Journal of Geomatics Vol. 13, No. 1, April 2019

Predicting soil organic carbon concentration using digital soil mapping techniques in eastern Mau forest - Nakuru County, Kenya

Ian Ocholla Ariko[1] and Patroba Achola Odera[2*]
[1]Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
[2]University of Cape Town, Private Bag X3, Rondebosch, South Africa
*Email: patroba.odera@uct.ac.za

(Received: Mar 13, 2018; in final form: May 01, 2019)

Abstract: Soil organic carbon (SOC) concentration is one of the most important indicators of soil fertility and soil quality. Accurate information about the spatial variation of SOC concentration is critical to sustainable soil utilization and management. Eastern Mau forest reserve is one of the major water towers in Eastern Africa, providing essential ecosystem services and it is critical in storage of organic carbon. The objective of this study is to predict the spatial distribution of SOC concentration using digital soil mapping (DSM) techniques. Two prediction techniques; multi-linear regression and multi-linear regression kriging have been used. Seventy-five percent (75%) of soil sample data have been used for model(s) calibration while 25% for model(s) validation. Various variables including terrain attributes, climate data, land use and soil properties data are used in the prediction. Results indicate that multi-linear regression model has a lower R^2 of 37.4% compared to multi-linear regression kriging, with R^2 of 42.3%. The results further show that the western parts of eastern Mau (largely forest land) has the highest concentration of SOC, while lowest SOC concentration is observed on the eastern section (largely crop land). The study demonstrates that multi-linear regression kriging performs better than multi-linear regression in capturing the spatial distribution of SOC across the study area.

Keywords: soil organic carbon, digital soil mapping, multi-linear regression, multi-linear regression kriging, Eastern Mau forest

1. Introduction

Forest reserves are most affected by human population for settlement and agricultural activities. The SOC pool, estimated at 1,550 Pg to 1-meter depth is about twice the atmospheric pool or 2.8 times the biotic pool (Batjes, 1996). Soil organic matter contains 58% carbon on average (Chan, 2008) and is essential in regulating climate, water supplies and biodiversity which are vital to human well-being (Kumar et al., 2016).

Countries which are signatories to the United Nations Framework Convention on Climate Change (UNFCC) under the Kyoto protocol are expected to monitor changes in the soil organic carbon (SOC storage) within their countries (Razakamanarivo et al., 2011). Consequently, better and accurate techniques for estimation of SOC concentration are necessary to inform policy making on actions that enhance organic carbon storage and suggest measures to counter areas with depreciating SOC storage. The recent development of geospatial technologies has allowed for a spatial quantitative prediction approach involving modeling of continuous soil properties (based on factors of soil formations) besides the assessment of accuracy and uncertainties of the predictions. This approach referred as digital soil mapping (Mora-Vallejo et al., 2008) is much better compared to the conventional mapping techniques which only generate qualitative maps whose accuracy cannot be assessed. The conventional method(s) use polygons which do not consider spatial variability within an area because a whole polygon is normally given a constant value.

Soil organic matter (SOM) makes up just 2-10% of soils mass but has a critical role in physical, chemical and biological function of agricultural soils. SOM is formed by decay of organic material that enters the soil system. Soil organic carbon (SOC) is the major constituent of SOM. SOC is normally expressed as a percentage carbon by weight, that is, g C per 100 g of soil (Chan, 2008).

In Kenya, deforestation has led to a decline in most of the forest reserves, leaving only five major water towers, which are closely monitored for ecosystem sustainability. This paper identifies possible significant covariates to be used in the prediction model(s) and generates predicted SOC concentration map using multi-linear regression and multi-linear regression kriging techniques in eastern Mau forest.

The study area (eastern Mau forest reserve) is in Nakuru County, Kenya. It lies between latitudes 0^0 15' S and 0^0 40' S and longitudes 35^0 40'E and 36^0 10'E as shown in figure 1. It has an area of approximately 650 km^2, with an elevation ranging from 2,210 to 3,070 m. The climate is cool and humid with an average of 93.5 mm of annual precipitation and a mean annual temperature ranging from 9.8 to 17.5^o C. The Njoro and Naishi rivers drain from the eastern slopes into Lake Nakuru, while River Nessuiet flows northwards into Lake Bogoria. The major land uses are forest, agriculture and grassland. Despite the rampant deforestation and degradation experienced since mid-1990s, because of illegal logging, charcoal burning and encroachment of approximately 61,023 ha for human settlement (UNEP, 2009), eastern Mau forest remains the largest Afromontane forest in Eastern Africa.

2. Theoretical background of digital soil mapping

Digital soil mapping (DSM) technique is defined by (Lagacherie and McBratney, 2007) as "the creation and population of spatial information systems by numerical models inferring the spatial and temporal variation of soil types and soils properties from soil observation and knowledge from related environmental variables". DSM is like conventional soil mapping except that the functional relationship between the soils attributes and model factors are formulated using statistical models rather than conceptual models (Thompson et al., 2012). These statistical models are fitted using geo-referenced soil data. We discuss some models in the following sections.

2.1 State factors or CLORPT model

For the latter half of 20th century, scientific rationale for soil mapping has been the state factors or CLORPT model, which stands for (CL=climate, O=organisms, R=relief, P=parent material, and T=time). The logic of CLORPT model was based on the equation of Jenny (Jenny, 1992) and formulated from the recognition of the factors of soil formation. The state factor equation can be expressed as,

$$S = f(cl, o, r, p, t) \quad (1)$$

where, S represents the soil, considered to be a function of (cl) climate, organism (o) or vegetation, relief (r), parent material (p) acting through time (t).

