Research Article

Soil-Landscape Modeling and Remote Sensing to Provide Spatial Representation of Soil Attributes for an Ethiopian Watershed

Nurhussen Mehammednur Seid,[1] Birru Yitaferu,[2] Kibebew Kibret,[3] and Feras Ziadat[4]

[1] *Burie Agricultural College, P.O. Box Rissneleden 19, Sundbyberg, 17453 Stockholm, Sweden*
[2] *Amhara Regional Agricultural Research Institute (ARARI), P.O. Box 527, Bahir Dar, Ethiopia*
[3] *School of Natural Resources Management and Environmental Science, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia*
[4] *International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box 950764, Amman 11195, Jordan*

Correspondence should be addressed to Feras Ziadat; f.ziadat@cgiar.org

Information about the spatial distribution of soil properties is necessary for natural resources modeling; however, the cost of soil surveys limits the development of high-resolution soil maps. The objective of this study was to provide an approach for predicting soil attributes. Topographic attributes and the normalized difference vegetation index (NDVI) were used to provide information about the spatial distribution of soil properties using clustering and statistical techniques for the $56 \, km^2$ Gumara-Maksegnit watershed in Ethiopia. Multiple linear regression models implemented within classified subwatersheds explained 6–85% of the variations in soil depth, texture, organic matter, bulk density, pH, total nitrogen, available phosphorous, and stone content. The prediction model was favorably comparable with the interpolation using the inverse distance weighted algorithm. The use of satellite images improved the prediction. The soil depth prediction accuracy dropped gradually from 98% when 180 field observations were used to 65% using only 25 field observations. Soil attributes were predicted with acceptable accuracy even with a low density of observations (1-2 observations/$2 \, km^2$). This is because the model utilizes topographic and satellite data to support the statistical prediction of soil properties between two observations. Hence, the use of DEM and remote sensing with minimum field data provides an alternative source of spatially continuous soil attributes.

1. Introduction

Quantitative information and spatial distribution of soil properties are among the main prerequisites for achieving sustainable land management. The accuracy of soil information determines, to a large extent, the reliability of land resources management decisions [1–3]. Conventional soil surveys are usually used to derive information about soils and their distribution [4]; however, limited areas are covered by detailed soil information mainly due to the high costs of surveys [5]. Furthermore, the spatial distribution of soil characteristics as represented by a conventional soil map does not reflect the distribution in the field because of the polygon-based mapping employed [6]. In polygon-based mapping, soils are assumed to be homogeneous within the polygon, and abrupt changes take place at the boundaries between polygons [7, 8]. Soils usually show a diffuse spatial distribution

that is hard to address in polygon-based soil maps [9]. Many researchers have suggested continuous raster maps as a better alternative to mapping soils and their properties [6, 7, 9].

In soil science, the implementation of geomatics—GIS, GPS, remote sensing, and digital elevation models (DEMs)—is suggesting new alternatives [3, 10]. Global coverage of high-resolution imagery and terrain data is increasing, and this enhances the popularity of digital soil mapping to generate accurate soil maps that capture many properties with reasonable effort [11]. The benefit of using GIS is the production of digital soil attributes and digital soil maps to optimize the need for costly field work and laboratory analysis [10, 12].

Various statistical models are used to investigate the relationships between the spatial distribution of soil attributes and that of landscape attributes. Predictive mapping techniques, such as geostatistics (i.e., kriging and cokriging),

fuzzy logic, linear and multiple regression, regression trees, and neural networks, have been used to develop soil maps [10, 13]. Ongoing research in digital soil mapping indicates that quantitative prediction models are promising tools to produce soil maps with acceptable accuracy [14, 15]. Research has provided optimistic results, and some researchers obtained better results than traditional soil surveys [16–18]. The use of satellite data to complement topographic information improves the mapping of natural resources [10, 13, 19]. These investigations are based on the relationship between landscape pattern and the differentiation of soil types and attributes [20]. The spatial distribution of specific soil properties has been predicted using soil-landscape modeling. This included sand, silt, and organic matter contents; A-horizon thickness; solum depth; extractable phosphorus; and pH [21, 22]. Using different terrain attributes, the empirical models developed in these studies explained 41–68% of the variation in soil properties [21, 22].

GIS is used to quantify the relationships between soils and topographic and remote sensing variables and to aid the digital soil mapping efforts [2, 3, 10, 23, 24]. Topographic parameters are derived from DEMs and used to predict the distribution of soil characteristics [5, 22, 25, 26]. The increasing availability of high-resolution remote sensing data provides a new opportunity for predicting soil characteristics with acceptable accuracy [11]. Many researchers have reported on the contribution of remote sensing data in providing acceptable prediction of soil characteristics [27–29]. Nevertheless, there are many issues that warrant further research: the conjunctive use of satellite data and topographic data to improve the prediction of soil characteristics; and the best time of year to acquire satellite data. The use of various numbers (densities) of observations, and how that affects accuracy, will help researchers decide on a reasonable number of observations to maintain acceptable accuracy with an appropriate amount of field work. The use of soil-landscape models to predict soil attributes is theoretically better than interpolation techniques, because the former uses topography and satellite data as a background to guide prediction, while interpolation relies solely on the spatial relationships between adjacent observations without considering the variations in factors that drive soil differentiation. However, the merits of using soil-landscape modeling compared with interpolation techniques need verification and quantification, especially when a limited number of observations are available. The objective of this study is to investigate the utility of a soil-landscape model using DEM and remote sensing to predict the spatial distribution of soil characteristics with an optimum number of field observations.

2. Materials and Methods

The study area is located in the Lake Tana basin in the Gumara-Maksegnit watershed of the Amhara region in Ethiopia, within $12°24'–12°31'$N and $37°34'–37°37'$E. It is centered 45 km southwest of Gondar town and covers an area of 56 km^2. Altitude within the watershed ranges from 1923 to 2851 m above sea level, with topography ranging from gentle
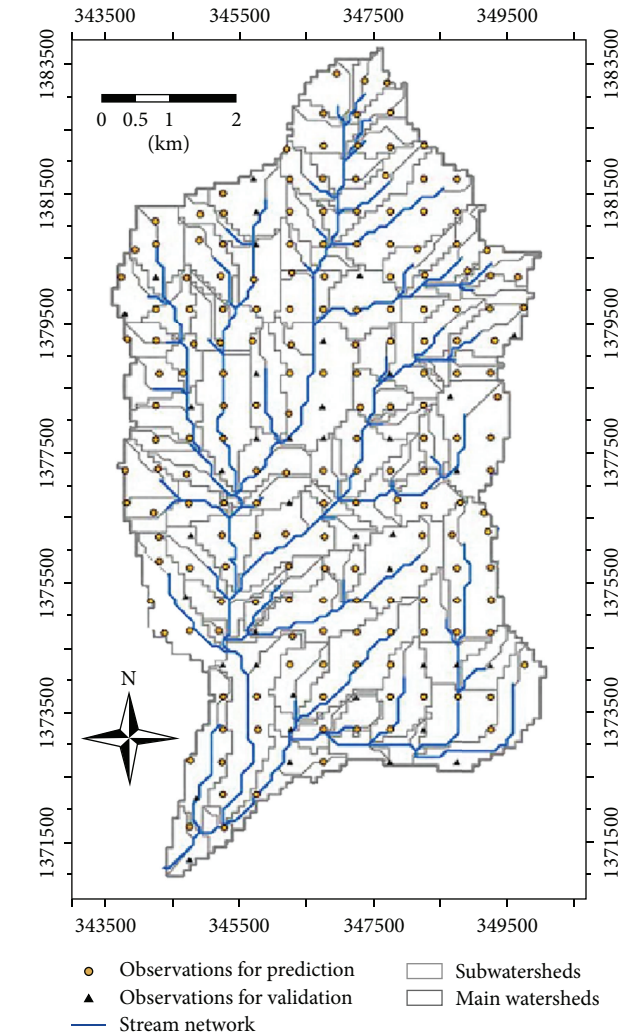

