Spatially distributed modelling and mapping of soil organic carbon and total nitrogen stocks in the Eastern Mau Forest Reserve, Kenya

Kennedy WERE[1,2], Bal Ram SINGH[3], Øystein Bjarne DICK[1]

1. Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432, Ås, Norway;
2. Kenya Agricultural and Livestock Research Organisation, Kenya Soil Survey, P.O. Box 14733-00800, Nairobi, Kenya;
3. Department of Environmental Sciences, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432, Ås, Norway

Abstract: Detailed knowledge about the estimates and spatial patterns of soil organic carbon (SOC) and total nitrogen (TN) stocks is fundamental for sustainable land management and climate change mitigation. This study aimed at: (1) mapping the spatial patterns, and (2) quantifying SOC and TN stocks to 30 cm depth in the Eastern Mau Forest Reserve using field, remote sensing, geographical information systems (GIS), and statistical modelling approaches. This is a critical ecosystem offering essential services, but its sustainability is threatened by deforestation and degradation. Results revealed that elevation, silt content, TN concentration, and Landsat 8 Operational Land Imager band 11 explained 72% of the variability in SOC stocks, while the same factors (except silt content) explained 71% of the variability in TN stocks. The results further showed that soil properties, particularly TN and SOC concentrations, were more important than that other environmental factors in controlling the observed patterns of SOC and TN stocks, respectively. Forests stored the highest amounts of SOC and TN (3.78 Tg C and 0.38 Tg N) followed by croplands (2.46 Tg C and 0.25 Tg N) and grasslands (0.57 Tg C and 0.06 Tg N). Overall, the Eastern Mau Forest Reserve stored approximately 6.81 Tg C and 0.69 Tg N. The highest estimates of SOC and TN stocks (*hotspots*) occurred on the western and northwestern parts where forests dominated, while the lowest estimates (*coldspots*) occurred on the eastern side where croplands had been established. Therefore, the *hotspots* need policies that promote conservation, while the *coldspots* need those that support accumulation of SOC and TN stocks.

Keywords: soil organic carbon; total nitrogen; carbon sequestration; climate change; digital soil mapping; Eastern Mau

1   Introduction

Soil organic carbon (SOC) and total nitrogen (TN) are key determinants of biogeochemical

cycling, as well as soil quality and properties (Obade and Lal, 2013; Wang *et al.*, 2013; Yang *et al.*, 2014). They vary spatially and temporally in response to a mix of climatic, edaphic, biotic, topographical, and lithological factors. Such dynamics also affect the contributions of SOC and TN to atmospheric greenhouse gases, particularly carbon dioxide (CO_2) and nitrous oxide (N_2O). The world soils contain about 1500 petagrams of carbon (Pg C) to 1 m depth (1 Pg = 10^{15} g), which is twice the amount of C in the atmospheric pool and three times the amount in the biotic pool (Lal, 2004; Smith, 2004, 2008). This implies that even slight changes in SOC pool can significantly affect global C cycle and climate. Therefore, current research is geared towards quantifying and mapping SOC and TN stocks, in space and over time, with a view to understanding climate change and land degradation processes. Unfortunately, the traditional soil mapping techniques are expensive, time-consuming, and yield coarse qualitative information (Mora-Vallejo *et al.*, 2008; Mehrjardi *et al.*, 2014). Consequently, there is increasing effort in the emerging field of digital soil mapping (DSM) to develop, evaluate, and apply new techniques for spatial prediction and mapping of soil properties.

The existing DSM techniques fall into two categories, namely (1) measure and multiply (MM) and (2) soil-landscape modelling (SLM) techniques (Mishra *et al.*, 2010; Cambule *et al.*, 2014). In MM approach, the study area is stratified and then the point estimates of a target soil property within a stratum are averaged and multiplied by the stratum's area. In contrast, in SLM approach, the spatial variability of a target soil property is explained by its relationships with soil-forming factors, such as topography, climate, land use, vegetation, parent material, and soil type. In particular, field observations and ancillary environmental data are used to calibrate an empirical model, which is then applied to generate a prediction surface of the target soil variable (Mishra *et al.*, 2010; Li *et al.*, 2013b; Cambule *et al.*, 2014). SLM approach has been boosted by the improvements in computer technology and accessibility to inexpensive environmental data from remote sensors and existing spatial databases. Although MM approach is simple, it ignores the complex interactions of environmental factors with the target soil variables, which account for the spatial variability. Thus, MM approach yields predictions with lesser accuracy than SLM approach. McKenzie and Ryan (1999), McBratney *et al.* (2003), and Scull *et al.* (2003) have provided detailed reviews of DSM.

Literature is replete with examples of SLM techniques that have been applied so far to model and map the spatial patterns of SOC and TN stocks. The techniques range from multiple linear regression (Lesch and Corwin, 2008; Meersmans *et al.*, 2008) and partial least square regression (Selige *et al.*, 2006; Amare *et al.*, 2013) to generalized linear models (Yang *et al.*, 2008), classification and regression trees (Kheir *et al.*, 2010; Martin *et al.*, 2011; Razakamanarivo *et al.*, 2011), kriging (Wu *et al.*, 2009; Zhang *et al.*, 2010; Liu *et al.*, 2011; Li *et al.*, 2013a; Cambule *et al.*, 2014), regression-kriging (Hengl *et al.*, 2004, 2007; Lamsal *et al.*, 2006; Mora-Vallejo *et al.*, 2008; Sumfleth and Duttmann, 2008; Li, 2010; Vasques *et al.*, 2010a, 2010b; Dorji *et al.*, 2014; Martin *et al.*, 2014), geographically weighted regression (Mishra *et al.*, 2010; Zhang *et al.*, 2011; Mishra and Riley, 2012; Kumar *et al.*, 2013; Wang *et al.*, 2013), geographically weighted regression-kriging (Kumar *et al.*, 2012), neural networks (Malone *et al.*, 2009; Jaber and Al-Qinna, 2011; Li *et al.*, 2013b), random forests (Grimm *et al.*, 2008; Vågen and Winowiecki, 2013a, 2013b), rule-based models

(Lacoste *et al*., 2014), and linear mixed models (Doetterl *et al*., 2013; Karunaratne *et al*., 2014). Of these techniques, multiple linear regression (MLR) is most popular because of its simplicity, computational efficiency, and straightforward interpretation (Li *et al*., 2013b). However, its assumptions of spatial stationarity in the effects of environmental variables and spatial independence in the target soil properties are mostly violated leading to misspecification of prediction models. Hybrid methods, particularly regression-kriging (MLRK), which combines ordinary kriging with MLR are also gaining currency in digital soil mapping because of their detailed results and lower prediction errors compared to pure geostatistical, or statistical methods (Hengl *et al*., 2004). Geographically weighted regression (GWR) is the most recent technique, which has drawn the attention of environmental scientists. GWR was designed to deal with the spatially varying relationships between the target and environmental variables (i.e., spatial non-stationarity); hence, the estimated parameters also vary spatially (Wang *et al*., 2013). Even though some comparative studies have shown that it outperforms MLRK in spatial prediction of SOC stocks (Mishra *et al*., 2010), the application of GWR is still limited. Few studies have also attempted to couple GWR with kriging (geographically weighted regression-kriging; GWRK) to predict the spatial distribution of environmental phenomena; for example, urban heat island in Wrocław, Poland (Syzomanoski and Kryza, 2012) and SOC stocks in Pennsylvania State, USA (Kumar and Lal, 2011; Kumar *et al*., 2012).

The objective of this study was to estimate and map the spatial distribution of SOC and TN stocks to 30 cm depth in the Eastern Mau Forest Reserve by integrating field sampling, remote sensing, geographical information systems (GIS), and statistical modelling. The 30 cm depth is consistent with the Intergovernmental Panel on Climate Change (IPCC) guidelines (IPCC, 2006). The Eastern Mau Forest Reserve was selected because it had undergone wanton deforestation and degradation since the mid-1990s owing to ill-advised forest excisions and illegal loggings, encroachments, and charcoal burning (Government of Kenya 2009; UNEP 2009). Despite this, no complete studies had been undertaken to quantify the storage and map the spatial patterns of SOC and TN. This study aimed to bridge this gap and contribute information for designing spatially-targeted, effective, and sustainable strategies for ecosystem restoration and management.

2   Materials and methods

2.1   Study area

The Eastern Mau Forest Reserve, which covers approximately 650 km^2, is part of East Africa's largest closed-canopy indigenous montane forest, and Kenya's key water catchment area. It is bounded by the latitudes 0º15´–0º40´S and the longitudes 35º40´–36º10´E (Figure 1) with the altitudes ranging from 2210 to 3070 m above sea level. The climate is cool and humid; that is, the average annual temperatures vary from 9.8°C to 17.5°C, while the average annual rainfall ranges between 935 and 1287 mm (Jaetzold *et al*., 2010). The rainfall distribution pattern is tri-modal with peaks in April, August, and November. The Njoro, Naishi, and Larmudiac Rivers drain the eastern slopes into Lake Nakuru, while the Nessuiet and Rongai flow northwards into Lake Bogoria, and Baringo, respectively. The area's