The Jenny equation illustrates that by correlating soil attributes with observable difference in one or more of the state factors, a function (f) or model can be developed that explains the relationship between the two, which can be used to predict soil attributes at new locations when the state factors are known. However, the state factors do not constitute factors that institute pedogenic processes.

2.2 SCORPAN model

In the last decade, McBratney et al., 2003 generalized and formulated a new equation with the objective of modeling the variables responsible for the processes of soil formation, through an empiric quantitative description of the relationships among other spatially geo-referenced factors which are used as spatial prediction functions. It is an improvement of Jenny equation, the scorpan model is expressed as,

$$S = f(s, c, o, r, p, a, n) \quad (2)$$

where, (s) is the soil attributes, (a) represent the age or time factor and (n) the space or spatial position. The other symbols have their usual meanings as given in equation 1.

This model differs from the clorpt as it is intended for quantitative spatial prediction rather than explanation, this distinction justifies the inclusion of soil and space as factors because soil attributes can be predicted from other soil attributes and spatial information (Thompson et al., 2012).

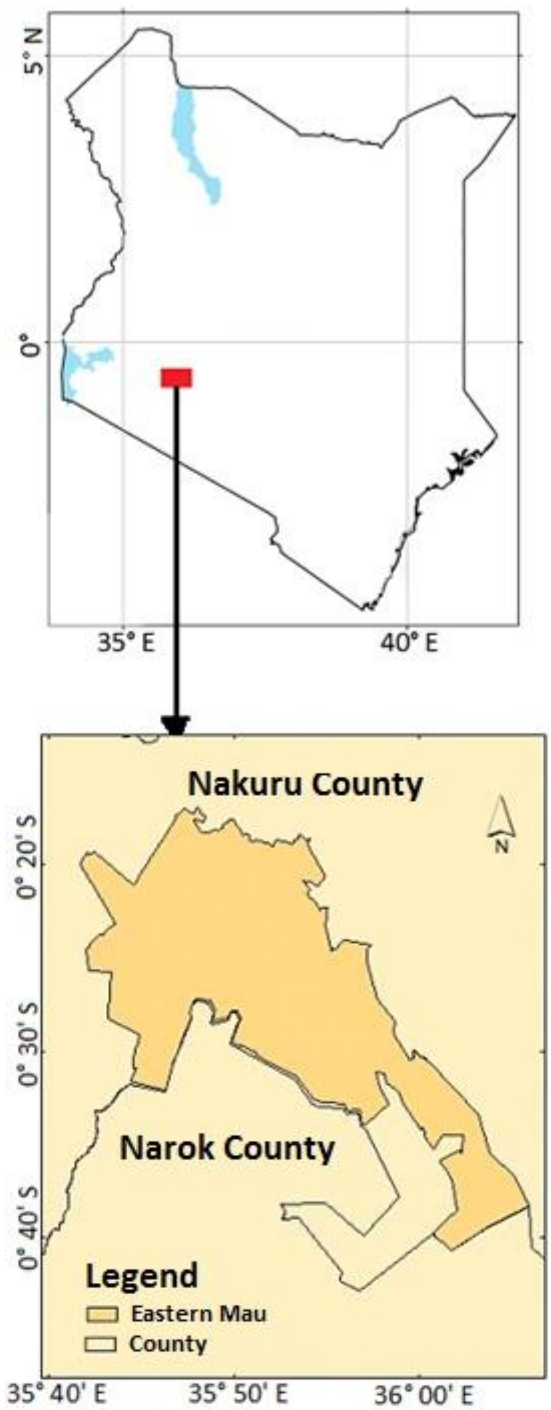

Figure 1: Eastern Mau forest reserve.

2.3 Geostatistics

Geostatistics offers a way of describing the spatial continuity of natural phenomena and provides adaptations of classical regression techniques (Hengl, 2009). Geostatistics is used to predict values of a sample variable over the whole area of interest as it is used in combination with various geospatial Information Systems (GIS) layers. Geostatistics differs from the conventional statistics in that in the later the samples taken from a statistical community are independent from each other and the presence of one sample does not show any information about the next sample. While in geostatistics the spatial structure or correlation among variables in a region are investigated (Abadi et al., 2015), with the use of semi-variograms to quantify spatial autocorrelation.

We discuss only one type of interpolation technique here, Kriging. Kriging is one of the geostatistical methods, based on the theory or regionalized variables and variogram model. It is considered as the best linear unbiased predictor (*BLUP*) that satisfies a certain optimality criterion. It is named after a South African mine engineer D.G. Krige who used the technique in the mining industry in the early 1950's as a means of improving ore reserve estimation (Krige, 1951). Kriging is a suitable method in the presence of spatial dependence as it is beneficial to model the deterministic component of soil spatial variation as a function of the environmental covariates and the stochastic component (Thompson et al., 2012).

We now describe ordinary kriging and regression kriging. Ordinary kriging is often regarded more appropriate interpolation technique, as it adapts to local fluctuation of the mean by limiting the domain of stationarity of the mean to the local neighborhood (Mulder et al., 2011). Ordinary kriging is used to improve the prediction by interpolating the environmental variables. On the other hand, Regression kriging is a hybrid method that combines either a simple or multi-linear regression (MLR) model with ordinary kriging of the regression residuals.