Figure 1: Distribution of field observations, watershed boundary, stream network, and subwatershed divisions for Gumara-Maksegnit watershed.

to sharp steep slopes. The mean annual rainfall is 1052 mm. The mean minimum and maximum temperatures are 13.3 and 28.5°C, respectively. The study area is characterized by an ustic moisture regime [30]. The natural resource base has been depleted to a great extent due to soil erosion and improper land use, which resulted in reduced tree cover, biodiversity, agriculture, range, and pasture productivity.

2.1. Soil Survey and Field Observations. The area was divided into square grids (500 m × 500 m). The total area of the watershed and the heterogeneity of topography and vegetation cover, and consequently the soils, were used to guide the selection of this grid size. Larger grid size (1 km) would result in a limited number of observations and would limit the representation of soil variability. Soil samples were taken within each grid in the watershed (Figure 1). This was a compromise between grid and free sampling techniques. While the surveyor was free to take the sample within the defined grid to represent the area, the distribution of sampling sites

still followed an unbiased grid. In cases where many grids shared common features, the surveyor reduced the number of samplings to represent these grids with one sampling site. Each site was excavated to a depth of 100 cm, or an impeding layer, using an auger. FAO terminology [30] was used to describe the sampling sites. The soil attributes and site characteristics recorded were GPS coordinates (easting, northing, and elevation), soil depth using an auger, slope steepness (percent) estimated using clinometers, surface stone cover (percent), and stone content of the top soil layer (0–25 cm). A soil sample for each soil observation was taken for laboratory analysis. The following soil attributes were analyzed: clay, silt, sand, organic matter, total nitrogen and available phosphorus contents, pH, and bulk density.

The hydrometer method outlined by the simplified procedure of Day was used to determine soil particle size distribution [31]. Hydrogen peroxide (H_2O_2) was used to destroy the organic matter, and sodium hexametaphosphate ($NaPO_3$) was used as a dispersing agent. The bulk density (Bd) of the soil was estimated from undisturbed (core) soil samples collected using a core sampler, weighed at field moisture content, and then determined following the procedures of Blake [32].

Soil pH was measured using a digital pH meter in the supernatant suspension of 1 : 2.5 soil : liquid ratio; where the liquids were water and 1 M KCl solution, and pH was calculated by subtracting soil pH (KCl) from soil pH (H_2O). The organic carbon content of the soil was analyzed following the wet digestion method described by Walkley and Black, which involves digestion of organic carbon in the soil samples with potassium dichromate ($K_2Cr_2O_7$) in sulfuric acid solution [33]. The Kjeldahl procedure was followed for the determination of total nitrogen as described by Sahlemeden and Taye by oxidizing the organic matter with concentrated sulfuric acid and converting the nitrogen in the organic compounds into ammonium sulfate during the oxidation [34].

Available phosphorus was determined by Olsen et al. method [35]. The soil samples were shaken with 0.5 M sodium bicarbonate at a nearly constant pH 8.5 in 1 : 2 soil : solvent ratio for 30 min, and the extract was obtained by filtering the suspension as indicated by Olsen et al.. The available phosphorus extracted was measured by spectrophotometer following the procedure outlined by Murphy and Riley [36].

2.2. DEM and Terrain Analysis. Shuttle Radar Topography Mission (SRTM) 90 m resolution DEM was used for the study. The free availability and accuracy of this product, checked against field observations, were the main reasons for its selection. Previous research recommended some terrain attributes as best predictors of soil characteristics [37, 38]. The following terrain parameters were derived using standard commands in ArcGIS: aspect; profile, plan, and mean curvatures; flow accumulation area and slope. The average upslope contributing area for each pixel (upslope flow accumulation) was calculated by multiplying the average flow accumulation grid by the area of the pixel (flow accumulation $\times$ 8100 m^2). The compound topographic index (CTI) for each pixel was calculated using the formula [21] CTI = ln (As/tan$D$), where As is the average upslope contributing area and D is the
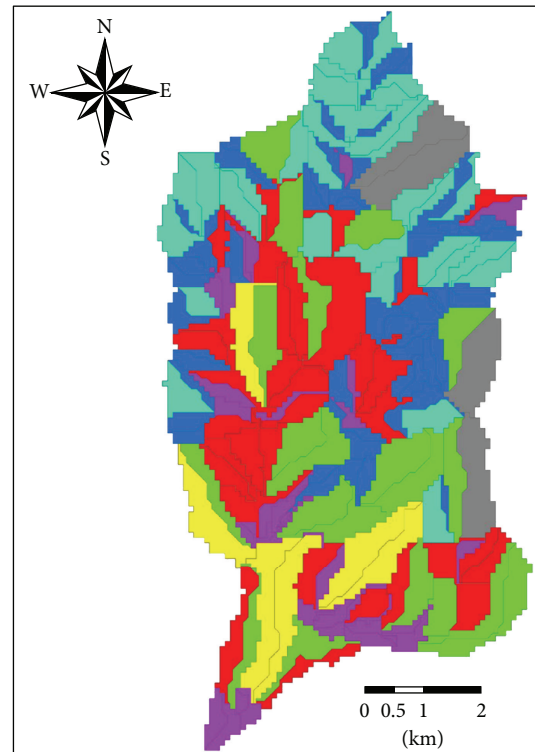
Class		Area (ha)	Slope (%)
■	1	3–25	20–60
■	2	45–70	0–40
■	3	3–25	0–20
■	4	70–100	0–20
■	5	80–90	30–50
■	6	20–130	0–20
■	7	18–45	20–60

FIGURE 2: Distribution and characteristics of the seven subwatersheds classes.

average slope degree. ARCSWAT was used to derive the stream network, and the watershed was also automatically divided into a number of subwatersheds (Figure 1).