physiography is characterized by major scarps and uplands comprising pyroclastic rocks (e.g., pumice tuffs) of Tertiary-Quaternary volcanic age. These soft light brown rocks have insets of yellow pumice and angular trachyte, which decompose into deep to very deep, and dark reddish brown clayey soil aggregates (McCall, 1967). The soils, classified as *Mollic Andosols* by FAO, are friable and smeary with humic topsoils (Jaetzold *et al*., 2010). The dominant land cover types are forests, grasslands, and croplands (Figure 1). The floristic composition of forests and grasslands comprise indigenous tree species, such as *Prunus africana, Arundinaria alpina, Juniperus procera, Olea europaea* ssp. *africana, Olea capensis* ssp. *hochstetteri, Podocarpus latifolius, Nuxia congesta, Clematis hirsuta, Schefflera volkensii*, and *Dombeya torrida*, exotic tree species like *Pinus patula* and *Cupressus lusitanica*, and grass species, such as *Pennisetum clandestinum*. The major crops grown are maize (*Zea mays*), beans (*Phaseolus vulgaris*), wheat (*Triticum aestivum*), and potatoes (*Solanum tuberosum*) (Were *et al*., 2015).
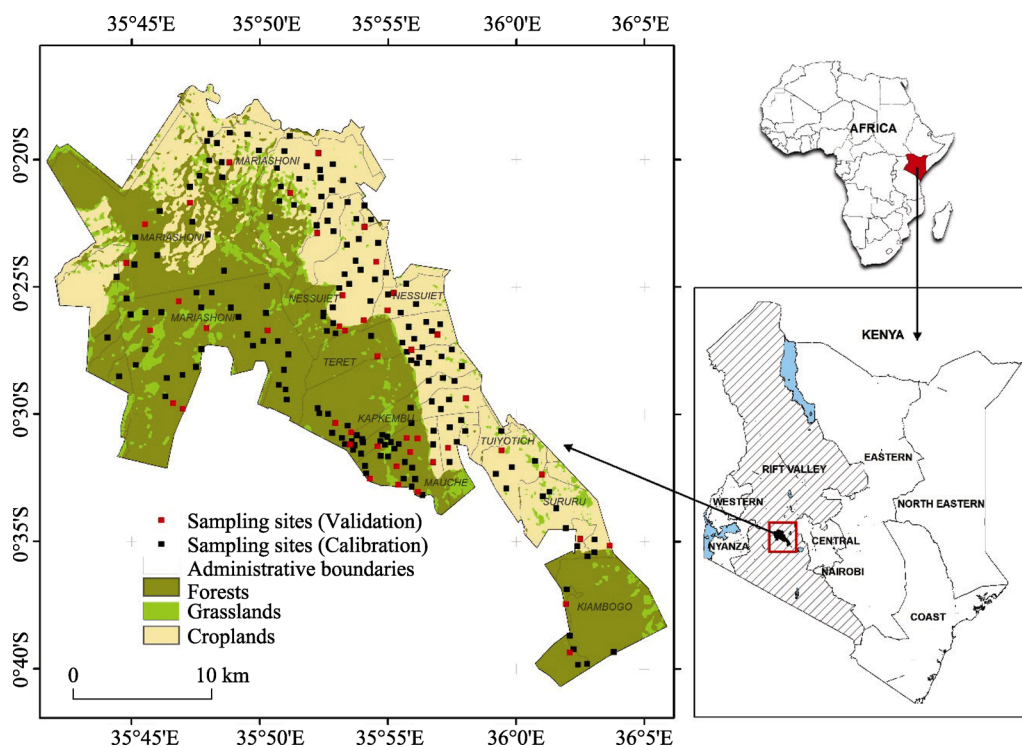

Figure 1   Geographical location of the study area

2.2   Data sources and pre-processing

Figure 2 summarizes the data sources and spatial modelling framework of this study. The overall methodology involved seven major steps: (1) soil sampling and analysis, (2) preparation of the environmental predictors and target soil variables, (3) calibration of the regression-based models, (4) application of the models, (5) interpolation of the regression-based residuals and their addition to the fitted trend surfaces, (6) validation, and (7) production of the thematic maps for SOC and TN stocks.
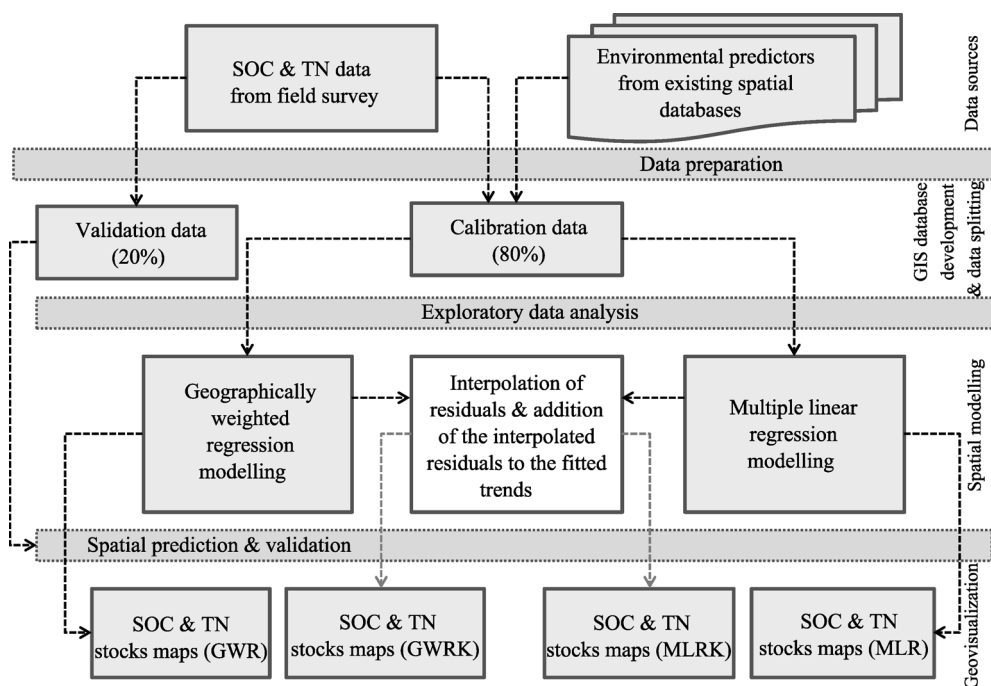
Figure 2   Illustration of the data sources and modelling framework

2.2.1   Soil sampling and analysis

The soil sampling campaign was conducted between June and August 2012. Before the campaign, sampling points were generated in a completely randomized design using agro-ecological zones map as the base in a GIS environment. A map showing the distribution of these sampling points was created and used in the field. At each sampling point, an auger was used to collect samples at 0–15 cm and 15–30 cm depths from the centre and corners of a plot measuring 30 m × 30 m. The samples taken from corresponding depths in a plot were mixed thoroughly and bulked into one composite sample of about 500 g. To determine bulk density (BD), a core sampler (5 cm in diameter and 5 cm in height) was used to collect one undisturbed sample at the centre of each plot and at each depth. The geographical coordinates, elevation, vegetation, and land management practices were also recorded. A total of 320 soil samples were collected from 160 sampling plots to analyze the chemical and physical properties, and a similar number to determine BD at the National Agricultural Research Laboratories. Supplementary soil data that had been collected similarly from 60 other sampling plots to assess the effects of land cover changes on SOC and TN stocks (Were *et al*., 2015) were also used. Overall, soil data from 220 sampling plots (Figure 1) were used to model the spatial distribution of SOC and TN stocks.

The soil samples were air-dried, ground, and sieved through a 2 mm mesh. SOC concentrations, TN concentrations, and BD were then determined using the Walkley-Black wet oxidation method (Nelson and Sommers, 1982), Kjeldahl digestion method (Bremner and Mulvaney, 1982), and core method (Blake, 1965), respectively. These three properties were used to calculate SOC and TN stocks (i.e., the target variables) at each depth. Additional soil properties were also analyzed. The hydrometer method (Day, 1965) was used to determine particle size distribution, while the Mehlich method (Okalebo *et al*., 2002) was used to esti-

mate phosphorous (P) content. A flame-photometer was used to measure potassium (K) content, an atomic absorption spectrophotometer to measure the contents of calcium (Ca) and magnesium (Mg), and a pH meter to measure pH (1:2.5 soil-water) (Okalebo *et al*., 2002).

Eq. (1) was used to calculate SOC stocks (Mg C ha^{-1}) for each depth (Aynekulu *et al*., 2011):

$$SOC_{st} = \frac{SOC}{100} \times BD \times D \times 100 \tag{1}$$

where SOC_{st} is the soil organic carbon stock (Mg C ha^{-1}), SOC is the soil organic carbon concentration (%, which is then converted to g C g^{-1} soil), BD is the bulk density (g cm^{-3}), D is the depth (cm), and 100 is the multiplication factor to convert the SOC per unit area from g C cm^{-2} to Mg C ha^{-1}. Stone contents were negligible due to the softness of the volcanic rocks; hence, are not accounted for in Eq. (1). Similarly, TN stocks (TN$_{st}$; Mg N ha^{-1}) for each depth were computed by substituting TN for SOC in Eq. (1). The SOC and TN stocks in the surface (0–15 cm) and subsurface soils (15–30 cm) were summed up to obtain the total stocks to 30 cm depth.

2.2.2   Remote sensing and GIS analysis

Twenty candidate environmental predictors that had been selected *a priori* based on the *scorpan* conceptual model (McBratney *et al*., 2003) were obtained from existing spatial databases and fieldwork. The *scorpan* model captures six key soil-forming factors; namely, soil properties (s), climate (c), organisms (o), topography (r), parent material (p), age (a), and space (n). Table 1 provides the sources of temperature, rainfall, land cover, elevation (digital elevation model; DEM), and Landsat 8 Operational Land Imager (OLI) data. Although Landsat 8 OLI imagery was captured in 2013, whereas fieldwork was conducted in 2012 (i.e., when Landsat 5 TM operational imaging had ended, and before the launch of Landsat 8 OLI), it was assumed that soil properties had not changed significantly within such a short time to affect their spectral response. Slope, curvature, aspect, and compound topographic index (CTI) were also extracted from the DEM. Eq. (2) was used to extract CTI:

$$CTI = \ln\left(\frac{A_s}{\tan\beta}\right) \tag{2}$$

where A_s is the upslope area and β is the slope (McKenzie and Ryan, 1999).

The Normalized Difference Vegetation Index (NDVI) (Eq. (3)) was derived after the digital numbers of OLI band 4 (red; R) and 5 (near infra-red; NIR) were converted to top-of-atmosphere reflectance (ρ) (http://landsat.usgs.gov/Landsat8_Using_Product.php).

$$NDVI = \frac{\rho NIR - \rho R}{\rho NIR + \rho R} \tag{3}$$

Moreover, principal component analysis was performed to reduce dimensionality, while capturing as much variability as possible from OLI bands 2, 3, 4, 5, 6, and 7. The first principal component (PC1), which explained 98% of the variability, was chosen for spatial modelling. All raster grids were transformed to Universal Transverse Mercator coordinate system (UTM WGS84 Zone 36S) prior to extracting the area of interest from each. The 1 km climatic grids were resampled to 30 m to synchronize them with the rest. Soil data from the

laboratory, including sand content, silt content, clay content, TN concentrations, C concentrations, pH, Mg, Ca, P, and K were also integrated into the GIS database both as points in vector format and raster grids after interpolation by ordinary kriging. Ordinary kriging method was chosen because it has been widely used to optimize the prediction of soil properties at unvisited locations in pedological studies (Chaplot *et al*., 2010; Pachomphon *et al*., 2010; Kumar and Lal, 2011; Tesfahunegn *et al*., 2011; Marchetti *et al*., 2012; Elbasiouny *et al*., 2014). Finally, the attribute values of all the other raster grids (e.g., slope, rainfall, temperature) were extracted to the points to allow the analysis of relationships between the target variables and environmental predictors.