Multi linear regression is commonly used in up-scaling approach to model the linear relationship between independent variable and secondary variables (predictors). However, MLR generates a process which is stationary and assumes the residuals are identical and independently distributed. Errors associated with MLR are large since the approach does not consider the varying relationship between the environmental variable and the SOC across space (Ge et al., 2007).

Regression kriging (RK) involves the use of environmental variables as it consists of three components, it can be expressed as (Zhang et al., 2012),

$$Z_{(S)} = Z^{*}_{(S)} + E^{'}_{(S)} + E^{''} \qquad (3)$$

where, $Z_{(S)}$ is the RK prediction formed by summing the regression prediction from the covariates, $Z^{*}_{(S)}$ is the deterministic component (ordinary kriging of the residuals), $E^{'}_{(S)}$ is the stochastic component and $E^{''}$ is the pure noise.

In this study, multi linear regression and multi-linear regression kriging techniques have been used to predict SOC concentration based on other similar studies (Mora-Vallejo et al., 2008; Sumfleth and Duttmann, 2008). Peng et al., 2013 indicate that the use of multi linear regression model is simple and direct.

3. Materials and methods

3.1 Soil data

Two hundred and twenty (220) soil samples of soil organic carbon, clay, sand and silt (up to 30 cm depth) were provided by the Kenya soil survey. The soil samples were georeferenced within the study area, the soil samples were then randomly divided into two, that is, 75% calibration set, and 25% validation set. Figure 2 shows the distribution of SOC data points in the study area.

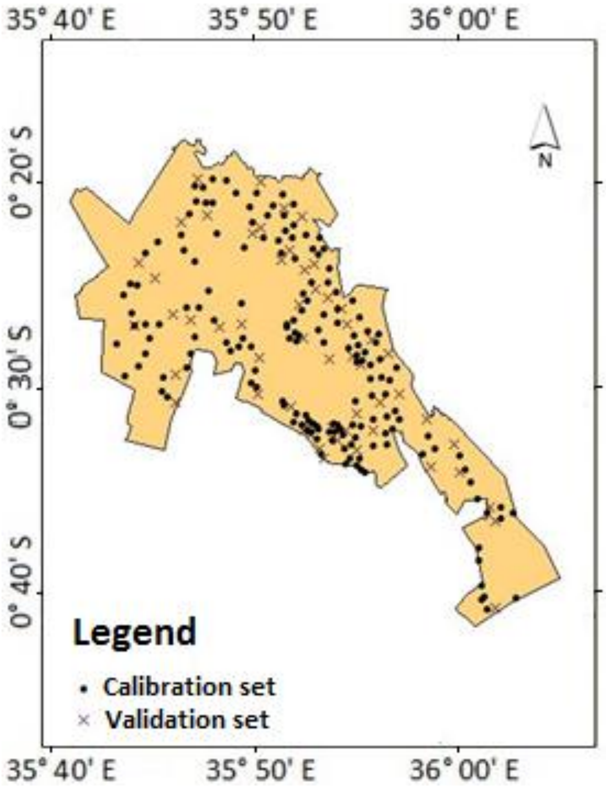

Figure 2: Distribution of SOC data points.

3.2 Auxiliary data

Various sources of data retrieved and analyzed to capture the spatial variation of the soil forming factors in the study area included climate, normalized difference vegetation index (NDVI) and digital elevation model (DEM). The mean annual temperature and precipitation on a 30 arc-second raster data were obtained from www.worldclim.org. The datasets were resampled to 30 m resolution (admittedly with some errors). Multi-spectral remote sensed satellite imagery data (Landsat 5) was obtained from http://earthexplorer.usgs.gov/. Landsat 5 was selected because soil sample were collected in 2011. Subsurface reflectance band values were used in calculating NDVI values using the following expression,

$$NDVI = \frac{NIR - R}{NIR + R} \qquad (4)$$

where, R and NIR are the Red and Near Infrared bands respectively (Rouse et al., 1993). The land cover of the study area was also generated from the relevant bands.

A 30 m Shuttle Radar Topographic Mission (SRTM) data was obtained from www.jpl.nasa.gov, it was used to derive different topographic attributes as aligned with relief as part of soil forming factors. The parameterized derivatives included elevation, slope, aspect, plan curvature, profile curvature and flow accumulation. Topographical Wetness Index (TWI) which is a secondary topographic derivative was also used. The equation used for TWI is given as (Wilson and Gallant, 2000),

$$TWI = \frac{In(flow\ accumulation)}{\tan(slope \div 57.29577951)} \quad (5)$$

3.3 Selection of explanatory variables for predictive model

The SOC calibration dataset was used to extract points from the co-variables through map overlay. The extracted co-variable-values were then linearly regressed against the response variable (SOC). Regression analysis is essential in characterizing the relationship between the predictors and response variable, also it is used in estimating their correlation. Stepwise regression analysis using Akaike Information Criterion (AIC) was applied on the initial model to obtain a reduced model, which accomplishes a desired level of prediction with as few predictor variables as possible (Akaike, 1973). The best model has the smallest AIC value. The coefficients of the reduced model were fitted into a multi-linear regression equation to map the spatial distribution of SOC concentration, using the following expression (Kumar and Lal, 2011),

$$Y = B_0 + B_1X_1 + B_2X_2 + ... + B_kX_k + e \quad (6)$$

where, Y is the SOC concentration, B_0 is the Y intercept (a constant term) B_k are model coefficients, X_k are independent variables, and e is an error of disturbance.