The watershed subdivisions were considered as a base unit in the prediction model analyses. This is because the correlation between terrain attributes and soil characteristics for the whole watershed was very low and therefore unsuitable to predict the latter from the former. The subdivision of the whole watershed into smaller subwatersheds grouped the soil observations into homogeneous units and so enabled the establishment of better statistical relationships [37]. Each small subwatershed was divided, by the passing streamline, into two facets (subdivisions) (Figure 1). The generated facets were grouped into classes (Figure 2), based on their characteristics of area and slope. If individual subwatersheds were considered directly, the number of observations within each subwatershed was not enough to establish a rigorous statistical relationship. The grouping of subwatersheds into classes, based on characteristics such as area and slope, increased the number of observations within each class and enabled good prediction. The selection of an appropriate number

of classes to cluster the subwatershed facets is important in generating good results [37]. The classification (grouping) of subwatershed facets was repeated to produce a classification that led to an approximately equal number of observations in each class (at least 15 observations). This allowed a reasonable regression model to be established for each class. The clusters were used in the prediction of soil attributes. Seven classes were found to be optimum for predicting soil attributes in the study area.

2.3. Generating the Normalized Difference Vegetation Index (NDVI) Map.

ASTER satellite images taken on January 30 and March 19, 2007 and SPOT satellite images taken on October 8 and November 23, 2007 were analyzed in order to select the best image(s) for the study. The images were corrected radiometrically and geometrically using ENVI software. Radiometric correction was done separately for ASTER and SPOT imagery on an individual band basis. Geometric correction for the images within the study area was done by taking the October SPOT image as a base image. The corrected images were resampled to the same spatial resolution (15 m × 15 m). The NDVI values for each pixel were calculated using the following formula [39]:

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R}, \quad (1)$$

where NIR is reflection in near infrared and R is reflection in red wavelength. NDVI values of the four images were compared based on their vegetation biomass distribution. Two images with maximum and minimum vegetation cover were selected: an ASTER image taken on March 19 when most of the watershed area was bare and a SPOT image taken on October 8 when the most was vegetated. This enabled assessment of the best time for satellite images to be taken for improving soil prediction. The other two images provided less extra information and were not considered in further analysis. Theoretically, the image in March reflected variations in soil appearance or color because the soil was bare at that time in Ethiopia (direct effect), while the image in October reflected the indirect effect of soil characteristics on vegetation cover and greenness because the soil was covered with vegetation at that time.

Terrain attributes and satellite data for each soil observation were extracted. Statistical analyses were implemented, within each class, between the derived terrain attributes, satellite data, and collected soil attributes of field observations using SPSS. Multiple linear regression models are usually employed to predict dependent (soil attributes) from independent variables (satellite data and terrain attributes). From a total of 220 observations, 180 were randomly selected for analysis, and the remaining 40 observations were used to assess the accuracy of the prediction model. A regression model was established for each class to predict soil attributes from terrain attributes and satellite data. Map algebra of ArcGIS was used to obtain predicted soil attribute grids using the regression models and the raster grids for each class (slope percent, contributing area, CTI, aspect classes,

curvature classes, and NDVI values of ASTER and SPOT images). An example of these equations follows:

$$
\begin{aligned}
\text{Clay content (class 1)} \\
= 17.69 + &((\text{grid\_slope}) * 0.07) \\
+ &((\text{grid\_flowaccumulation}) * 0.00002) \\
+ &((\text{grid\_CTI}) * 0.12) \\
+ &((\text{grid\_aspect}) * 0.25) \\
+ &((\text{grid\_meancurvature}) * 0.54) \\
+ &((\text{grid\_NDVI-March}) * -0.004) \\
+ &((\text{grid\_NDVI-October}) * -0.006).
\end{aligned}
$$
$$(2)$$

Predictions were first executed for individual classes and were then merged together to generate predicted values for the whole watershed. The prediction accuracy was verified in two ways: first, by comparing the predicted and observed values using 40 randomly selected field observations (Figure 1); and second, by comparing the predicted soil attributes with those derived using the inverse distance weighted (IDW) technique.

The role of satellite data in improving the prediction accuracy of soil attributes using multiple linear regression models was assessed by repeating the previous analysis without using the satellite images as an independent variable. This was done for one of the soil attributes (clay content). Regression coefficients (R^2) of the models without including remote sensing data were compared with those that included satellite data. The quality of the predictions with and without satellite data was also tested by examining the relationship between observed and predicted values when some differences between them were allowed. This comparison was implemented by allowing a difference between predicted and observed clay content of 3, 5, or 7%—because it was not expected that predicted values would exactly equal the observed values (where the difference allowed was zero). In fact, within a distance of 1 m in the field, there are some differences in some soil attributes within these ranges.

2.4. Selection of the Optimum Number of Observations for Soil Prediction.

The effect of the number of field observations used to build the regression models on the accuracy of the prediction model (sensitivity analysis) was investigated using soil depth as an example. This was to determine the minimum number of observations to maintain acceptable prediction accuracy of soil attributes. Various numbers of observations were selected, starting from 180 and gradually decreasing: 150, 120, 90, 60, 40, 30, 25, and finally 20 observations. The prediction model was implemented until a threshold was reached where the prediction accuracy declined sharply to a low level. The classes used in the earlier modeling were amalgamated to get an optimum number of observations per class. The accuracies of predicting soil depth as a result of

TABLE 1: Descriptive statistics of the soil attributes and site characteristics for soil observations.

Variable	Minimum	Maximum	Mean	Mode	Std. deviation
Soil depth (cm)	0	101	43.5	101	33.0
Elevation (m)	1954	2852	2252	2020	220
Slope (%)	2	90	34.8	30	23.3
Stone cover at the surface (%)	0	65	15.4	10	14.2
Stone in the soil (%)	0	73.8	17.4	0	13.2
Clay content (%)	11.6	67.8	28.6	20.3	12.6
Silt content (%)	4.0	49.8	34.0	39.8	7.7
Sand content (%)	14.6	72.6	37.3	42.6	9.3
Organic matter (%)	0.21	9.97	2.82	1.12	1.9
Total nitrogen (%)	0.03	9.06	0.26	0.22	0.6
Available phosphorus (mg kg^{-1})	0.5	97.2	14.1	4.5	15.3
Bulk density (g cm^{-3})	0.81	1.70	1.25	1.21	0.2
pH	5.34	7.98	6.72	6.62	0.41
Erosion status	Low	Severe	—	Moderate	—
Erosion type	Sheet	Gully	—	Rill	—

using different numbers of soil observations (density) were estimated using different validation indices [27, 40], such as the root mean square error (RMSE) and the mean absolute estimation error (MAEE). As the difference between predicted values and observed values increases, so do both RMSE and MAEE. Consider the following:

$$\mathrm{RMSE} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\left(\acute{Y}_i - y_i\right)^2},$$

$$\mathrm{MAEE} = \frac{1}{n} \sum_{i=1}^{n} \left|\acute{Y}_i - y_i\right|,$$

(3)

where n is the number of verification points, $\acute{Y}_i$ is predicted soil depth, and y_i is observed soil depth. The quality of the prediction using different densities of observations was also tested by examining the relationship between observed and predicted values by allowing some differences between them. This enabled easy observation of the gradual change in percentage of observations predicted correctly when different observation densities were used.