Table 1 Properties of the environmental predictors for spatial modelling

Variables	Data format	Date	Source	Scale	Soil-forming factor
Target variables					
1. SOC stocks	Points	2012	Field work		
2. TN stocks	Points	2012	Field work		
Predictor variables					
1. SOC concentration	Raster	2012	Interpolated field data	30 m	S
2. TN concentration	Raster	2012	Interpolated field data	30 m	S
3. Magnesium	Raster	2012	Interpolated field data	30 m	S
4. Potassium	Raster	2012	Interpolated field data	30 m	S
5. Calcium	Raster	2012	Interpolated field data	30 m	S
6. Clay content	Raster	2012	Interpolated field data	30 m	S
7. Silt content	Raster	2012	Interpolated field data	30 m	S
8. Sand content	Raster	2012	Interpolated field data	30 m	S
9. pH	Raster	2012	Interpolated field data	30 m	S
10. Elevation	Raster	–	ASTER GDEM http://gdem.ersdac.jspacesystems.or.jp/	30 m	R
11. Slope	Raster	–	ASTER GDEM	30 m	R
12. Aspect	Raster	–	ASTER GDEM	30 m	R
13. Curvature	Raster	–	ASTER GDEM	30 m	R
14. CTI	Raster	–	ASTER GDEM	30 m	S
15. Temperature	Raster	1950–2000	www.worldclim.org	1 km	C
16. Rainfall	Raster	1950–2000	www.worldclim.org	1 km	C
17. Surface reflectance & thermal emission	Raster	30.05.2013	Landsat 8 OLI (bands 2, 3, 4, 5, 6, 7, 10 & 11) http://earthexplorer.usgs.gov/	30 m	C, S
18. NDVI	Raster	30.05.2013	Landsat 8 OLI (bands 4 & 5)	30 m	O
19. PC bands	Raster	30.05.2013	Landsat 8 OLI (bands 2, 3, 4, 5, 6 & 7)	30 m	S
20. Land cover	Raster	17.01.2011	Landsat 5 TM; Were *et al*. (2013)	30 m	O

Note: SOC=soil organic carbon; TN=total nitrogen; CTI=compound topographic index; NDVI=normalized difference vegetation index; PC= principal component; S=soil properties; C=climate; O=organisms; and, R=topography

2.3 Spatial modelling

2.3.1 Exploratory data analysis

Firstly, descriptive statistics of the target variables were estimated. This was followed by

pairwise Pearson's product-moment correlation analysis to detect collinearity between the predictor variables, as well as their correlation with the target variables. Predictors entered the model only if their correlation with the target variable was, or exceeded 0.2. Again, two highly correlated predictors ($r \geqslant 0.8$) were retained in a model only if their variance inflation factors (VIFs) did not exceed 10; otherwise, one was removed (Montgomery *et al.*, 2006).

2.3.2   Model development

The processed point dataset from the GIS database ($n=220$) was randomly split into two: (i) training data ($n=176$) to calibrate the models of SOC and TN stocks and, subsequently, create prediction surfaces, and (ii) test data ($n=44$) to validate the surfaces. Multiple linear regression (MLR), multiple linear regression-kriging (MLRK), geographically weighted regression (GWR), and geographically weighted regression-kriging (GWRK) techniques were used to calibrate the models.

2.3.2.1 MLR and MLRK

Eq. 4 gives the form of MLR model used to define the relationship between the target variables and predictors at the sampled locations (Montgomery *et al.*, 2006):

$$y_i = \beta_0 + \sum\nolimits_k \beta_k x_{ik} + \varepsilon_i \tag{4}$$

where y_i is the value of the target variable at ith location, β_0 is the regression coefficients, x_i is the value of the predictor variable at ith location, k is the number of predictors, and ε_i is the error term.

Full MLR models were fitted by ordinary least square estimator, after which the best subset models were ranked based on Mallow's C_p using all possible regressions variable selection method. The final reduced model for each target variable was selected from the three best subset models after scrutiny for physical correctness. T-tests were used to determine significance of the model parameters, while analysis of variance *F*-tests were used to determine significance of the regression at a level of 5%. The adequacy of the models was checked using residual plots, normal probability plots, measures of influence and leverage (e.g., Cook's *D*), VIFs, and coefficients of determination (R^2). Finally, the models were applied to create prediction surfaces of the target variables.

To develop MLRK models and prediction surfaces, the deterministic component of the target variable modelled by MLR (Eq. 4) and the spatially correlated stochastic component modelled by kriging the MLR residuals were summed up. Eq. 5 summarizes the MLRK model (Hengl *et al.*, 2004; Vasques *et al.*, 2010a):

$$y(u_i v_i) = m_{\mathrm{mlr}}(u_i v_i) + \varepsilon'_{\mathrm{ok}}(u_i v_i) + \varepsilon''(u_i v_i) \tag{5}$$

where $y(u_i v_i)$ is the target variable at location $(u_i v_i)$, $(u_i v_i)$ is the coordinates of the ith location, $m_{mlr}(u_i v_i)$ is the deterministic component, $\varepsilon'_{ok}(u_i v_i)$ is the spatially correlated random component, and $\varepsilon''(u_i v_i)$ is the spatially independent residuals error (noise).

2.3.2.2 GWR and GWRK

Similar predictors were used to build GWR models to allow comparison with MLR models. Unlike MLR that assumes spatial stationarity and locational independence, GWR takes into account the spatial location of samples. This allows the estimated parameters to vary locally; hence, representing the spatially varying relationships between the target and predictor variables much better (Zhang *et al.*, 2011). Eq. 6 expresses the form of GWR model (Fotheringham *et al.*, 2002):

$$y_i = \beta_0(u_i v_i) + \sum\nolimits_k \beta_k(u_i v_i)x_{ik} + \varepsilon_i \tag{6}$$

where y_i is the value of the target variable at ith location, $(u_i v_i)$ is the coordinates of the ith location, $\beta(u_i v_i)$ are the regression coefficients, x_i is the value of the predictor variable at ith location, k is the number of predictors, and ε_i is the error term.

   The GWR parameters were estimated using adaptive (*bi*-square) spatial kernel functions, where the bandwidth of the samples included for estimation varied with sample density (Fotheringham *et al*., 2002; Wang *et al*., 2013). The optimal bandwidth was determined by Akaike Information Criterion (AICc). Lastly, the estimated parameters were applied to create spatially distributed maps of the target variables.

   To develop GWRK models, the deterministic component of the target variable modelled by GWR (Eq. 6) and the spatially correlated stochastic component represented by kriged GWR residuals were added. Eq. 7 provides the form of GWRK model (Kumar *et al*., 2012):

$$y(u_i v_i) = m_{gwr}(u_i v_i) + \varepsilon'_{ok}(u_i v_i) + \varepsilon''(u_i v_i) \tag{7}$$

where $y(u_i v_i)$ is the target variable at location $(u_i v_i)$, $m_{gwr}(u_i v_i)$ is the deterministic component, $\varepsilon'_{ok}(u_i v_i)$ is the spatially correlated random component, and $\varepsilon''(u_i v_i)$ is the spatially independent residuals error.

   Additionally, Moran's *I* was calculated to measure spatial autocorrelation in the residuals. The Moran's *I* values range from −1 to +1, with 0 indicating absence of spatial autocorrelation, positive values indicating positive autocorrelation, and negative values indicating the opposite (Overmars *et al*., 2003).

2.3.3 Model evaluation

A ten-fold validation procedure was employed to evaluate the prediction surfaces produced by the fitted models. In this procedure, the original dataset (*n*=220) was randomly split into training (*n*=176) and testing (*n*=44) datasets ten times. The training data were used to calibrate models and generate prediction surfaces, while the testing data were used to validate them. Root mean squared error (RMSE) and mean error (ME) were computed from the differences between the predicted and measured values to determine the precision and bias of the predictions, respectively (Eqs. 8 and 9):

$$\text{RMSE} = \sqrt{\sum\nolimits_1^n (y_i - \hat{y}_i)^2 / n} \tag{8}$$

$$\text{ME} = \sum\nolimits_1^n (y_i - \hat{y}_i)/n \tag{9}$$

where $\hat{y}_i$ is the estimated value, y_i is the measured value, and n is the number of measured values in the testing data. The ME should be close to zero, while RMSE should be as small as possible. Average ME and RMSE values of the ten-fold validation are reported in this paper. Statistical validation was supplemented by visual inspection of the spatial patterns of the target variables.

   The method with the lowest prediction error indices was chosen to provide the final estimates of SOC and TN stocks for the Eastern Mau Forest Reserve. To estimate the stocks under different land cover types, the prediction surfaces were overlaid with the land cover map of the area and zonal statistics extracted. All data management, analyses, and geovisualization functions were performed using ArcGIS® 10.1, ERDAS IMAGINE® 2013, GWR4, Microsoft Excel® 2010, and R version 3.0.1 (R Core Team, 2013) with its add-in packages

"sp" (Pebesma *et al*., 2013) and "automap" (Hiemstra, 2013).