3.4 Spatial structure of the model

The residuals obtained from the reduced regressed model were assessed and fitted into a semi variogram. The nugget to sill ratio was then used to characterize the importance of the random component and provide quantitative measures of spatial dependence. The residuals were interpolated using kriging to incorporate the spatial correlation of the errors of the multi-linear regression model.

The final prediction map was obtained by spatial overlay of the multi-linear regression model surface with the kriged interpolated residual surface in a regression kriging approach (equation 3). Validation test using mean error (ME), root mean square error (RMSE) and coefficient of determination (R^2) were performed to both models to evaluate the prediction accuracy.

4. Results and discussion

4.1 Modelling results

Results of the multi linear regression analysis of various covariates indicated that the covariables could account for 56% of SOC concentration variability within the study area. The results from the stepwise regression analysis are shown in table 1.

Model number four (4) had the lowest AIC values (Table 1). The coefficients of the model parameters were determined and the reduced multi linear regression model obtained as,

$$SOC = -7.2832 + (-0.0895468 \times Silt) + (-0.771998 \times band4) + (68.5149 \times band1) + (9.2718 \times NDVI) + (0.0019903 \times flow\ accumulation) + (0.003212 \times elevation) \quad (7)$$

where SOC and NDVI have the same meanings as earlier defined.

Model parameters such as silt concentration, NDVI, flow accumulation and elevation are given in figures 3, 4, 5 and 6 respectively. We also present rainfall, sand soil concentration and major land use/cover in figures 7, 8 and 9 respectively.

Table 1: Model parameters with their AIC values.

Model	Covariables	AIC values
1	TWI, Slope, Curvature, plan curvature, rainfall, sand, NDVI, band 1,4,5, Elevation, Silt, flow accumulation	39.34
2	Plan curvature, rain, flow accumulation, NDVI, Silt, band 1,4,5, Elevation	36.50
3	Rain, Elevation, band 1,4,5, NDVI, Silt, Sand, flow accumulation	35.15
4	Rain, Sand, Band 1,4, flow accumulation, NDVI, Silt, Elevation	34.88

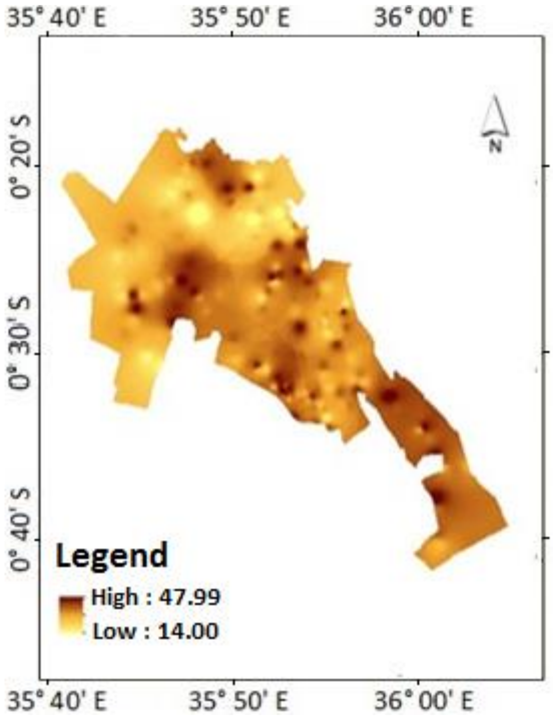

Figure 3: Silt concentration distribution (units in %).

Using equation 7, the reduced multi linear regressed SOC surface was generated as shown in figure 10. The residuals were then assessed if they fulfilled the requirements for spatial dependence i.e. they should not be biased and should have a uniform distribution. This assessment was done through fitting the residuals into a histogram and normal quarter quantile plot, the residuals certified the requirements. The residuals were then fitted into an experimental semi-variogram with a spherical function, to measure the spatial dependence using the nugget effect (N) to sill (S) ratio.

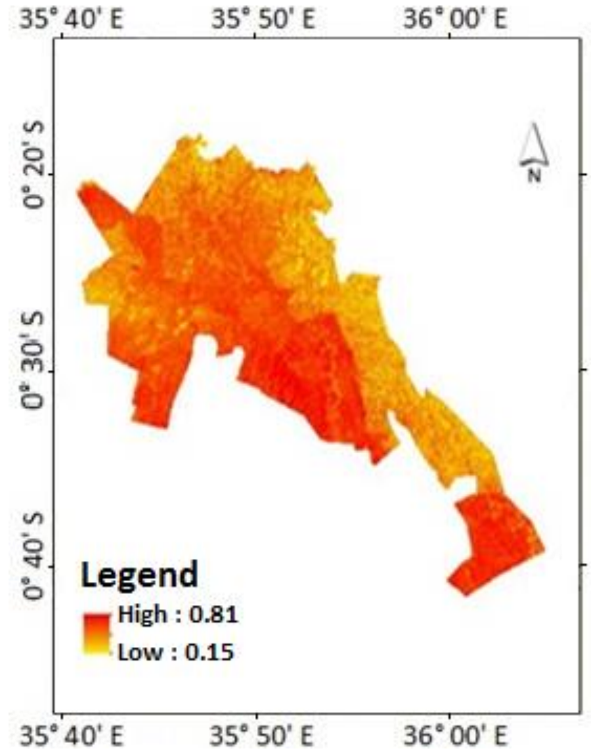

Figure 4: Normalized difference vegetation index (NDVI).