3. Results and Discussion

Descriptive statistics for soil attributes and site characteristics were recorded for each site (Table 1). The correlation between soil attributes and terrain attributes and satellite data, when considering the whole set of soil observations, was very low (0.0–0.50); thus, building a regression model to predict soil attributes would not produce acceptable results. However, the range of R^2 between soil attributes and terrain attributes and satellite data when the whole watershed was divided into smaller subwatersheds was 0.06–0.85 (Table 2). R^2 depended on the type of soil attribute and the subwatershed facet class for which the relationship was being established. In general, the R^2 values were acceptable compared with previous studies

[40–42] and were used to generate predictions of the various soil attributes within the seven classes. However, the final judgment of the prediction accuracy is made through the comparison between predicted and measured values of soil characteristics. Significant correlations between terrain attributes and satellite images with soil attributes varied for the same soil attribute for different classes (Table 3). For example, in class 2, clay content was significantly correlated with slope percent, upslope contributing area, and the ASTER image taken in March, whereas, in classes 3 and 7, clay content was highly correlated with slope percent, aspect, curvature, the ASTER image taken in March, and the SPOT image taken in October. These relationships indicate that classifying the watershed into classes improved the prediction of soil attributes. The results were supported by findings of other researchers who reported low R^2 (0.0–0.19) between various soil attributes and terrain attributes for the whole watershed, which was also improved significantly for the classified subwatersheds and consequently improved the prediction accuracy of soil attributes [37].

The regression coefficients (Table 3) indicated that digital terrain attributes had a stronger influence on soil properties. This was supported by previous studies [43, 44]. In nature, soil properties are highly spatially variable [45, 46], and, for accurate estimation of soil properties, this continuous variability should be considered. In addition, in nature, land is not flat (two dimensional) as it is represented on the map. Hence, prediction will be more reliable if a three-dimensional model is used in which terrain attributes provide a quantification of landform shape and connectivity that define geomorphometry and water flow patterns [37, 47–49].

The RMSE between predicted soil characteristics and those observed in the field (Table 4) indicated good accuracy of prediction using the regression models for most soil characteristics when compared with previous studies [38, 40, 50]. Bayer et al. found that two regression techniques had similar capabilities to provide significant prediction models

TABLE 2: Regression coefficients (R^2) of the predicted soil attributes for different classes.

Class	Soil depth (cm)	Clay content (%)	Silt content (%)	Sand content (%)	Organic matter (%)	Bulk density (gm cm^{-3})	pH	Total nitrogen (%)	Available phosphorus (ppm)	Stone at surface (%)	Stone in soil (%)
1	0.45	0.12	0.21	0.16	0.25	0.34	0.39	0.19	0.16	0.10	0.13
2	0.45	0.50	0.26	0.53	0.32	0.21	0.18	0.09	0.07	0.37	0.58
3	0.72	0.80	0.73	0.81	0.64	0.54	0.48	0.72	0.68	0.26	0.32
4	0.76	0.54	0.42	0.57	0.85	0.65	0.37	0.78	0.71	0.69	0.79
5	0.53	0.38	0.26	0.06	0.26	0.48	0.79	0.19	0.50	0.21	0.54
6	0.34	0.52	0.24	0.52	0.18	0.26	0.14	0.26	0.18	0.07	0.28
7	0.25	0.33	0.33	0.26	0.25	0.36	0.32	0.19	0.15	0.17	0.18

TABLE 3: Regression models to predict clay percent of the surface layer using terrain attributes and satellite images.

Class no.	Constant	Slope (%)	As	CTI	Aspect	Curvature	ASTER March	SPOT October	R^2
1	17.69	0.07	0.00002	0.12	0.25	0.54	−0.004	−0.006	0.12
2	42.93	−0.33**	0.0001*	−0.75	0.31	1.08	−0.08**	0.04	0.50
3	39.16	−0.53*	0.0001	1.34	3.93**	−2.63**	0.04	−0.20	0.80
4	22.12	−1.37*	−0.0001	3.68	2.79	0.76	−0.02	−0.13	0.54
5	−2.82	−0.06	−0.0001	2.74	1.94	0.15	−0.01	−0.04	0.38
6	53.77	−0.25**	0.0001**	−1.20	−1.72*	−0.20	−0.05**	0.007	0.52
7	12.29	−0.18*	−0.00003	1.75	1.40	1.62	−0.01*	−0.04*	0.33

*Significant at the 0.05 probability level; **significant at the 0.01 probability level.
As: upslope contributing area; CTI: compound topographic index.

for soil organic carbon and iron oxides [51]. In the present study, comparing this RMSE with that derived from spatial interpolation using the IDW method indicated favourably comparable accuracy of the prediction model; RMSE of the prediction model was lower than that of the IDW in all cases. Similar results were also observed when the number of observations used was reduced to 60 (Table 4). In addition, comparing the RMSE differences (between the prediction model and the IDW) using 180 with that of 60 observations indicated higher increments in error term differences as the number of observations used decreased (Table 4). As a result, the prediction model was highly preferred due to higher accuracy than IDW, especially when a limited number of observations (60) were used. This is because the soil-landscape model provides estimates based on the characteristics of the terrain between two observations and therefore improves accuracy. This is not the case when interpolation is done with the assumption that closest observations are also closer in their soil characteristics, which is not always true. The soil-landscape model uses the background topography and satellite data to characterize each observation and then predict their soil attributes with consideration of the characteristics of these observations. This is consistent with soil genesis theories that are not considered in interpolation techniques. Previous research indicated that the use of different regression techniques enables the prediction of different topsoil parameters in a rapid and nondestructive manner, while avoiding the spatial accuracy problems associated with spatial interpolation [52].

The results indicated that soil attributes were predicted with acceptable accuracy using SRTM 90 m DEM and also provide a visual representation of the spatial distribution of soil attributes (Figure 3). This indicated that soil attributes can be predicted with low resolution DEMs. These conclusions were similar to those of Thompson et al. [53] and Wechsler [54], who concluded that high-resolution DEMs are not always necessary for soil-landscape modeling. Chabrillat et al. (2002) reported that although higher spatial resolution provided a purer image in more heterogeneous sites, it did not identify new features that a lower spatial resolution data set would miss in homogeneous terrain [55].

When only terrain attributes without satellite images were used to build multiple regression models (Table 5), the R^2 declined for all classes compared with those obtained using both terrain and satellite data (Table 3). For example, there were large decreases in R^2 in class 5 from 0.38 (Table 3) to 0.18 (Table 5) and in class 2 from 0.8 (Table 3) to 0.68 (Table 5). This indicated that satellite images could improve the prediction of soil attributes when used together with terrain attributes in building regression models. Gerighausen et al. found that the use of multiannual image data enhanced the prediction of soil parameters [56]. The percent of agreement between the predicted and observed clay content using satellite images and terrain attributes was compared with those derived using terrain attributes only. The prediction accuracy of clay content (percent of correctly predicted observations) increased as a result of using satellite images with terrain attributes compared to using terrain attributes only. For example, 50% of the observations showed agreement in predicting clay contents within a difference of ±7% between observed and predicted values when only terrain attributes were used to build the model (Table 6). The above agreement

TABLE 4: Root mean square error between field observations and predicted soil attributes for multiple regression and IDW and using different densities of observations.