3   Results

3.1   Exploratory data analysis

Table 2 presents the numerical summaries of SOC and TN stocks at the sampled locations. Soil organic carbon stocks range from 42.0 to 193.4 Mg C ha^{-1} with a mean of 102.7 Mg C ha^{-1}. The standard deviation is 24.6 Mg C ha^{-1} and coefficient of variation is 23.9%, which suggests moderate variability. The skewness is 0.39 indicating an approximately normal distribution of the data, whereas kurtosis is 0.97 implying less peaked values in the distribution of the data. Similarly, TN stocks vary from 4.2 to 19.1 Mg N ha^{-1} with a mean of 10.3 Mg C ha^{-1}. The standard deviation is 2.4 Mg C ha^{-1} and coefficient of variation is 23.8%, while skewness and kurtosis are 0.28 and 0.76, respectively. Again, this shows moderate variability and minimal departure from normality. Hence, spatial modelling of both SOC and TN stocks was performed using the raw, non-transformed data. Pearson's correlation analysis shows that some of the predictors were highly correlated ($r \geqslant 0.80$), and that only 13 met the threshold correlation ($r \geqslant 0.20$) with the target variables (Table 3). Thus, the candidate

Table 2   Descriptive statistics of SOC and TN stocks (0–30 cm)

Variable	n	Mean	Median	SD	CV (%)	Min.	Max.	Range	Skewness	Kurtosis
SOC$_{st}$	220	102.7	103.2	24.6	23.9	42.0	193.4	151.4	0.39	0.97
TN$_{st}$	220	10.3	10.3	2.4	23.8	4.2	19.1	14.9	0.28	0.76

SD=standard deviation; CV=coefficient of variation; n=number of observations

Table 3   Pearson's correlation coefficient between the predictors and target variables selected for spatial modelling

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. SOC stock	1.00														
2. TN stock	0.99	1.00													
3. TN content	0.84	0.85	1.00												
4. SOC content	0.85	0.84	0.99	1.00											
5. Silt	−0.41	−0.42	−0.56	−0.55	1.00										
6. Magnesium	0.35	0.35	0.44	0.44	−0.36	1.00									
7. Clay	0.28	0.29	0.40	0.39	−0.61	0.08	1.00								
8. Temperature	−0.50	−0.50	−0.63	−0.63	0.28	−0.04	−0.34	1.00							
9. Rainfall	0.44	0.45	0.56	0.55	−0.23	0.25	0.10	−0.61	1.00						
10. Elevation	0.51	0.51	0.65	0.65	−0.30	0.06	0.35	−0.99	0.65	1.00					
11. Aspect	0.22	0.23	0.18	0.17	−0.08	0.01	0.02	−0.16	0.16	0.16	1.00				
12. NDVI	0.30	0.30	0.39	0.39	−0.25	0.07	0.24	−0.50	0.25	0.50	0.11	1.00			
13. Land cover	−0.48	−0.48	−0.54	−0.53	0.31	0.00	−0.41	0.83	−0.46	−0.84	−0.16	−0.56	1.00		
14. PC1	−0.48	−0.48	−0.52	−0.52	0.15	−0.03	−0.23	0.71	−0.50	−0.73	−0.28	−0.32	0.74	1.00	
15. Landsat 8 OLI band 11	−0.58	−0.58	−0.65	−0.65	0.35	−0.09	−0.37	0.81	−0.56	−0.84	−0.29	−0.63	0.89	0.82	1.00

Note: SOC=soil organic carbon; TN=total nitrogen; NDVI=normalized difference vegetation index; PC= principal component. Bold form shows that the correlation coefficient between the predictors and target variables exceeded the threshold value ($r > 0.2$).

predictors for developing full models reduced from 20 to 13; namely, elevation, aspect, rainfall, temperature, TN, Mg, silt, clay, land cover, PC1, NDVI, and OLI band 10 and 11 (Table 3). The predictors that were highly correlated include: temperature and elevation, temperature and land cover, elevation and land cover, elevation and OLI band 11, land cover and OLI band 11, and PC1 and OLI band 11. Therefore, VIFs of the predictors in the reduced models were also checked for multi-collinearity.

3.2   Spatial models

3.2.1   MLR

The subset SOC and TN stock models selected by all possible regressions method had Mallow's C_p values (3.8 and 10.3, respectively) that were lower than the number of model parameters. Table 4 provides the summaries of the models. Elevation, silt content, TN concentration, and OLI band 11 have significant effects on SOC stocks explaining 72% of its variability (adjusted R^2=0.72), whereas OLI band 11, elevation, and SOC concentration have significant effects on TN stocks explaining 71% of its variability (adjusted R^2=0.71). Total nitrogen concentrations have the largest magnitude of effect on SOC stocks, while SOC concentrations have the largest magnitude of effect on TN stocks. OLI band 11 has the smallest magnitude of effect on both TN and SOC stocks. Visual analysis of the residual and normal probability plots indicated equality of variance and normality in the distribution of error terms, as well as linearity in the model parameters. The few outliers that were evident on these plots were not sufficiently influential to warrant their removal from the data because Cook's D indices were less than 1. Despite the high correlation between elevation and OLI band 11 (r=0.84), the associated VIFs do not exceed 10 in the models. Moran's indices are very low, but statistically significant; that is, 0.11 (p=0.0141) and 0.08 (p=0.0550) for SOC and TN stocks models, respectively. This shows very weak tendency for clustering of similar residuals. The high nugget-to-sill ratios (NSRs) of 78.6% for the residuals of SOC stock model, and 73.7% for the residuals of TN stock model to a range of 4 km (Table 6 and Figure 3) also demonstrate this weak spatial structure. However, the spatial dependency of SOC and TN stocks data are moderate (NSRs of 58.1% and 45.6%, respectively) to a range

Table 4   Parameter estimates of the MLR models

Parameter	SOC stocks model					TN stocks model				
	Estimate	SE	t value	Pr (>|t|)	VIF	Estimate	SE	t value	Pr (>|t|)	VIF
Intercept	143.502	50.757	2.827	0.0053**	–	16.741	5.131	3.263	0.0013**	–
Silt	0.443	0.202	2.191	0.0298*	1.531	–	–	–	–	–
Band 11	–0.003	0.001	–2.360	0.0194*	3.511	–0.000	0.000	–2.475	0.0143*	3.489
Elevation	–0.022	0.009	–2.503	0.0133*	3.613	–0.002	0.001	–2.305	0.0223*	3.558
TN	178.200	12.269	14.524	0.0000***	2.471	–	–	–	–	–
SOC	–	–	–	–	–	1.597	0.106	15.103	0.0000***	1.807
Adjusted R^2	0.72					0.71				
RMSE	13.07					1.33				
Moran's I	0.11					0.08				

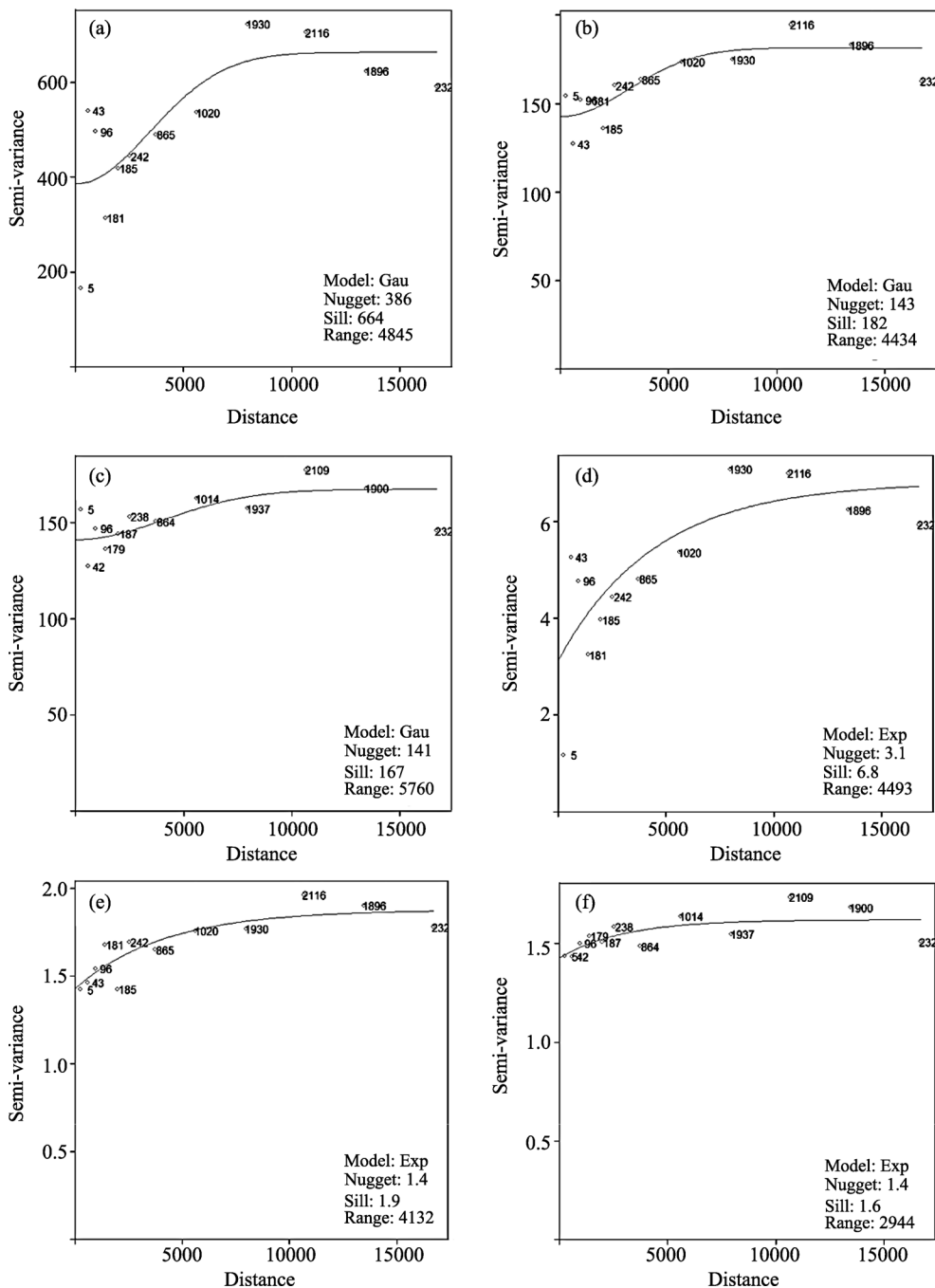Significance codes: 0 '***' 0.001 '**' 0.01 '*'

Figure 3   (a) Experimental variograms (points) and fitted models (lines) of SOC stocks (b) MLR$_{soc}$ residuals (c) GWR$_{soc}$ residuals (d) TN stocks (e) MLR$_{tn}$ residuals, and (f) GWR$_{tn}$ residuals

of 4.8 km. Total sills for the residuals are 182 Mg C ha^{-1} and 1.9 Mg N ha^{-1}, which are close to the variance (σ^2) estimates of the respective MLR models (170.6 Mg C ha^{-1} and 1.7 Mg N ha^{-1}).