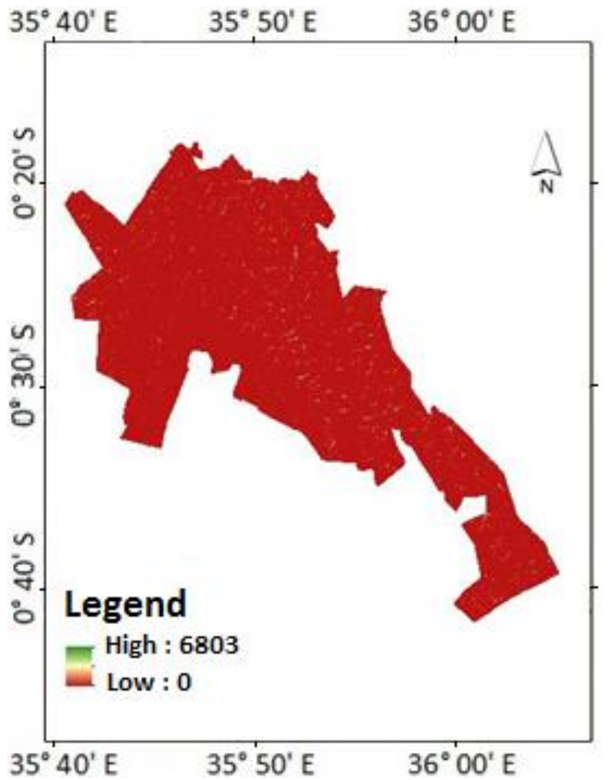

Figure 5: Flow accumulation.

The N:S ratio was 56.34% which indicated a moderate spatial dependence of the residuals. The high nugget value as shown in table 2 indicated the measurement error within the data. It also meant that the distance between sample points were far apart lowering the spatial dependence of the residuals.

Table 2: Semi-variogram parameters of the residuals.

	Nugget	Sill	Nugget/sill
SOC residuals	0.54	0.9657	0.56

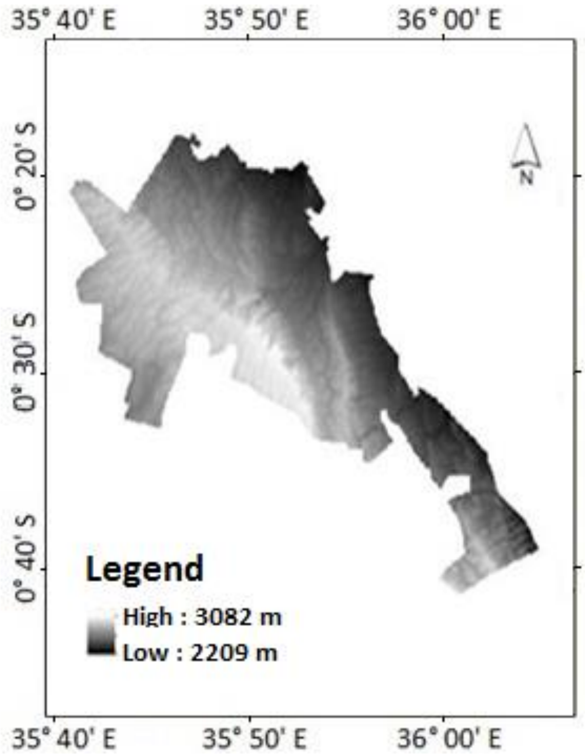

Figure 6: Elevation (units are in m).

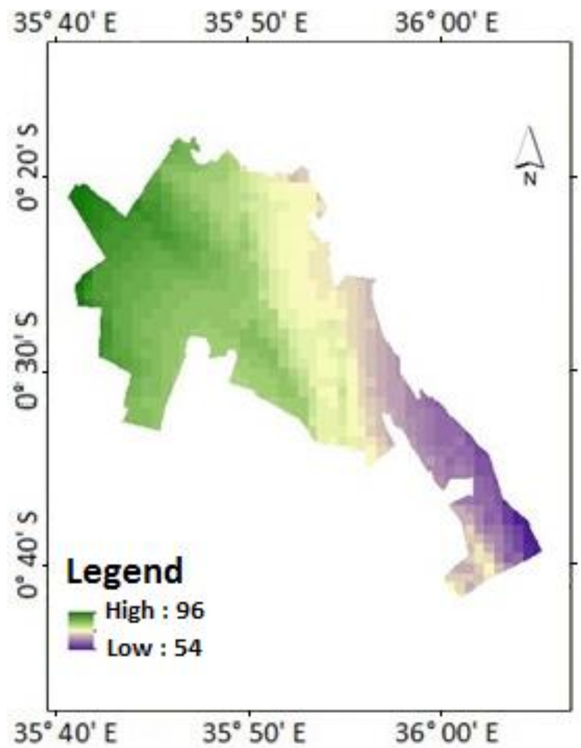

Figure 7: Mean annual rainfall (units are in mm).

