Geoderma 311 (2018) 143–148

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning

Boniface H.J. Massawe [a,b,*], Sakthi K. Subburayalu [a], Abel K. Kaaya [b], Leigh Winowiecki [c], Brian K. Slater [a]

[a] School of Environment and Natural Resources, The Ohio State University, 210 Kottman Hall, 2021 Coffey Road, Columbus, OH 43210, USA
[b] Department of Soil and Geological Sciences, Sokoine University of Agriculture, PO Box 3008, Morogoro, Tanzania
[c] World Agroforestry Centre, United Nations Avenue, Gigiri, Nairobi, Kenya

ARTICLE INFO

Article history:
Received 21 March 2016
Received in revised form 11 November 2016
Accepted 14 November 2016
Available online 24 November 2016

Keywords:
Kilombero Valley
Numerical classification
Machine learning
Soil mapping
Decision tree analysis
DEM

ABSTRACT

Inadequacy of spatial soil information is one of the limiting factors to making evidence-based decisions to improve food security and land management in the developing countries. Various digital soil mapping (DSM) techniques have been applied in many parts of the world to improve availability and usability of soil data, but less has been done in Africa, particularly in Tanzania and at the scale necessary to make farm management decisions. The Kilombero Valley has been identified for intensified rice production. However the valley lacks detailed and up-to-date soil information for decision-making. The overall objective of this study was to develop a predictive soil map of a portion of Kilombero Valley using DSM techniques. Two widely used decision tree algorithms and three sources of Digital Elevation Models (DEMs) were evaluated for their predictive ability. Firstly, a numerical classification was performed on the collected soil profile data to arrive at soil taxa. Secondly, the derived taxa were spatially predicted and mapped following SCORPAN framework using Random Forest (RF) and J48 machine learning algorithms. Datasets to train the model were derived from legacy soil map, RapidEye satellite image and three DEMs: 1 arc SRTM, 30 m ASTER, and 12 m WorldDEM. Separate predictive models were built using each DEM source. Mapping showed that RF was less sensitive to the training set sampling intensity. Results also showed that predictions of soil taxa using 1 arc SRTM and 12 m WordDEM were identical. We suggest the use of RF algorithm and the freely available SRTM DEM combination for mapping the soils for the whole Kilombero Valley. This combination can be tested and applied in other areas which have relatively flat terrain like the Kilombero Valley.

1. Introduction

The Kilombero Valley in Tanzania presents great potential for the expansion and intensification of rice production. This valley, covering an area of about 11,600 km^2 (Kato, 2007), has been identified by the Government of Tanzania for financial and technological investments to expand and intensify rice production (TIC, 2013). Rice is the second most important cereal crop in Tanzania after maize (Bucheyeki et al., 2011), and its demand has been increasing following shift in preference by local population from traditional staples to rice, and increased market demands from neighboring countries. To develop and promote sustainable rice production intensification; farmers and policy makers need to identify the most suitable areas and respective management options. However, updated and detailed soil information to this support decision-making process is currently lacking.

Accurate soil information is crucial for informing management recommendations aimed to increase agricultural productivity and overall food security, especially in developing countries where the GDP is heavily dependent on the agricultural sector (Cook et al., 2008; Msanya et al., 2002). Relatively longer time is required to gather such information through conventional soil inventory and generally, larger amount of resources are required for such exercises (McBratney et al., 2003). Recent developments in remote and proximal sensing, computational methods and information technology, have provided means by which soil information can be collected, shared, communicated and updated more efficiently (Malone, 2013; McBratney et al., 2003; Scull et al., 2003; Vågen et al., 2013; Vågen et al., 2016; Winowiecki et al., 2016a, 2016b). Predictive soil landscape model frameworks such as the SCORPAN approach (McBratney et al., 2003) could be used to predict continuous soil classes and soil attributes that better represent soil spatial variability. The increased availability of high resolution digital elevation models (DEMs) that provide predictive variables in digital soil mapping together with the advances in machine learning techniques add to the ease of generating spatial soil information and depicting uncertainty (Hansen et al., 2009; Haring et al., 2012; Subburayalu and Slater, 2013; Subburayalu et al., 2014).