3.2.2   GWR

Table 5 shows the summaries of parameter estimates of the GWR models for SOC and TN

stocks. Once again, TN concentrations have the highest magnitude of effect on SOC stocks, while SOC concentrations have the highest magnitude of effect on TN stocks. OLI band 11 has the lowest magnitude of effect on both TN and SOC stocks. Unlike the MLR models, the GWR models show that the magnitude of the effects of predictors varies with sampling location. This means that the interactions between the target variables and environmental factors are spatially non-stationary. Hence, the summaries of GWR estimates are given in ranges instead of mean values. Although the magnitudes vary spatially, the directions of the effects are constant. The Moran's indices are lower than for MLR models; that is, 0.06 (p=0.1798) and 0.02 (p=0.1620) for SOC and TN stocks models, respectively. This indicates that the GWR residuals are approximately uncorrelated, and that the models are better specified than the MLR models. The high NSRs of 84.4% for the GWR residuals of SOC stocks, and 87.5% for the GWR residuals of TN stocks to a range of 6 km and 3 km, respectively (Table 6 and Figure 3), also reveal this weak spatial dependency. Total sills for the residuals are 167 Mg C ha^{-1} and 1.6 Mg N ha^{-1}, which are lower than for the MLR residuals. However, the range is shorter for the GWR residuals than for the MLR residuals only in the case of TN stocks.

Table 5 Parameter estimates of the GWR models

Parameter	SOC stocks model					TN stocks model				
	Mean	SD	Min.	Max.	Range	Mean	SD	Min.	Max.	Range
Intercept	129.525	31.412	50.031	199.881	149.851	14.790	4.771	8.037	26.423	18.387
Silt	0.436	0.115	0.203	0.614	0.411	–	–	–	–	–
Band 11	−0.003	0.001	−0.004	−0.001	0.004	−0.000	0.000	−0.000	−0.000	0.000
Elevation	−0.021	0.007	−0.041	−0.004	0.037	−0.002	0.001	−0.005	−0.001	0.004
TN	177.230	23.072	142.790	238.558	95.768	–	–	–	–	–
SOC	–	–	–	–	–	1.576	0.197	1.087	1.899	0.812
Global adjusted R^2	0.73					0.72				
Global RMSE	12.86					1.29				
Moran's I	0.06					0.02				

Table 6 Parameters of the fitted variogram models for SOC and TN stocks, and the residuals of the respective GWR and MLR models

Variable	Model	Nugget Mg ha^{-1}	Partial sill Mg ha^{-1}	Total sill Mg ha^{-1}	Range (m)	Nugget- to-sill ratio (%)	Spatial dependence
SOC stocks	Gaussian	386	278	664	4845	58.1	Moderate
MLR$_{soc}$ residuals	Gaussian	143	39	182	4434	78.6	Weak
GWR$_{soc}$ residuals	Gaussian	141	26	167	5760	84.4	Weak
TN stocks	Exponential	3.1	3.7	6.8	4493	45.6	Moderate
MLR$_{tn}$ residuals	Exponential	1.4	0.5	1.9	4132	73.7	Weak
GWR$_{tn}$ residuals	Exponential	1.4	0.2	1.6	2944	87.5	Weak

3.3 Model evaluation

Table 7 presents the results of ten-fold validation procedure used to quantify the errors attached to the prediction maps of SOC and TN stocks. The average MEs for all prediction

models are close to 0, which indicate a small tendency for over- or underestimation. In addition, the average RMSEs range from 16.7 to 19.9 Mg C ha^{-1} and from 1.5 to 1.9 Mg N ha^{-1}, which are slightly higher than the RMSEs of the fitted models (13.07 Mg C ha^{-1} and 1.33 Mg N ha^{-1} for MLR models, and 12.86 Mg C ha^{-1} and 1.29 Mg N ha^{-1} for GWR models). This suggests that the models do not predict new data as precise as they fit the original ones. However, the differences in RMSEs are slight, and it should also be noted that the proportions of observations used for calibration and validation were not equal. The GWR models show better performance in predicting SOC and TN stocks at new locations than MLR models given their lower average RMSEs and MEs. The average RMSEs are also slightly lower than the standard deviations of the measured values (Table 3), which means that incorporation of the predictors and spatial correlation gives better estimations than what would be achieved by just using the measured values for predictions. However, addition of the stochastic part (kriged residuals) to the GWR and MLR outputs does not reduce the prediction errors. The RMSEs of the GWRK and GWR models are similar, and so are the RMSEs of the MLRK and MLR models.

Table 7   Summary statistics of the spatial prediction errors

	SOC$_{st}$ (Mg C ha^{-1})		TN$_{st}$ (Mg N ha^{-1})	
	ME	RMSE	ME	RMSE
GWRK	−0.48	16.74	0.04	1.53
GWR	−0.86	16.66	−0.03	1.51
MLRK	0.39	19.42	−0.31	1.93
MLR	0.30	19.89	−0.33	1.93

ME= mean error; RMSE=root mean squared error

3.4   Spatial distribution and estimates of SOC and TN stocks

Figures 4 and 5 display the different prediction surfaces of SOC and TN stocks produced by MLR, MLRK, GWR, and GWRK models. The maps reveal similar spatial patterns of SOC and TN stocks meaning that SOC and TN stocks respond similarly to the environmental factors. There is a general decrease of SOC and TN stocks from west to east. The highest estimates of SOC and TN stocks occur in the western and northwestern parts, which according to the environmental data, have higher forest cover, elevations, and SOC and TN concentrations, but lower silt contents and surface temperatures. These *hotspots* are parts of the Logoman, Nessuiet, Kiptunga, and Baraget forests that are undisturbed. The lowest estimates, on the other hand, occur on the eastern side where croplands have been established, including Teret, Nessuiet, Kapkembu, Tuiyotich, and Sururu locations. These *coldspots* are areas with higher crop cover, silt contents, and surface temperatures, but lower elevations, and SOC and TN concentrations. In the northern and southeastern parts where crop cover is also high, the SOC and TN stocks are moderate to high. The GWR and GWRK prediction surfaces give more realistic pictures of the moderate to high SOC and TN stocks at Sururu forest in the southeasternmost part, which is more degraded than the forests in the western and northwestern parts.

The models generated minimum and maximum values that approximate the measured values (cf. Table 3). The MLR and MLRK estimates of TN stocks range from 5.8 to 15.1 Mg

N ha^{-1}, whereas the GWR and GWRK estimates vary from 5.3 to 15.8 Mg N ha^{-1}. Similarly, the MLR and MLRK estimates of SOC stocks range from 56.7 to 146 Mg C ha^{-1}, while the GWR and GWRK estimates vary from 55.6 to 146 Mg C ha^{-1}.

Table 8 gives the magnitude of SOC and TN stocks under different land cover categories

Table 8 Soil organic carbon and nitrogen stocks under different land cover types

Land cover	Area	SOC stocks				TN stocks			
		Min.	Max.	Mean	Total	Min.	Max.	Mean	Total
	(Ha)	(Mg ha^{-1})			(Tg)	(Mg ha^{-1})			(Tg)
Forests	32228.4	75.5	142.9	110.4	3.78	7.5	15.3	11.1	0.38
Grasslands	5509.4	66.7	129.8	103.5	0.57	6.7	12.6	10.4	0.06
Croplands	25828.1	62.9	126.9	95.2	2.46	6.5	12.2	9.6	0.25
Total	65565.9				6.81				0.69
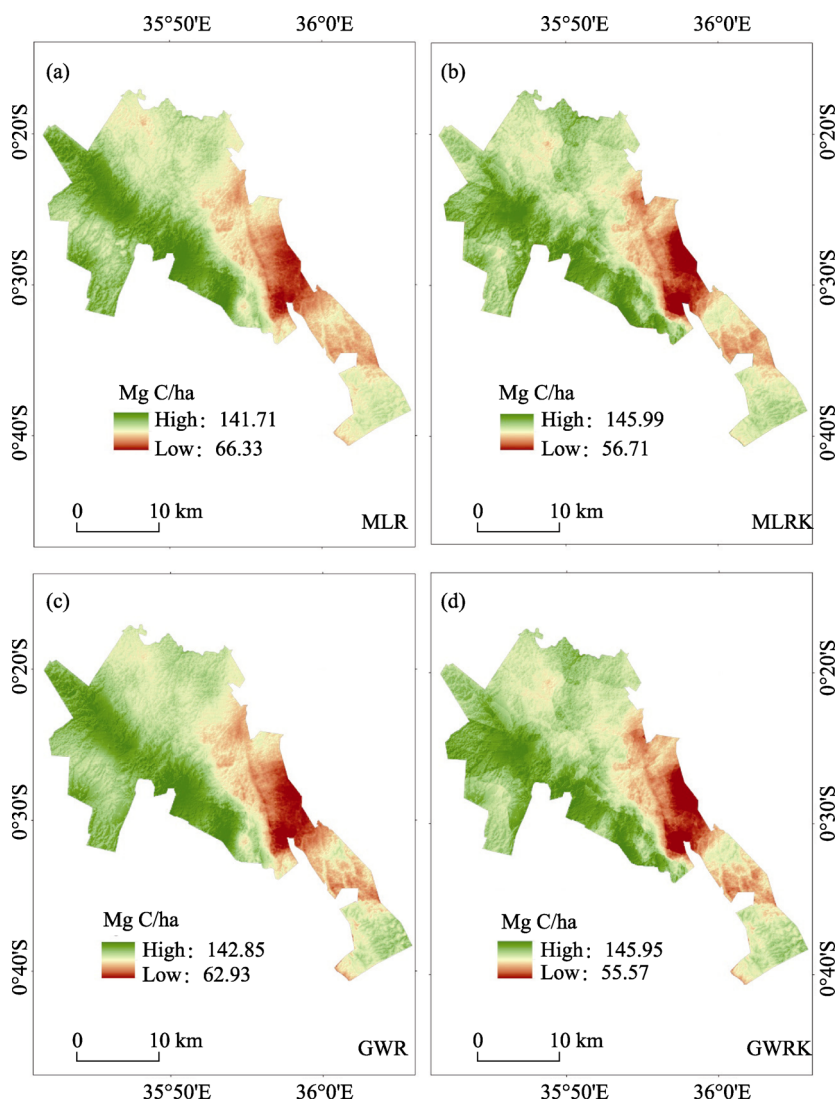
Figure 4 Maps showing the spatial patterns of the predicted SOC stocks using MLR, MLRK, GWR and GWRK