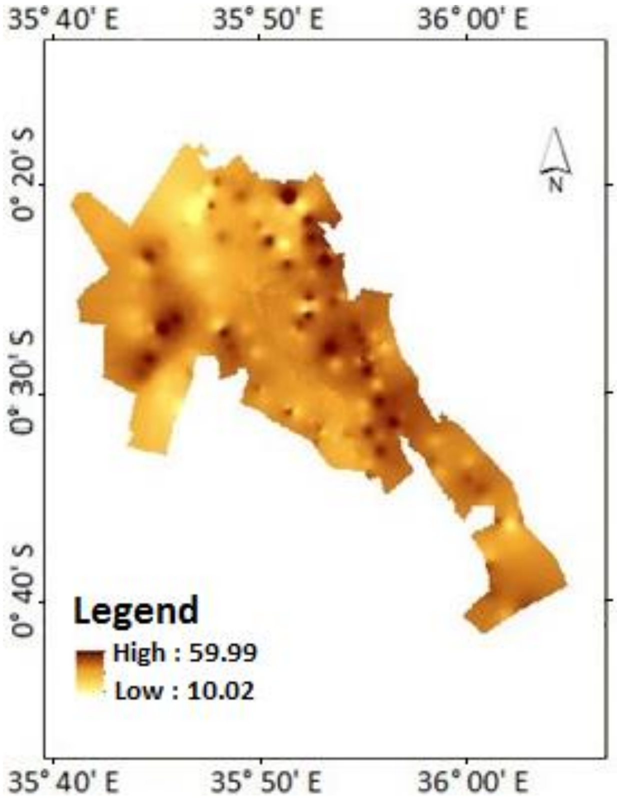

Figure 8: Sand soil concentration (units are in %).

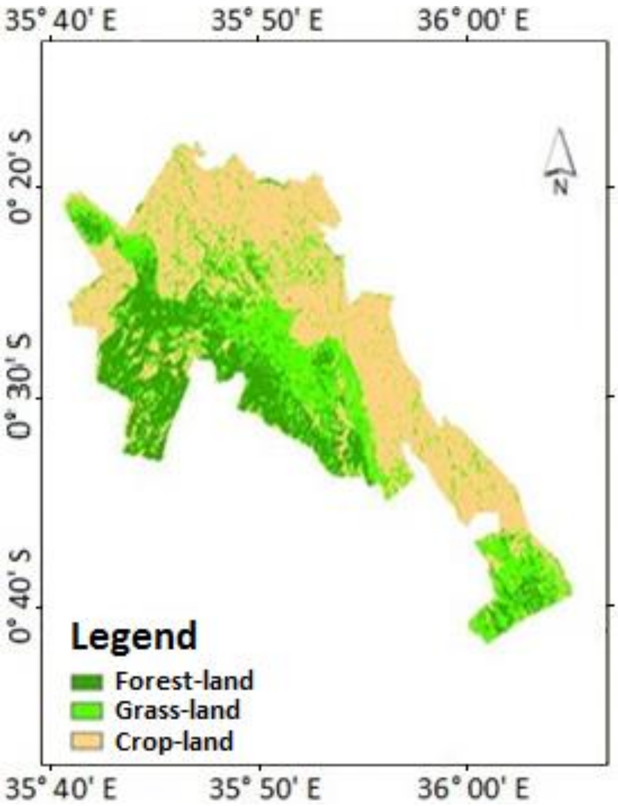

Figure 9: Major land use/cover.

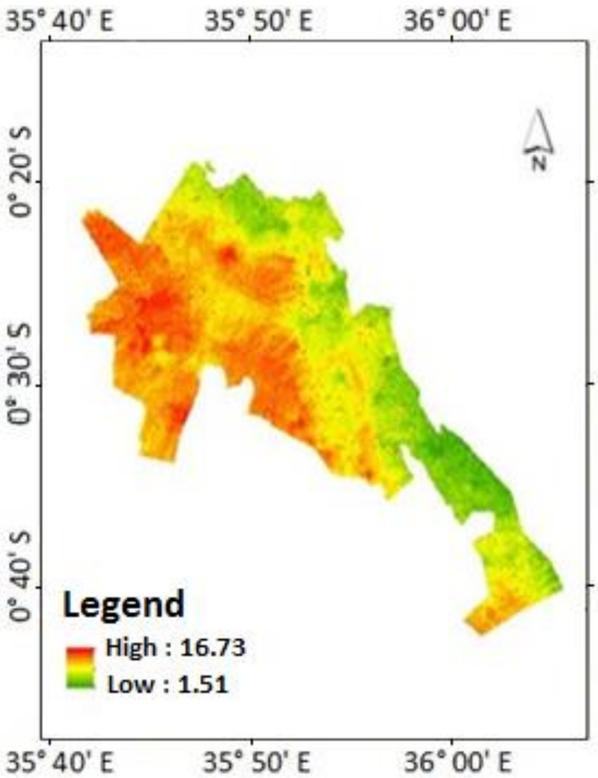

Figure 10: Reduced regressed surface of SOC concentration (units are in %).

The residuals were then interpolated using ordinary kriging as shown in figure 11. The reduced multi-linear regressed model surface and the kriged surface maps were spatially overlaid to obtain the final regression kriged surface of SOC concentration (Figure 12).