Predicted soil attributes	R_180 (A)	I_180 (B)	R_60 (C)	I_60 (D)	B-A	D-C
Soil depth (cm)	26.4	32.64	33.7	43.37	6.24	9.67
Clay (%)	12.6	16.64	12.70	20.51	4.04	7.81
Silt (%)	7.3	10.25	8.59	14.64	2.95	6.05
Sand (%)	9.4	11.47	10.24	15.44	2.07	5.2
Organic matter (%)	1.39	1.53	1.55	1.82	0.14	0.27
Bulk density (g cm^{-3})	0.18	0.25	0.24	0.36	0.07	0.12
pH	0.38	0.54	0.46	0.83	0.16	0.37
Total N (%)	0.29	0.46	0.09	0.34	0.17	0.25
Available P (mg kg^{-1})	19.41	20.35	17.12	21.22	0.94	4.1
Surface stone cover (%)	12.2	14.53	14.15	20.62	2.33	6.47
Stone in the soil (%)	12.4	15.71	17.13	26.38	3.31	9.25

R_180 (A): root mean square error calculated from those predicted using 180 observations by multiple regression models, I_180 (B): root mean square error calculated from spatial interpolation (IDW: inverse distance weighted) using 180 observations, R_60 (C): root mean square error calculated from those predicted using 60 observations by multiple regression models, I_60 (D): root mean square error calculated from spatial interpolation (IDW: inverse distance weighted) using 60 observations, B-A: the difference in root mean square errors between those interpolated from 180 observations and those predicted using 180 observations by multiple regression models, and D-C: the difference in root mean square errors between those interpolated from 180 observations and those predicted using 180 observations by multiple regression models.

TABLE 5: Regression models to predict clay percent of the surface layer using terrain attributes only (without satellite images).

Class No.	Constant	Slope (%)	As	CTI	Aspect	Curvature	R^2
1	15.06	0.06	0.00001	0.23	0.27	0.61	0.10
2	46.18	−0.53**	0.0001*	−1.29	0.88	1.07	0.37
3	19.31	−0.28*	0.00006	0.29	4.54**	−4.55**	0.68
4	3.44	−1.38*	−0.00008	2.68	3.17	2.37	0.46
5	0.97	−0.11	−0.0002	1.78	1.25	1.10	0.18
6	66.35	−0.40**	0.0002**	−2.30	−2.11*	−1.41	0.45
7	3.46	−0.22*	−0.00004	1.85	1.25	2.52	0.24

*Significant at the 0.05 probability level; **significant at the 0.01 probability level.
As: upslope contributing area; CTI: compound topographic index.

TABLE 6: Agreement between observed and predicted values of soil attributes using terrain attributes with and without satellite images.

Predicted clay content using terrain attributes and satellite images		Predicted clay content using terrain attributes only (without satellite images)	
D*	%§	D*	%§
0	5.0	0	0
3	27.5	3	22.5
5	50.0	5	35
7	62.5	7	50

*Magnitude of difference between observed and predicted value; §percent of observations where predicted values agreed with observed values.

was improved to 62.5% when satellite images were used with the terrain attributes. The results showed that satellite images improved the prediction of soil attributes when used with terrain attributes to build multiple regression models. This is because satellite images provide more information about the factors controlling soil variability in addition to the topographic information provided by DEM-derived terrain attributes. Chabrillat et al. (2002) indicated that imaging spectrometry aids in the detection and mapping of expansive clays [55]. The images taken at two different dates provided direct and indirect information. The direct information was through the variation in the soil surface reflectance, which is related to soil variability (when the soil surface is bare during March). The indirect information was through the reflectance of the different vegetation cover, which reflects the variation in soil type through the effect on vegetation cover characteristics during October. Many researchers have suggested that remote sensing helps in providing accurate prediction of soil properties [27–29]. Bartholomeus et al. (2012) showed that although variation of soil properties within vegetation classes is large, the vegetation composition is useful to estimate soil properties [57].

Since the spectral reflectance of both satellite images was significantly correlated with clay content (Table 3), we can conclude that two images taken during dry and rainy seasons could be sufficient, with terrain attributes, to predict soil attributes using multiple linear regression models. During the dry period, there was a correlation between soil properties (clay content) and spectral reflectance in the visible range [58]. During the rainy season, most of the area was covered with vegetation that reflects variation in soil properties such
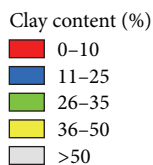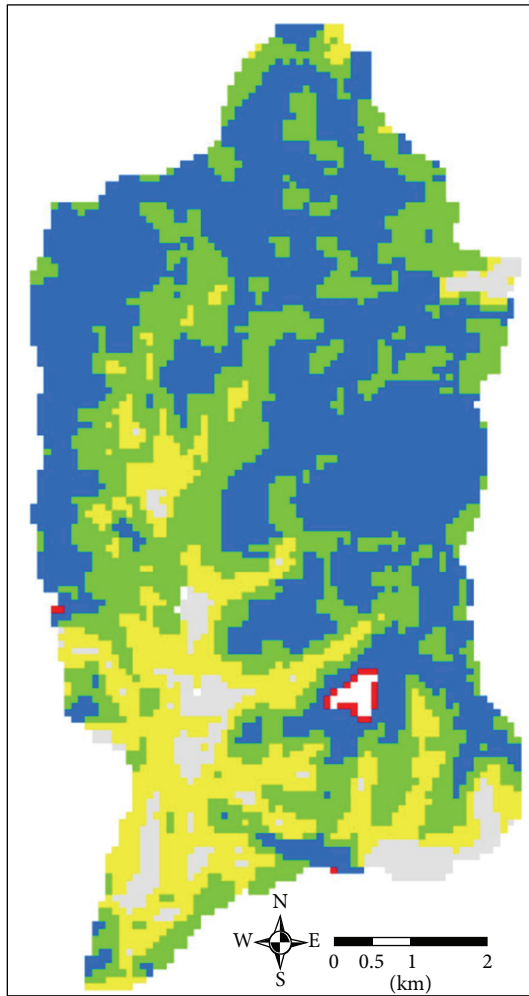
Clay content (%)
- ■ 0–10
- ■ 11–25
- ■ 26–35
- ■ 36–50
- ■ >50

FIGURE 3: Predicted clay content (%) of the surface soil using terrain attributes, satellite images, and 180 field observations.

TABLE 7: Percent of correctly predicted observations within ±50 cm of the observed soil depth, RMSEs, and MAEEs calculated using different field observation densities.

Number of observations used	Percent of correctly predicted observations within ±50 cm	RMSE (cm)	MAEE (cm)
180	97.5	26.4	21.7
150	95	29.7	25.5
120	92.5	31.4	26.5
90	87.5	31.7	26.5
60	87.5	33.7	29.4
40	72.5	40.9	31.3
30	67.5	54.3	42.5
25	65	57.5	46.4
20	35	126.0	88.2

first into five classes, and several trials were done to predict soil depth using 150, 120, and 90 observations to build regression models. The analysis was repeated based on two classes to build regression models using only 60, 40, 30, 25, and 20 observations. This was because, when a limited number of observations were used (≤60 observations), there were insufficient observations if five or seven classes were used (the number of observations for each class was too low to build good regression models). The regression models (and R^2) to predict soil depth at different densities of field observations using terrain attributes and satellite images were used to get predicted soil-depth grids (Figure 4). The proportion of predicted soil-depth of zero (no soil) increased dramatically as the number of observations used reduced from 25 to 20 (Figure 4).