based on GWR method, which has lower prediction error indices compared to other methods. Forests stores the highest amounts of SOC and TN (3.78 Tg C and 0.38 Tg N) followed by croplands (2.46 Tg C and 0.25 Tg N), and grasslands (0.57 Tg C and 0.06 Tg N) (1 Tg = 10^{12} g = 1 million tons). This is because forests cover the largest area (32,228 ha), while grasslands cover the smallest area (5509 ha). In total, the Eastern Mau Forest Reserve stores about 6.81 Tg and 0.69 Tg of SOC and TN, respectively.
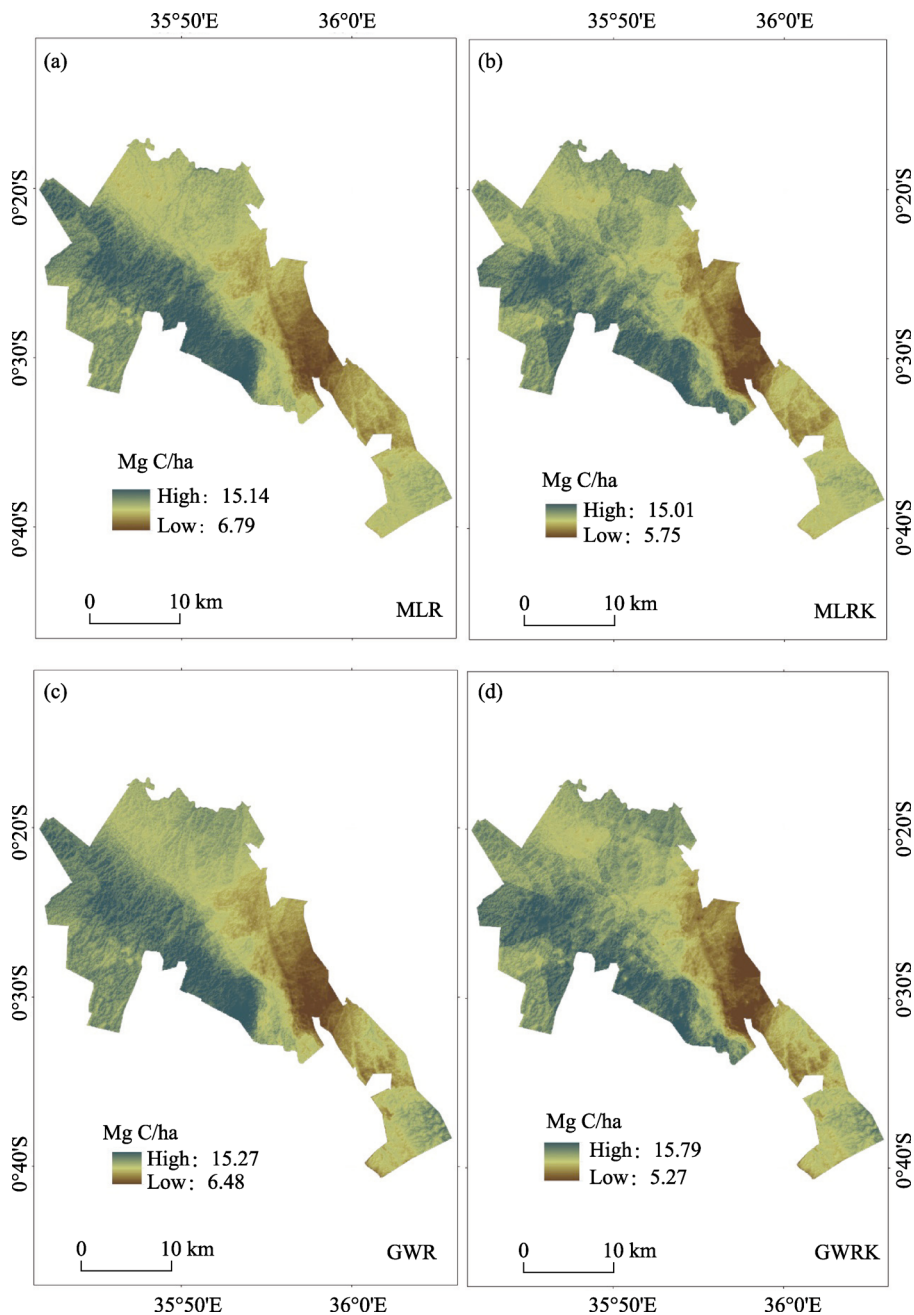

Figure 5   Maps showing the spatial patterns of the predicted TN stocks using MLR, MLRK, GWR and GWRK

4 Discussion

4.1 Spatial models

The significant effects of elevation, silt content, TN concentration, and OLI band 11 in the SOC models, and elevation, SOC concentration, and OLI band 11 in the TN models implies that topographical, edaphic, and climatic factors control the spatial patterns of SOC and TN stocks in the Eastern Mau Forest Reserve. Pachomphon *et al*. (2010) and Li and Shao (2014) reported similar combination of controlling factors in Laos and north-western China, respectively. The magnitudes of the effects of these predictors indicates that soil properties, particularly TN and SOC concentrations, are more important than the other factors in determining the observed variability of SOC and TN stocks, respectively. This was expected because of the high statistical correlation between them and the tight coupling of C and N cycles. Nitrogen supply increases the net uptake of C in terrestrial ecosystems, which in turn leads to higher inputs of C and N to the soils (Zaehle *et al*., 2011). The high coefficient of determination (i.e., R^2>0.70) obtained for the fitted MLR and GWR models further confirms the explanatory power of these soil properties. In a similar study at four contrasting East African landscapes, Vågen and Winowiecki (2013a) also concluded that intrinsic soil properties determined more the SOC dynamics than other environmental factors alone. The significant effects of OLI band 11 (proxy for surface temperature) and elevation on SOC and TN stocks suggests that: (i) some solutions to the problem of up-scaling soil survey data to landscape level in the region exist in the freely available remotely sensed and topographical data, and (ii) computationally intensive remote sensing- and DEM-derived parameters (e.g., NDVI, CTI) do not always improve the spatial prediction of soil properties. Generally, poor prediction performance (R^2<0.50) has been the norm in the region. For instance, Mora-Vallejo *et al*. (2008) developed MLR and MLRK models using topographical and geomorphological variables that explained less than 25% of SOC variability in south-eastern Kenya. But recently, Vågen *et al*. (2013b) and Winowiecki *et al*. (2015) achieved better performances (i.e., R^2>0.70) when they predicted SOC stocks using remotely-sensed imagery and random forest models in Ethiopia and Tanzania, respectively.

In terms of spatial structure, the NSRs of raw SOC and TN stocks data revealed moderate spatial dependency (Table 6), which compare with the findings of Sumfleth and Duttmann (2008). This suggests that in the short-range, random and structural processes are equally influential in explaining the spatial variability of SOC and TN stocks. The structural processes that determine the variability of SOC and TN stocks in the Eastern Mau Forest Reserve are the natural soil-forming factors, including topography, soil properties, and climate, while the random processes that explain the remaining variability are human activities, such as illegal loggings, encroachments, and charcoal burning, as well as land management practices. In contrast, the residuals obtained from the GWR and MLR models exhibited weaker patterning as evidenced by the low Moran's indices and high NSRs. This means that the global trend models partly explained the variability and spatial correlation of SOC and TN stocks leaving only a small, less structured, short-range variation unexplained (Vasques *et al*., 2010a, 2010b). The unexplained short-range spatial variation reflects the inherent data errors and spatial sources of variations at distances smaller than the shortest sampling interval. Theoretically, this can be resolved by increasing the sampling intensity, but practically, this

may be difficult to implement due to resource constraints. The NSRs also hint at the proportion of variation that can be explained by the spatial models. As expected, the NSRs for the MLR models of SOC and TN stocks (79% and 74%) were close to the proportions of variation that the models explained (adjusted R^2=72% and 71%).

4.2   Model evaluation

The GWR-based models were better than MLR-based models in predicting new data (Table 7); thus, GWR models were chosen to quantify the total stocks of SOC and TN in the area and under different land cover types (Table 8). Mishra *et al*. (2010), Zhang *et al*. (2011), and Syzomanoski and Kryza (2012) also obtained similar results. Basically, the MLR approach assumed that the environmental factors, which affected the variability of SOC and TN stocks, were spatially stationary. Hence, it represented their relationships using a global statistic. However, such global values can lead to large errors and be misleading since most of the variability in SOC and TN stocks stem from local interaction of processes (Kumar *et al*., 2012). In contrast, the GWR approach applied regressions locally, which accounted for both the spatial trends and local variations resulting in superior estimations of SOC and TN stocks. The major weakness of GWR approach was that even though the variation of regression coefficients locally implied the selection of different predictors at different locations (Zhang *et al*., 2011; Kumar *et al*., 2013), this did not happen; hence, some predictors may have been redundant at some locations. Unlike other studies (Mishra *et al*., 2010; Kumar *et al*., 2012; Syzomanoski and Kryza, 2012; Zhang *et al*., 2012), the addition of stochastic component (kriged residuals) to the MLR and GWR outputs did not yield lower prediction errors in this case. The small proportion of spatially correlated random component in the residuals as indicated by the low Moran's indices and NSRs (Table 6) explains this. Mora-Vallejo *et al*. (2008) and Li *et al*. (2013b) also reported that MLRK did not outperform MLR in their study.