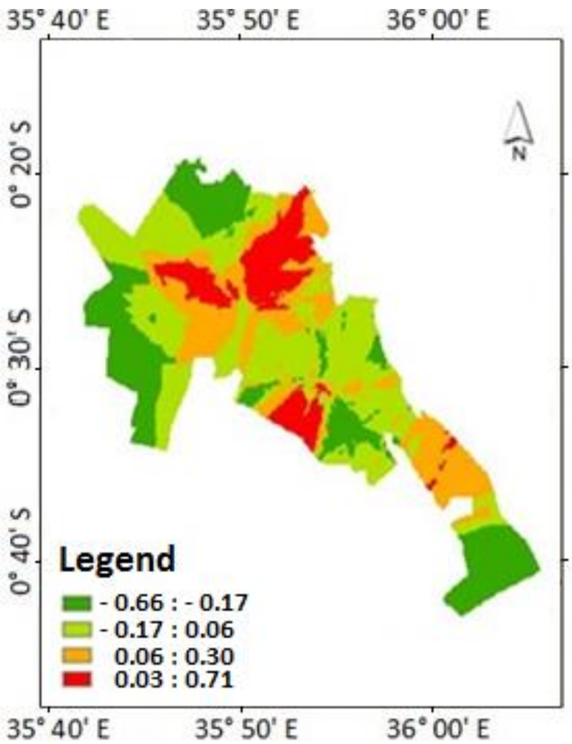

Figure 11: Interpolated residual surface.

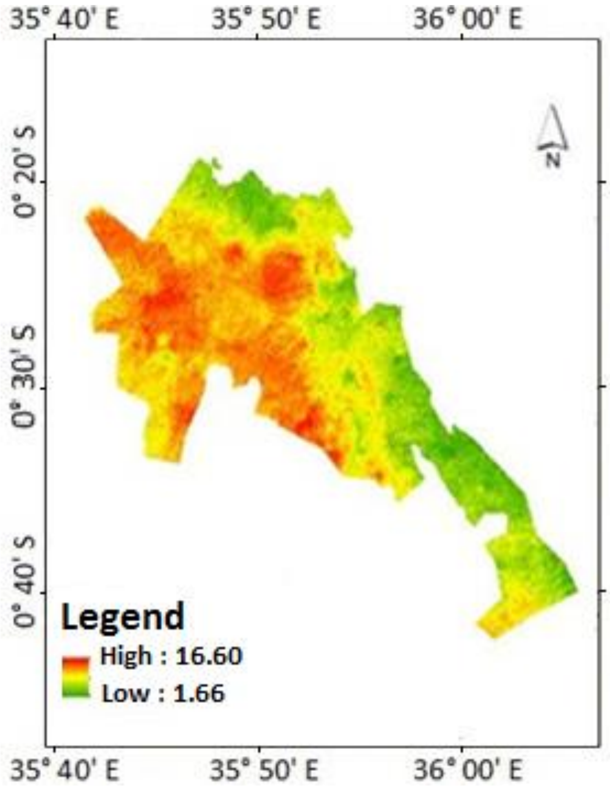

Figure 12: Regression kriged surface of SOC concentration (units are in %).

4.2 Model assessment

Fifty-five independent SOC data points were used for validation of the two models. The results of the validation are given in table 3. The R^2 of the regression and regression kriging models are 37.38 and 42.25%, respectively. The better performance by regression kriging (RK) is because it considers uncertainty due to regression and measurement in the form of kriged variables compared to the regression model.

The low R^2 (42.25%) of the overall RK model can be attributed to the moderate spatial structure of the residuals and errors due to the interpolation of datasets. In table 3, RK model had a lower ME (0.2794) compared to the regression model (0.3304). The RK model also had a better RMSE value (1.3970) compared to the regression surface RMSE value (1.4590). From the validation, the regression kriging model proved to be a better prediction technique in the study area.

Table 3: Statistics of models' assessment.

	Regression	Regression kriging
Minimum difference	0.0392	0.0690
Maximum difference	5.5494	4.9121
Mean difference	0.3304	0.2794
RMSE	1.4590	1.3970
R^2	0.3738	0.4225

4.3 Discussion

The spatial distribution of regression-kriging showed that SOC concentration decreased from the western section to the south-eastern section in the forest reserve, as shown in figure 12. The spatial distribution of SOC concentration in the study area could be predicted using terrain attributes, soil, texture, climate and land use.

Band 4 and band 1 (Landsat data) were significant variables in estimation of SOC concentration, band 4 captures the near infrared reflectance which distinguishes vegetation varieties and conditions while band 1 provides information capable of differentiating soil and rock surfaces from vegetation. From the two bands, areas which have vegetation cover tend to have higher concentration of SOC compared to bare soil and rock surfaces.

A comparison of the final SOC prediction map (Figure 12) with the land use/cover map (Figure 9), reveals that SOC concentration is relatively higher in the forest areas and lower in the cropland areas. These results show that land use/cover type has a significant impact on the spatial SOC concentration patterns. This can be attributed to the fact that forest land fix plentiful SOC because of the flourishing soil plant roots and thicker forest litter layers, which are easily absorbed and beneficial to SOC. In cropland, there is continuous tillage exposing the nutrient for decomposition leading to decrease in the SOC concentration.

High elevation regions also recorded high SOC concentration, this is because the high-altitude areas had low temperatures which lower nutrient decomposition and these areas have low human interference, compared to areas of lower elevation which have increased anthropogenic activities which expose soil organic carbon for decomposition.