The optimum number of observations to produce an acceptable prediction of soil attributes was assessed using percent of observations where the soil depth was correctly predicted within an acceptable range of agreement (using RMSE and MAEE). The accuracy of predicting soil depth decreased from 97.5% when 180 field observations were used to 95% for 150 observations, reached 65.0% using 25 observations, and then dropped drastically to 35% for 20 observations (Table 7). There were gradual increases in RMSE and MAEE from the predictions using 180 observations (26.4 and 21.7 cm, resp.) to the predictions using 25 observations (57.5 and 46.4 cm, resp.); then, both increased sharply for 20 observations (126 and 88.2 cm, resp.; Table 7). Since there was a high increment in RMSE and MAEE values at 20 compared with 25 observations, it can be considered that 25 observations (or one observation/2 km^2) were optimum to build multiple regression models in soil-landscape modeling for this particular watershed. The accuracy assessment methods used for selecting the optimum number of observations to predict soil depth indicated that 25 observations were the minimum required for multiple linear regression models to produce acceptable soil prediction. The results also indicated that the prediction accuracy was the same when 60 or 90 observations were used (87.5%). However, a close look at

as clay content. As a result, soil properties could be indirectly predicted by determining the spectral reflectance of the vegetation cover. However, the satellite image taken when the field was bare showed higher correlation with all soil attributes compared with the satellite image taken when the field was covered by vegetation. Hence, the best time to acquire satellite images to improve the prediction of soil attributes would be during the dry period when most of the area is bare. Mulder et al. [29] also reported similar findings and successfully used airborne, space borne, and in situ measurements using various instruments.

Selecting Optimum Number of Field Observations to Predict Soil Attributes. The seven classes that were used in the earlier modeling (and using 180 observations) were amalgamated
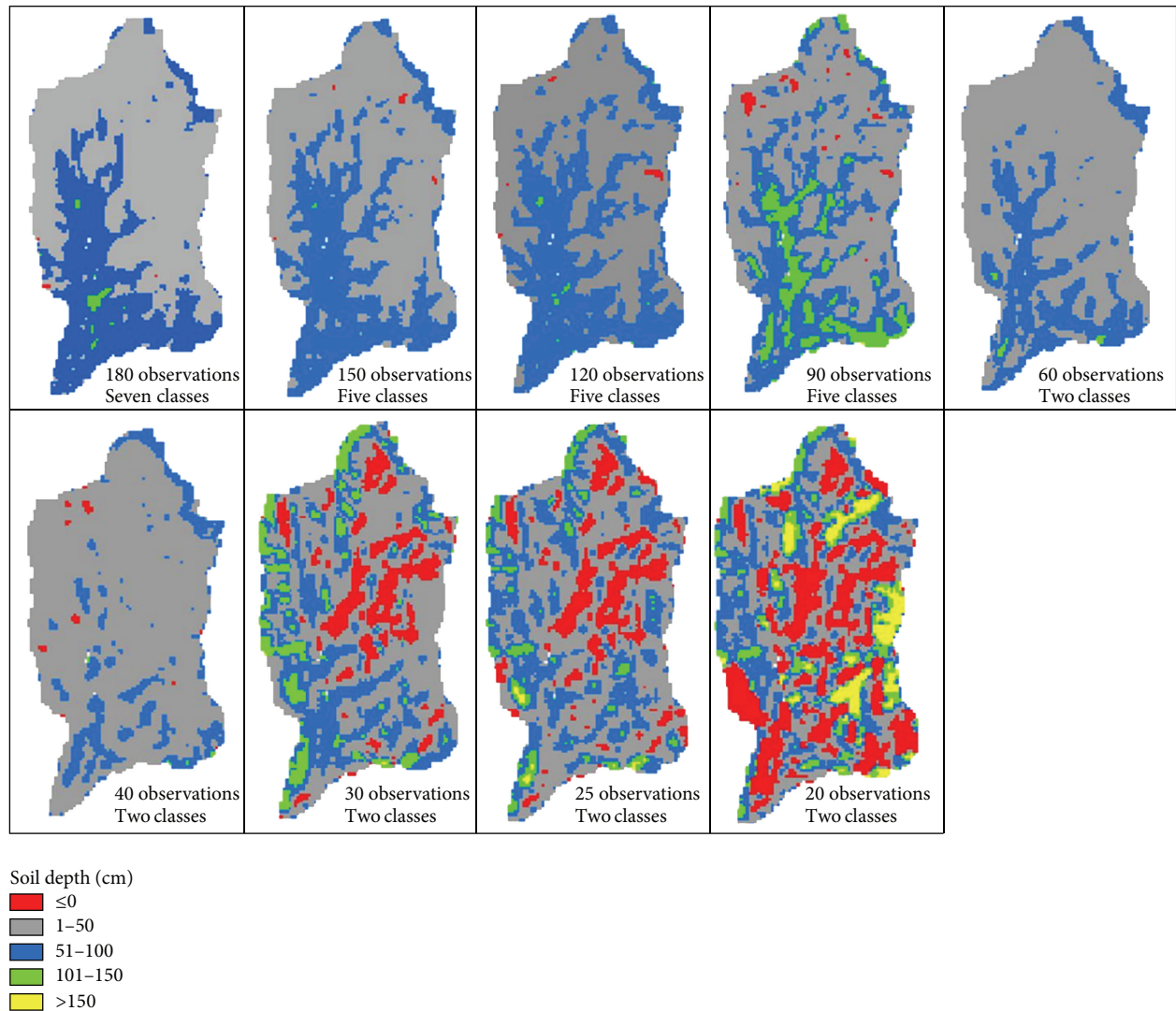
Soil depth (cm)
- ≤0
- 1–50
- 51–100
- 101–150
- >150

FIGURE 4: Predicted soil depth (cm) using terrain attributes, satellite images, and different density of field observations.

Figure 4 revealed that the spatial distribution of the predicted soil depth classes drastically changed when the number of observations used was <60. Therefore, it can be suggested that, for this study area, the optimum number of observations to generate an acceptable prediction is somewhere between 60 and 25 observations (1-2 observations/2 km^2). The users can select between these observation densities to produce an optimum results. The approach indicated that soil-landscape modeling could be used with a low number of observations to produce accurate predictions of soil attributes with an acceptable representation of the spatial distribution of soil attributes over the landscape.