4.3   Spatial distribution and estimates of SOC and TN stocks

The prediction maps revealed spatial patterns of SOC and TN stocks that were similar and reflected the environmental predictors. The given characteristics of the *hotspots* of SOC and TN stocks in the western and northwestern parts, as well as the highly fertile *Andosols* of the area favour accumulation of SOC and TN stocks. For instance, the high rainfall and low temperatures associated with higher altitudes increase net primary productivity of the forests and decrease SOC turnover. The lower silt content relative to clay content in the forest soils also indicates the presence of organo-complexes, or allophane, imogolite, and ferrihydrite clay minerals, which stabilize organic matter and plant nutrients (Lemenih *et al*., 2005; Chaplot *et al*., 2010). The smaller pore spaces of clay particles also promote aggregation and physical protection of SOC. In contrast, the characteristics of the *coldspots* of SOC and TN stocks on the eastern side are unfavourable for accumulation of SOC and TN stocks. For example, the higher crop cover is attributed to the conversion of forests to croplands, which began in the mid-1990s. In these croplands, biomass removal after harvesting, erosion, and frequent tillage, which breaks up the soil aggregates and alters aeration, can explain the lower SOC and TN stocks (Murty *et al*., 2002; Smith 2008; Eclesia *et al*., 2012; Wiesmeier *et al*., 2012). Thus, the *coldspots* of SOC and TN stocks also highlight human-induced soil degradation and sources of C and N emissions. The altitudinal gradient in SOC stocks men-

tioned above corresponds with previous studies in the Tana River basin, Kenya (Tamooh *et al*., 2012), while the highest SOC and TN stocks under forests coincide with other studies in the region (Bewketa and Stroosnidjer, 2003; Lemenih *et al*., 2005; Girmay and Singh, 2012; Demessie *et al*., 2013). The *hotspots* and *coldspots* on the prediction maps of SOC and TN stocks are the sites to target the best management practices (BMPs) for climate change mitigation and sustainable land management. For example, the western and northwestern parts need practices that promote retention, whereas the eastern part requires those that enhance accumulation of SOC and TN stocks.

The SOC and TN stocks in the Eastern Mau Forest Reserve to 30 cm depth were estimated at 6.81 Tg and 0.69 Tg, respectively. This accounts for 0.36% of the total SOC stock to 30 cm depth reported for Kenya (Batjes, 2004). Batjes (2004) further reported that *Andosols* of the humid and semi-humid regions in Kenya stored an average of 9.1 Kg C m^{-2} (91 Mg C ha^{-1}) to 30 cm depth, which slightly differs from the present findings (i.e., 10.3 Kg C m^{-2} or 102.7 Mg C ha^{-1}). This difference can be ascribed to the properties of the data used in the two studies. Batjes (2004) used coarse resolution legacy data from SOil and TERrain (SOTER) database and Africa Land Cover Characteristics (ALCC) database, while the present study used newly collected, fine resolution field data to estimate SOC stocks.

4.4   Limitations of the study

We acknowledge the limitations of this study. Firstly, the soil properties used as predictors were themselves products of interpolation by ordinary kriging. Thus, interpolation errors may have been propagated to the subsequent prediction of SOC and TN stocks. These predictors would have enhanced the prediction accuracy more had they been sampled more intensely than the target variables. Similarly, estimation of SOC and TN stocks under different land cover types was based on a land cover map that had been produced through classification of Landsat 5 TM satellite imagery. Thus, the inherent classification errors may have influenced the estimates of SOC and TN stocks under the different land cover classes. Additionally, the auxiliary spatial data (e.g., DEM, Landsat imagery, and climate) were sourced from various databases; hence, their quality was different. Poor coverage of samples in the southeastern most and central parts, which were dominated by thick impenetrable bamboo forests may have also affected prediction accuracy in these areas. Lastly, some soil-forming factors (e.g., parent material and age) were omitted owing to lack of suitable data. Their inclusion, if significant, may improve the predictive power of future models. The foregoing factors introduced uncertainties, the quantification of which was beyond the scope of this study. Future work will assess the implications of error propagation through sensitivity analysis of model parameters estimated using multi-source auxiliary spatial data with varied accuracy.

5   Conclusions and recommendations

This study has demonstrated an integrated approach of field sampling, GIS, remote sensing, and statistical analysis to quantify and map SOC and TN stocks to 30 cm depth in the Eastern Mau Forest Reserve, Kenya. Based on the results, the conclusions drawn are: (1) Forests have the largest SOC and TN pools followed by croplands and grasslands. Altogether, the

Eastern Mau Forest Reserve stores about 6.81 Tg of C and 0.69 Tg of N. (2) The *hotspots* of SOC and TN stocks are the native systems in the western and northwestern parts, namely Logoman, Nessuiet, Kiptunga, and Baraget forests, while the *coldspots* are the human-dominated landscapes in the eastern part, including Teret, Nessuiet, Kapkembu, Tuiyotich, and Sururu locations. Thus, conversion of forests to croplands is a driver of soil degradation in this area. (3) Climatic, edaphic, and topographic factors control the observed spatial patterns of SOC and TN stocks; however, soil properties, particularly TN and SOC concentrations are the most important determinants. Despite the limitations, this study provides the first spatially exhaustive soil information for Eastern Mau forest reserve at a finer scale. The resultant outputs will assist to monitor SOC and TN stocks, as well as to formulate spatially targeted climate change mitigation and sustainable land management policies. Also, the approach used offers a cost-effective framework to derive knowledge of soil processes and multi-purpose soil information in other data-poor environments in Eastern Africa.

Acknowledgements

References

Amare T, Hergarten C, Hurni H *et al*., 2013. Prediction of soil organic carbon for Ethiopian highlands using soil spectroscopy. ISRN Soil Science, 720589 (11 pp), http://dx.doi.org/10.1155/2013/720589.

Aynekulu E, Vågen T-G, Shepherd K *et al*., 2011. A protocol for measurement and monitoring soil carbon stocks in agricultural landscapes. Version 1.1. World Agroforestry Centre, Nairobi.

Batjes N H, 2004. Soil carbon stocks and projected changes according to land use and management: A case study for Kenya. *Soil Use and Management*, 20: 350–356.

Bewketa W, Stroosnijder L, 2003. Effects of agro-ecological land use succession on soil properties in Chemoga watershed, Blue Nile basin, Ethiopia. *Geoderma*, 111: 85–98.

Blake G R, 1965. Bulk density. In: Black C A (ed.). Methods of Soil Analysis, Part 1. Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling, American Society of Agronomy, Inc., Madison, Wisconsin, USA.

Bremner J M, Mulvaney C S, 1982. Nitrogen - total. In: Page A L (ed.). Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties. 2nd ed. American Society of Agronomy, Inc., Madison, Wisconsin, USA.

Cambule A H, Rossiter D G, Stoorvogel J J *et al*., 2014. Soil organic carbon stocks in the Limpopo National Park, Mozambique: Amount, spatial distribution and uncertainty. *Geoderma*, 213: 46–56.

Chaplot V, Bouahom B, Valentin C, 2010. Soil organic carbon stocks in Laos: Spatial variations and controlling factors. *Global Change Biology*, 16: 1380–1393.

Day P R, 1965. Particle fractionation and particle size analysis. In: Black C A (ed.). Methods of Soil Analysis, Part 1. Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling. American Society of Agronomy, Inc., Madison, Wisconsin, USA.

Demessie A, Singh B R, Lal R, 2013. Soil carbon and nitrogen stocks under chronosequence of farm and traditional agro-forestry uses in Gambo district, southern Ethiopia. *Nutr. Cycl. Agroecosys.*, 95: 365–375.

Doetterl S, Stevens A, van Oost K *et al*., 2013. Spatially explicit regional scale prediction of soil organic carbon stocks in cropland using environmental variables and mixed model approaches. *Geoderma*, 204/205: 31–42.

Dorji T, Odeh I O A, Field D J *et al*., 2014. Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. *Forest Ecology and Management*, 318: 91–102.

Eclesia R P, Jobbagy E G, Jackson R B *et al*., 2012. Shifts in soil organic carbon for plantation and pasture establishment in native forests and grasslands of South America. *Global Change Biology*, 18: 3237–3251.

Elbasiouny H, Abowaly M, Abu_Alkheir A *et al*., 2014. Spatial variation of soil carbon and nitrogen pools by using ordinary kriging method in an area of north Nile delta, Egypt. *Catena*, 113: 70–78.

Fotheringham A S, Brunsdon C, Charlton M E, 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. England: John Wiley & Sons Inc.

Girmay G, Singh B R, 2012. Changes in soil organic carbon stocks and soil quality: Land use system effects in northern Ethiopia. Acta Agriculturae Scandinavica, Section B – *Soil & Plant Science*, 62(6): 519–530.

Government of Kenya, 2009. Report of the prime minister's task force on the conservation of the Mau forest complex. [Online]. Available:

http://www.kws.org/export/sites/kws/info/maurestoration/maupublications/Mau_Forest_Complex_Report.pdf [Accessed 2014, January 19].

Grimm R, Behrens T, Märker M *et al*., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island: Digital soil mapping using Random Forests analysis. *Geoderma*, 146: 102–113.

Hengl T, Heuvelink G B M, Rossiter D G, 2007. About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33: 1301–1315.

Hengl T, Heuvelink G B M, Stein A, 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120: 75–93.

Hiemstra P, 2013. Classes and methods for spatial data in R. [Online]. Available: http://cran.r-project.org/web/packages/automap/automap.pdf. [Accessed 2013, December 15].

IPCC, 2006. IPCC Guidelines for national greenhouse gas inventories, prepared by the national greenhouse gas inventories programme, Eggleston H S, Buendia L, Miwa K *et al*. (eds.). Published: IGES, Japan.

Jaber S M, Al-Qinna M I, 2011. Soil organic carbon modelling and mapping in a semi-arid environment using thematic mapper data. *Photogrammetric Engineering & Remote Sensing*, 77(7): 709–719.

Jaetzold R, Schmidt H, Hornetz B *et al*., 2010. Farm management handbook of Kenya, Vol. II. Natural conditions and farm management information. 2nd ed., Part B Central Kenya, Subpart B1a Southern Rift Valley Province. Ministry of Agriculture, Kenya and German Agency for Technical Cooperation (GTZ), Nairobi.