5. Conclusions

This study compared two prediction techniques (multi-linear regression and multi-linear regression kriging) in predicting soil organic carbon concentration in the eastern Mau forest reserve. The two models were calibrated and verified using independent validation datasets. Results show that multi-linear regression model has a lower R^2 of 37.4% compared to multi-linear regression kriging (42.3%). The western part of eastern Mau, largely forest land, has the highest concentration of SOC, while lowest SOC concentration is observed on the eastern section, largely crop land. Given the characteristics of the study area, the number of observations used and data distribution, the prediction of SOC by multi-linear regression kriging is satisfactory. These results can be improved by accurate selection and representation of soil formation factors and related spatial residuals.

Acknowledgements

We would like to thank the Kenya Soil Survey Department for providing valuable datasets. We are grateful to Dr. Kennedy Were for his support in terms of data and technical expertise in carrying out this study. Finally, we appreciate the valuable comments by two anonymous reviewers, that have been used to improve the quality of the manuscript.

References

Abadi, K.M., R. Jafarinia and J. Varvani (2015). Analyzing kriging and cokriging methods by using ArcGIS software in preparing SP map of Farahan plain soil. Biological Forum-An International Journal, 7(1), 836–846.

Akaike, H. (1973). Information theory and an extension of themaximum likelihood principle. In Petrov, B.N. and F. Csaki (eds.), 2nd International Symposium on Information Theory, Akademiai Kiado, Budapest, pp. 267–281.

Batjes, N.H. (1996). Total carbon and nitrogen in the soils of the world. European Journal of Soil Science, 47, 151–163.

Chan, Y. (2008). Increasing soil organic carbon of agricultural land. In: Prime facts, NSW Department of Primary Industries, available online at: https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0003/210756/Increasing-soil-organic-carbon.pdf.

Ge, Y., J.A. Thompson, C.L. Morgan and S.W. Searcy (2007). VNIR diffuse reflectance spectroscopy for agricultural soil property determination based on regression-kriging. Transactions of the American Society of Agricultural and Biological Engineers, 50, 1081–1092.

Hengl, T. (2009). A practical guide to geostatistical mapping. Office for Official Publications of the European Communities, Luxembourg.

Jenny, H. (1992). Factors of soil formation: A system of quantitative pedology. Dover, New York.

Krige, D.G. (1951). A Statistical approach to some basic mine valuation problems on Witwatersrand. Journal of chemical, Metallurgical and Mining Society of South Africa, 52(6), 119–139.

Kumar, P., P.C. Pandey, B.K. Singh, S. Katiyar, V.P. Mandal, M Rani, V. Tomar and S. Patairiya (2016). Estimation of accumulated soil organic carbon stock in tropical forest using geospatial strategy. Egyptian Journal of Remote Sensing and space Sciences, 19(1), 109–123.

Kumar, S. and R. Lal (2011). Mapping the organic carbon stocks of surface soils using local spatial interpolator. Journal of environmental monitoring, 13, 3128–3135.

Lagacherie, P and A.B. McBratney (2007). Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In Lagacherie, P., A.B. McBratney and M.Voltz, (eds.), Digital soil mapping: An Introductory Perspective, Developments in Soil Science, 31,pp. 389–399.

McBratney, A.B., M.L. Mendonca-Santos and B. Minasny (2003). On digital soil mapping. Geoderma, 117 (1-2), 3–52.

Mora-Vallejo, A., L. Claessens, J. Stoorvogel and G.B.M. Heuvelink (2008). Small scale digital soil mapping in southeastern Kenya. Catena, 76(1), 44–53.

Mulder, V.L., S. De Bruin, M.E Schaepman and T.R. Mayr (2011). The use of remote sensing in soil and terrain mapping-A Review. Geoderma, 162, 1–19.

Peng, G., W. Bing, G. Guangpo and Z. Guangcan (2013). Spatial distribution of soil organic carbon and total nitrogen based on GIS and geostatistics in a small watershed in a hilly area of northern China. PLoS ONE, 8(12), 1–9.

Razakamanarivo, R.H., C. Grinand, M.A. Razafindrakoto, M. Bernoux and A. Albrecht (2011). Mapping organic carbon stocks in eucalyptus plantations of the central highlands of Madagascar: A multiple regression approach. Geoderma,162, 335–346.

Rouse, J.W., R.H. Haas, J.A. Schell and D.W. Deering (1973). Monitoring vegetation systems in the great plains with ERTS. In proceedings of the third Earth Resources Technology Satellite symposium, Goddard Space Flight Center, National Aeronatics and space Administration, SP-351, Science and Technical Information Office, NASA, Washington, D.C., pp. 309–317.

Sumfleth, K. and R. Duttmann (2008). Prediction of soil property distribution in paddy soil landscape using terrain data and satellite information as indicators. Ecological Indicators, 8, 485–501.

Thompson,J.A., S. Roecker, S. Grunwald and P.R. Owens (2012). Digital Soil Mapping: Interactions with and applications for hydropedology. Academic Press, Boston, pp. 665–709.

UNEP. (2009). Kenya: Atlas of our changing environment. Division of Early Warning and Assessment (DEWA), United Nations Environment Programme (UNEP).

Wilson, J.P. and J.C. Gallant (2000). Secondary topographic attributes. In Wilson, J.P. and J.C. Gallant (eds.), Terrain Analysis: Principles and Applications, John Wiley and Sons, Inc., New York, pp. 87–132.

Zhang, S., Y. Huang, C. Shen, H. Ye and Y. Du (2012). Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. Geoderma, 171, 35–43.