4. Conclusions

The use of satellite image data together with terrain attributes improved the prediction of soil attributes. Two satellite images taken at different dates, one in the dry season and the other in the rainy season, were sufficient to capture the variation in soil reflectance and improve the prediction accuracy. If one image is available, the best acquisition time to facilitate soil-landscape modeling is during the dry season when most of the soil surface is bare. Generally, the results indicated that soil attributes were predicted with acceptable accuracy using multiple linear regression models from freely available DEM (90 m resolution) and satellite images with a minimum of field work. In addition, the prediction model was highly preferred due to comparable accuracy with the spatial interpolation using IDW, especially when a limited number of observations were used, which is usually the case in data-scarce areas. For this study, the optimum number of observations to predict soil attributes using multiple regression models and using terrain attributes and remote sensing data was between 60 and 25 observations (approximately 1-2 observations/2 km^2). The produced predictions will be very useful to provide information for detailed modeling activities, especially for a country like Ethiopia in which information on the detailed spatial distribution of soil attributes is very

scarce. Further investigations should be made for applying the model over larger areas with diverse soils and landscape features to validate the results and to out-scale their application.

Conflict of Interests

The authors declare no conflict of interests with any trademark or software mentioned in this paper.
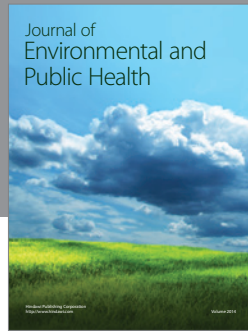
Acknowledgments

References

[1] J. Bouma, "The role of soil science in the land use negotiation process," *Soil Use and Management*, vol. 17, no. 1, pp. 1–6, 2001.

[2] A. R. Mermut and H. Eswaran, "Some major developments in soil science since the mid-1960s," *Geoderma*, vol. 100, no. 3-4, pp. 403–426, 2001.

[3] M. H. Salehi, M. K. Eghbal, and H. Khademi, "Comparison of soil variability in a detailed and a reconnaissance soil map in central Iran," *Geoderma*, vol. 111, no. 1-2, pp. 45–56, 2003.

[4] C.-W. Ahn, M. F. Baumgardner, and L. L. Biehl, "Delineation of soil variability using geostatistics and fuzzy clustering analyses of hyperspectral data," *Soil Science Society of America Journal*, vol. 63, no. 1, pp. 142–150, 1999.

[5] N. J. McKenzie, P. E. Gessler, P. J. Ryan, and D. A. O'Connell, "The role of terrain analysis in soil mapping," in *Terrain Analysis: Principles and Applications*, J. P. Wilson and J. C. Gallant, Eds., Chapter 10, John Wiley and Sons, New York, NY, USA, 2000.

[6] A.-X. Zhu, "Mapping soil landscape as spatial continua: the neural network approach," *Water Resources Research*, vol. 36, no. 3, pp. 663–677, 2000.

[7] P. A. Burrough, P. F. M. Van Gaans, and R. Hootsmans, "Continuous classification in soil survey: spatial correlation, confusion and boundaries," *Geoderma*, vol. 77, no. 2–4, pp. 115–135, 1997.

[8] A.-X. Zhu, "A similarity model for representing soil spatial information," *Geoderma*, vol. 77, no. 2–4, pp. 217–242, 1997.

[9] J. Balkovič, G. Čemanová, J. Kollár, M. Kromka, and K. Harnová, "Mapping soils using the fuzzy approach and regression-kriging—case study from the Považský Inovec Mountains, Slovakia," *Soil and Water Research*, vol. 2, no. 4, pp. 123–134, 2007.

[10] A. B. McBratney, M. L. Mendonça Santos, and B. Minasny, "On digital soil mapping," *Geoderma*, vol. 117, no. 1-2, pp. 3–52, 2003.

[11] D. M. Browning and M. C. Duniway, "Digital soil mapping in the absence of field training data: a case study using terrain attributes and semi automated soil signature derivation to distinguish ecological potential," *Applied and Environmental Soil Science*, vol. 42, pp. 1904–1910, 2011.

[12] S. Lamsal, S. Grunwald, G. L. Bruland, C. M. Bliss, and N. B. Comerford, "Regional hybrid geospatial modeling of soil nitrate-nitrogen in the Santa Fe River Watershed," *Geoderma*, vol. 135, pp. 233–247, 2006.

[13] P. Scull, J. Franklin, and O. A. Chadwick, "The application of classification tree analysis to soil type prediction in a desert landscape," *Ecological Modelling*, vol. 181, no. 1, pp. 1–15, 2005.

[14] E. Giasson, R. T. Clarke, A. V. Inda Jr., G. H. Merten, and C. G. Tornquist, "Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil," *Scientia Agricola*, vol. 63, no. 3, pp. 262–268, 2006.

[15] A. K. Stum, *Random forests applied as a soil spatial predictive model in arid Utah [M.S. thesis]*, Utah State University, 2010.

[16] X. Lui, J. Peterson, Z. Zhang, and S. Chandra, *Improving Soil Salinity Prediction with Resolution DEM Derived for LIDAR Data*, Center for GIS; School of Geography and Environmental Science Monash University, Melbourne, Australia, 2006.

[17] E. Aksoy, G. Özsoy, and M. S. Dirim, "Soil mapping approach in GIS using landsat satellite imagery and dem data," *African Journal of Agricultural Research*, vol. 4, no. 11, pp. 1295–1302, 2009.

[18] A. Al-Shamiri and F. M. Ziadat, "Soil-landscape modeling and land suitability evaluation: the case of rainwater harvesting in a dry rangeland environment," *International Journal of Applied Earth Observation*, vol. 18, pp. 157–164, 2012.

[19] D. G. Rossiter, *Digital Soil Mapping: Towards a Multiple-Use Soil information System*, Santa fe Bogota, Colombia, 2005.

[20] J. D. Phillips, "Spatial structures and scale in categorical maps," *Geographical and Environmental Modelling*, vol. 6, no. 1, pp. 41–57, 2002.

[21] I. D. Moore, P. E. Gessler, G. A. Nielsen, and G. A. Peterson, "Soil attribute prediction using terrain analysis," *Soil Science Society of America Journal*, vol. 57, no. 2, pp. 443–452, 1993.

[22] P. E. Gessler, I. D. Moore, N. J. McKenzie, and P. J. Ryan, "Soil-landscape modelling and spatial prediction of soil attributes," *International Journal of Geographical Information Systems*, vol. 9, no. 4, pp. 421–432, 1995.

[23] H.-T. Jiang, F.-F. Xu, Y. Cai, and D.-Y. Yang, "Weathering characteristics of sloping fields in the Three Gorges Reservoir Area, China," *Pedosphere*, vol. 16, no. 1, pp. 50–55, 2006.

[24] X.-Y. Zhang, Y.-Y. Sui, X.-D. Zhang, K. Meng, and S. J. Herbert, "Spatial variability of nutrient properties in black soil of northeast china1 1 project supported by the National Basic Research Program (973 Program) of China (No. 2005CB121108) and the heilongjiang provincial natural science foundation of China (No. C2004-25)," *Pedosphere*, vol. 17, no. 1, pp. 19–29, 2007.

[25] I. Esfandiarpoor Borujeni, M. H. Salehi, N. Toomanian, J. Mohammadi, and R. M. Poch, "The effect of survey density on the results of geopedological approach in soil mapping: a case study in the Borujen region, Central Iran," *Catena*, vol. 79, no. 1, pp. 18–26, 2009.