Karunaratne S B, Bishop T F A, Baldock J A *et al*., 2014. Catchment scale mapping of measureable soil organic carbon fractions. *Geoderma*, 219/220: 14–23.

Kheir R B, Greve M H, BØcher P K *et al*., 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification tree-based models: The case study of Denmark. *Journal of Environmental Management*, 91: 1150–1160.

Kumar S, Lal R, 2011. Mapping the organic carbon stocks of surface soils using local spatial interpolator. *Journal of Environmental Monitoring*, 13: 3128–3135.

Kumar S, Lal R, Liu D, 2012. A geographically weighted regression kriging approach for mapping soil organic carbon stock. *Geoderma*, 189/190: 627–634.

Kumar S, Lal R, Liu D, 2013. Estimating the spatial distribution of organic carbon density for the soils of Ohio, USA. *Journal of Geographical Sciences*, 23(2): 280–296.

Lacoste M, Minasny B, McBratney A *et al*., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma*, 213: 296–311.

Lal R, 2004. Soil carbon sequestration to mitigate climate change. *Geoderma*, 123: 1–22.

Lamsal S, Grunwald S, Bruland G L *et al*., 2006. Regional hybrid geospatial modeling of soil nitrate–nitrogen in the Santa Fe River watershed. *Geoderma*, 135: 233–247.

Lemenih M, Karltun E, Olsson M, 2005. Assessing soil chemical and physical property responses to deforestation and subsequent cultivation in smallholders farming system in Ethiopia. *Agriculture, Ecosystems and Environment*, 105: 373–386.

Lesch S M, Corwin D L, 2008. Prediction of spatial soil property information from ancillary sensor data using ordinary linear regression: Model derivations, residual assumptions and model validation tests. *Geoderma*, 148: 130–140.

Li D, Shao M, 2014. Soil organic carbon and influencing factors in different landscapes in an arid region of north-western China. *Catena*, 116: 95–104.

Li M, Zhang X, Pang G *et al*., 2013b. The estimation of soil organic carbon distribution and storage in a small catchment area of the Loess Plateau. *Catena*, 101: 11–16.

Li Q, Yue T, Wang C *et al*., 2013a. Spatially distributed modeling of soil organic matter across China: An application of artificial neural network approach. *Catena*, 104: 210–218.

Li Y, 2010. Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? *Geoderma*, 159: 63–75.

Liu Z, Shao M, Wang Y, 2011. Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. *Agriculture, Ecosystems and Environment*, 142: 184–194.

Malone B P, McBratney A B, Minasny B *et al*., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, 154: 138–152.

Marchetti A, Piccini C, Francaviglia R *et al*., 2012. Spatial distribution of soil organic matter using geostatistics: A key indicator to assess soil degradation status in central Italy. *Pedosphere*, 22(2): 230–242.

Martin M P, Orton T G *et al*., 2014. Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma, http://dx.doi.org/10.1016/j.geoderma.2014.01.005.

Martin M P, Wattenbach M, Smith P *et al*., 2011. Spatial distribution of soil organic carbon stocks in France. *Biogeosciences*, 8: 1053–1065.

McBratney A B, Santos M L M, Minasny B, 2003. On digital soil mapping. *Geoderma*, 117: 3–52.

McCall G J H, 1967. Geology of the Nakuru-Thomson's falls-Lake Hannington area: Degree sheet No. 35, S.W. Quarter and 43 N.W. Quarter, Report No. 78. Government Printer, Nairobi.

McKenzie N J, Ryan P J, 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89: 67–94.

Meersmans J, de Ridder F, Canters F *et al*., 2008. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, 143: 1–13.

Mehrjardi R T, Minasny B, Sarmadian F *et al*., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*, 213: 115–128.

Mishra U, Lal R, Liu D *et al*., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. *Soil Science Society of America Journal*, 74: 906–914.

Mishra U, Riley W J, 2012. Alaskan soil carbon stocks: Spatial variability and dependence on environmental factors. *Biogeosciences*, 9: 3637–3645.

Montgomery D C, Peck E A, Vining G G, 2006. Introduction to Linear Regression Analysis. John Wiley & Sons, Inc., New Jersey.

Mora-Vallejo A, Claessens L, Stoorvogel J *et al*., 2008. Small-scale digital soil mapping in southeastern Kenya. *Catena*, 76: 44–53.

Murty D, Kirschbaum M F, McMurtrie R E *et al*., 2002. Does conversion of forest to agricultural land change soil carbon and nitrogen? A review of the literature. *Global Change Biology*, 8: 105–123.

Nelson D W, Sommers L E, 1982. Total carbon, organic carbon and organic matter. In: Page A L (ed.) Methods of Soil Analysis, Part 2, Chemical and Microbiological Properties. 2nd ed. American society of agronomy, Inc., Madison, Wisconsin, USA.

Obade V P, Lal R, 2013. Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). *Catena*, 104: 77–92.

Okalebo J R, Gathna K W, Woomer P L, 2002. Laboratory methods for soil and plant analysis: A working manual. 2nd ed. Tropical Soil Biology and Fertility Programme, Nairobi.

Overmars K P, Verburg P H, 2005. Analysis of land use drivers at the watershed and household level: Linking two paradigms at the Philippine forest fringe. *International Journal of Geographical Information Science*, 19(2): 125–152.

Pachomphon K, Dlamini P, Chaplot V, 2010. Estimating carbon stocks at regional level using soil information and easily accessible auxiliary variables. *Geoderma*, 155: 372–380.

Pebesma E, Bivand R S, Rowlingson B *et al*., 2013. Classes and methods for spatial data in R. Available: http://cran.r-project.org/web/packages/sp/sp.pdf. [Accessed 2013, December 15].

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://wwwR-projectorg/.

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

Scull P, Franklin J, Chadwick O A *et al*., 2003. Predictive soil mapping: A review. *Progress in Physical Geography*, 27(2): 171–197.

Selige T, Böhner J, Schmidhalter U, 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, 136: 235–244.

Smith P, 2004. Soils as carbon sinks: The global context. *Soil Use and Management*, 20: 212–218.

Smith P, 2008. Land use change and soil organic carbon dynamics. *Nutr. Cycl. Agroecosyst.*, 81: 169–178.

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

Szymanowski M, Kryza M, 2012. Local regression models for spatial interpolation of urban heat island: An example from Wrocław, SW Poland. *Theor. Appl. Climatol.*, 108: 53–71.

Tamooh F, van den Meersche K, Meysman F *et al.*, 2012. Distribution and origin of suspended matter and organic carbon pools in the Tana River basin, Kenya. *Biogeosciences*, 9: 2905–2920.

Tesfahunegn G B, Tamene L, Vlek P L G, 2011. Catchment scale spatial variability of soil properties and implications on site-specific soil management in northern Ethiopia. *Soil & Tillage Research*, 117: 124–139.

UNEP, 2009. Kenya: Atlas of Our Changing Environment. Division of Early Warning and Assessment (DEWA), United Nations Environment Programme (UNEP). [Online]. Available: http://www.unep.org/dewa/africa/kenyaatlas/. [Accessed 2013, August 28].

Vågen T G, Winowiecki L A, 2013a. Mapping of soil organic carbon stocks for spatially explicit assessments of climate change mitigation potential. *Environmental Research Letters* 8, 015011 (9pp). doi: 10.1088/1748-9326/8/1/015011.

Vågen T G, Winowiecki L A, Abegaz A *et al.*, 2013b. Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia. *Remote Sensing of Environment*, 134: 266–275.

Vasques G M, Grunwald S, Comerford N B *et al.*, 2010a. Regional modelling of soil carbon at multiple depths within a subtropical watershed. *Geoderma*, 156: 326–336.

Vasques G M, Grunwald S, Sickman J O *et al.*, 2010b. Up-scaling of dynamic soil organic carbon pools in a north-central Florida watershed. *Soil Science Society of America Journal*, 74. doi: 10.2136/sssaj2009.0242.

Wang K, Zhang C, Li W, 2013. Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and co-kriging. *Applied Geography*, 42: 73–85.

Were K O, Dick Ø B, Singh B R, 2013. Remotely sensing the spatial and temporal land cover changes in Eastern Mau forest reserve and Lake Nakuru drainage basin, Kenya. *Applied Geography*, 41: 75–86.

Were K O, Singh B R, Dick Ø B, 2015. Effects of land cover changes on soil organic carbon and nitrogen stocks in the Eastern Mau Forest Reserve, Kenya. In: Lal R, Singh B R, Mwaseba D L *et al.*, (eds.). Sustainable Intensification to Advance Food Security and Enhance Climate Resilience in Africa. Springer International Publishing, Switzerland, 113–133.

Wiesmeier M, Spörlein P, Geuß U *et al.*, 2012. Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling depth. *Global Change Biology*, 18: 2233–2245.

Winowiecki L, Vågen T G, Huising J, 2015. Effects of land cover on ecosystem services in Tanzania: A spatial assessment of soil organic carbon. *Geoderma*, http://dx.doi.org/10.1016/j.geoderma.2015.03.010.

Wu C, Wu J, Luo Y *et al.*, 2009. Spatial prediction of soil organic matter content using co-kriging with remotely sensed data. *Soil Science Society of America Journal*, 73: 1202–1208.

Yang R, Su Y Z, Wang M *et al.*, 2014. Spatial pattern of soil organic carbon in desert grasslands of the diluvial-alluvial plains of northern Qilian Mountains. *Journal of Arid Land*, 6(2): 136–144.

Yang Y, Fang J, Tang Y *et al.*, 2008. Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. *Global Change Biology*, 14: 1592–1599.

Zaehle S, Ciais P, Friend A D *et al.*, 2011. Carbon benefits of anthropogenic reactive nitrogen offset by nitrous oxide emissions. *Nature Geoscience*, 4: 601–605.

Zhang C, Tang Y, Xu X *et al.*, 2011. Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. *Applied Geochemistry*, 26: 1239–1248.

Zhang S, Huang Y, Shen C *et al.*, 2012. Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma*, 171/172: 35–43.

Zhang Z, Yu C, Shi X *et al.*, 2010. Application of categorical information in the spatial prediction of soil organic carbon in the red soil area of China. *Soil Science and Plant Nutrition*, 56: 307–318.