soil Sci 33:141–163

Whalen JK, Chang C (2002) Macroaggregate characteristics in cultivated soils after 25 annual manure applications. Soil Sci Soc Am J 66:1637–1647

Yao R, Yang J, Wu D, Li F, Gao P, Wang X (2015) Evaluation of pedotransfer functions for estimating saturated hydraulic conductivity in coastal salt-affected mud farmland. J Soils Sediments 15:902–916. https:// doi. org/ 10. 1007/ s11368- 014- 1055-5

Zhao W, Hu Z, Yang H, Zhang L, Guo Q, Wu Z, Liu D, Li S (2016) Carbon density characteristics of sparse Ulmus pumila forest and Populus simonii plantation in Onqin Daga Sandy Land and their relationships with stand age. Chin J Plant Ecol 40:318–326

Zhou W, Han G, Liu M, Zeng J, Liang B, Liu J, Qu R (2020) Determining the distribution and interaction of soil organic carbon, nitrogen, pH and texture in soil profiles: a case study in the lancangjiang River Basin, Southwest China. Forests 11:532. https:// doi. org/ 10. 3390/ f1105 0532

Zhuang J, Jin Y, Miyazaki T (2001) Estimating water retention characteristic from soil particle-size distribution using a non-similar media concept. Soil Sci 166(5):308–321