[26] S. L. Kuriakose, S. Devkota, D. G. Rossiter, and V. G. Jetten, "Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India," *Catena*, vol. 79, no. 1, pp. 27–38, 2009.

[27] U. Mishra, *Predicting storage and dynamics of soil organic carbon at a regional scale [Ph.D. thesis]*, Ohio State University, 2009.

[28] A. Marchetti, C. Piccini, S. Santucci, I. Chiuchiarelli, and R. Francaviglia, "Simulation of soil types in Teramo province (Central Italy) with terrain parameters and remote sensing data," *Catena*, vol. 85, no. 3, pp. 267–273, 2011.

[29] V. L. Mulder, S. de Bruin, M. E. Schaepman, and T. R. Mayr, "The use of remote sensing in soil and terrain mapping—a review," *Geoderma*, vol. 162, no. 1-2, pp. 1–19, 2011.

[30] Food and Agriculture Organization (FAO), *Guidelines for Soil Descriptions*, FAO, Rome, Italy, 2006.

[31] P. R. Day, "Hydrometer method of particle size analysis," in *Methods of Soil Analysis*, C. A. Black, Ed., Agronomy Part II, No. 9, pp. 562–563, American Society of Agronomy, Madison, Wis, USA, 1965.

[32] G. R. Blake, "Bulk density," in *Methods of Soil Analysis Agron*, C. A. Black, Ed., Part I, No. 9., pp. 374–399, American Society of Agronomy, Madison, Wis, USA, 1965.

[33] A. Walkley and C. A. Black, "An examination of different methods for determining soil organic matter and the proposed modification by the chromic acid titration method," *Soil Science*, vol. 37, pp. 29–38, 1934.

[34] S. Sahlemeden and B. Taye, *Procedure for Soil and Plant Analysis*, National Soil Research Center; Ethiopian Agricultural Research Organization, Addis Ababa, Ethiopia, 2000.

[35] S. R. Olsen, C. V. Cole, F. S. Watanabe, and L. A. Dean, *Estimation of Available Phosphorus in Soil by Extraction with Sodium Bicarbonate*, vol. 939 of *Circular*, USDA, 1954.

[36] J. Murphy and J. P. Riley, "A modified single solution method for the determination of phosphate in natural waters," *Analytica Chimica Acta*, vol. 27, pp. 31–36, 1962.

[37] F. M. Ziadat, "Analyzing digital terrain attributes to predict soil attributes for a relatively large area," *Soil Science Society of America Journal*, vol. 69, no. 5, pp. 1590–1599, 2005.

[38] F. M. Ziadat, "Prediction of soil depth from digital terrain data by integrating statistical and visual approaches," *Pedosphere*, vol. 20, no. 3, pp. 361–367, 2010.

[39] P. Roudgarmi, M. Monavari, J. Feghhi, J. Nouri, and N. Khorasani, "Environmental impact prediction using remote sensing images," *Journal of Zhejiang University A*, vol. 9, no. 3, pp. 381–390, 2008.

[40] M. L. Kunkel, A. N. Flores, T. J. Smith, J. P. McNamara, and S. G. Benner, "A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain," *Geoderma*, vol. 165, no. 1, pp. 1–11, 2011.

[41] K. Sumfleth and R. Duttmann, "Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators," *Ecological Indicators*, vol. 8, no. 5, pp. 485–501, 2008.

[42] J. Van de Wauw, G. Baert, J. Moeyersons et al., "Soil-landscape relationships in the basalt-dominated highlands of Tigray, Ethiopia," *Catena*, vol. 75, no. 1, pp. 117–127, 2008.

[43] P. E. Gessler, O. A. Chadwick, F. Chamran, L. Althouse, and K. Holmes, "Modeling soil-landscape and ecosystem properties using terrain attributes," *Soil Science Society of America Journal*, vol. 64, no. 6, pp. 2046–2056, 2000.

[44] I. V. Florinsky, R. G. Eilers, G. R. Manning, and L. G. Fuller, "Prediction of soil properties by digital terrain modelling," *Environmental Modelling and Software*, vol. 17, no. 3, pp. 295–311, 2002.

[45] P. Burrough, "Soil variability: a late 20th century view," *Soils and Fertilizers*, vol. 56, pp. 529–562, 1993.

[46] B. P. Umali, D. P. Oliver, S. Forrester et al., "The effect of terrain and management on the spatial variability of soil properties in an apple orchard," *Catena*, vol. 93, pp. 38–48, 2012.

[47] E. Dobos, E. Micheli, M. F. Baumgardner, L. Biehl, and T. Helt, "Use of combined digital elevation model and satellite radiometric data for regional soil mapping," *Geoderma*, vol. 97, no. 3-4, pp. 367–391, 2000.

[48] Y. A. Pachepsky, D. J. Timlin, and W. J. Rawls, "Soil water retention as related to topographic variables," *Soil Science Society of America Journal*, vol. 65, no. 6, pp. 1787–1795, 2001.

[49] Y. Sulaeman and H. Subagyo, "Modeling soil landscape relationships," *Jurnal Ilmu Tanah Dan Lingkungan*, vol. 5, no. 2, pp. 1–14, 2005.

[50] S. Zhang, Y. Huang, C. Shen, H. Ye, and Y. Du, "Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information," *Geoderma*, vol. 171-172, pp. 35–43, 2012.

[51] A. Bayer, M. Bachmann, A. Müller, and H. Kaufmann, "A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded south african ecosystem," *Applied and Environmental Soil Science*, vol. 2012, Article ID 971252, 20 pages, 2012.

[52] T. Selige, J. Böhner, and U. Schmidhalter, "High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures," *Geoderma*, vol. 136, no. 1-2, pp. 235–244, 2006.

[53] J. A. Thompson, J. C. Bell, and C. A. Butler, "Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling," *Geoderma*, vol. 100, no. 1-2, pp. 67–89, 2001.

[54] S. Wechsler, "Uncertainties associated with digital elevation models for hydrologic applications: a review," *Hydrology and Earth System Sciences*, vol. 3, pp. 2343–2384, 2006.

[55] S. Chabrillat, A. F. H. Goetz, L. Krosley, and H. W. Olsen, "Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution," *Remote Sensing of Environment*, vol. 82, no. 2-3, pp. 431–445, 2002.

[56] H. Gerighausen, G. Menz, and H. Kaufmann, "Spatially explicit estimation of clay and organic carbon content in agricultural soils using multi-annual imaging spectroscopy data," *Applied and Environmental Soil Science*, vol. 2012, Article ID 868090, 23 pages, 2012.

[57] H. Bartholomeus, G. Schaepman-Strub, D. Blok, R. Sofronov, and S. Udaltsov, "Research article spectral estimation of soil properties in siberian tundra soils and relations with plant species composition," *Applied and Environmental Soil Science*, vol. 2012, Article ID 241535, 13 pages, 2012.

[58] J. A. Thomasson, R. Sui, M. S. Cox, and A. Al-Rajehy, "Soil reflectance sensing for determining soil properties in precision agriculture," *Transactions of the American Society of Agricultural Engineers*, vol. 44, no. 6, pp. 1445–1453, 2001.