* Corresponding author at: Department of Soil and Geological Sciences, Sokoine University of Agriculture, PO Box 3008, Morogoro, Tanzania.
E-mail addresses: bonmass@yahoo.com (B.H.J. Massawe), L.A.WINOWIECKI@CGIAR.ORG (L. Winowiecki).

http://dx.doi.org/10.1016/j.geoderma.2016.11.020
0016-7061/

The overall goal of this study was to develop a predictive soil map for a portion of Kilombero Valley, Tanzania to serve as a basis for quantitative land evaluation for intensified rice production. Machine learning informed by legacy soil map, new field data collection, and multiple sources for environmental correlates were combined and used for mapping of numerically derived soil classes. In this paper we report comparisons of two machine learning algorithms and three sources of terrain data.

2. Methods

2.1. Legacy soil map

The base map used to guide soil sampling was a reconnaissance legacy soil map developed in the late 1950s at a scale of 1:125,000 (FAO, 1961). The map was obtained in a scanned format from the World Soil Survey Archive and Catalog (WOSSAC). The legacy map was prepared based on aerial photo interpretation. The air photography at a scale of 1:30,000 was done by the British Royal Air Force in years 1948, 1949, and 1950; Hunting Aerosurveys Ltd. in 1955; Fairey Air Surveys Ltd. in 1956; and Air Survey Division of Tanganyika in 1957. The scanned legacy soil map was georeferenced in QGIS version 2.2 software (QGIS Development Team, 2014) using World Geodetic Survey 1984 (WGS-84) datum and Universal Transverse Mercator (UTM) coordinate system's zone 37 south. The soil units for the study area portion were digitized on screen. The legacy map covering the study area included 10 soil groups (Fig. 1).

2.2. Field data collection

FAO Guidelines for Soil Description (FAO, 2006) were used to describe and take samples from new 33 soil profiles in the study area. The number of soil profiles exceeded the number of soil groups in the legacy soil base map by 23 because it was necessary to describe and sample on areas which appeared not to be well represented with the legacy soil map. There has been some land cover/land use change since the preparation of the legacy map in the late 1950s due to deforestation and draining of swampy areas for agriculture, grazing and settlements. Terrain data and field observations were used to locate sites for additional soil profiles. The numbers of soil profiles per legacy soil map unit here shown in parentheses were: A (5), C (1), D (4), E (1), F (5), K (8), P (2), SW (4), U (1) and W (2). The soil units have been described in Fig. 1.

In most cases it was not possible to properly describe the soil profiles below 80 cm depth because of the high water table. The soil profile morphological characteristics recorded for this study included genetic horizon designation, depths, and soil colour. Soil colour determination was done in the field by recording the hue, value, and chroma in the Soil Munsell Colour Charts book (Munsell Colour Co., 1992) corresponding to a moistened soil clod.

2.3. Laboratory analysis

Soil samples were submitted to the soil analysis lab at Sokoine University of Agriculture for wet chemistry. Soil attributes that were analyzed in the laboratory included: soil pH (McLean, 1986); electrical conductivity (Rhoades, 1982); soil texture (Gee and Bauder, 1986); total nitrogen by Kjeldahl method (Bremner and Mulvaney, 1982) and organic carbon by Walkley and Black wet oxidation method (Nelson and Sommers, 1982). Available phosphorus was extracted by Bray and Kurtz-1 method (Bray and Kurtz, 1945) for soils with pH_{water} less than 7 and Olsen method for soils with pH_{water} above 7 (Watanabe and Olsen, 1965). Cation exchange capacity of the soil (CEC) and exchangeable bases were determined by saturating soil with neutral 1 M NH_4OAc (ammonium acetate) and the adsorbed NH_4^+ were displaced using 1 M KCl. The exchangeable bases (Ca^{2+}, Mg^{2+}, Na^+, K^+) were determined by atomic absorption spectrophotometer (Thomas, 1982) while CEC was determined by Kjeldahl distillation method. Diethylenetriaminepenta-acetic acid (DTPA) method (Lindsay and Norvel, 1978) was used to extract four micronutrients: iron, manganese, copper and zinc.

2.4. Numerical classification

Numerical soil classification techniques have been suggested and tested by many authors to generate horizon classes, soil classes, and to define taxonomic distance between the classes for existing soil classification systems in different studies (Carré and McBratney, 2005; Carré and Jacobson, 2009; Rayner, 1966; McBratney et al., 2010; Muir et al., 1970; Odgers et al., 2011a; Odgers et al., 2011b; Rizzo et al., 2014).

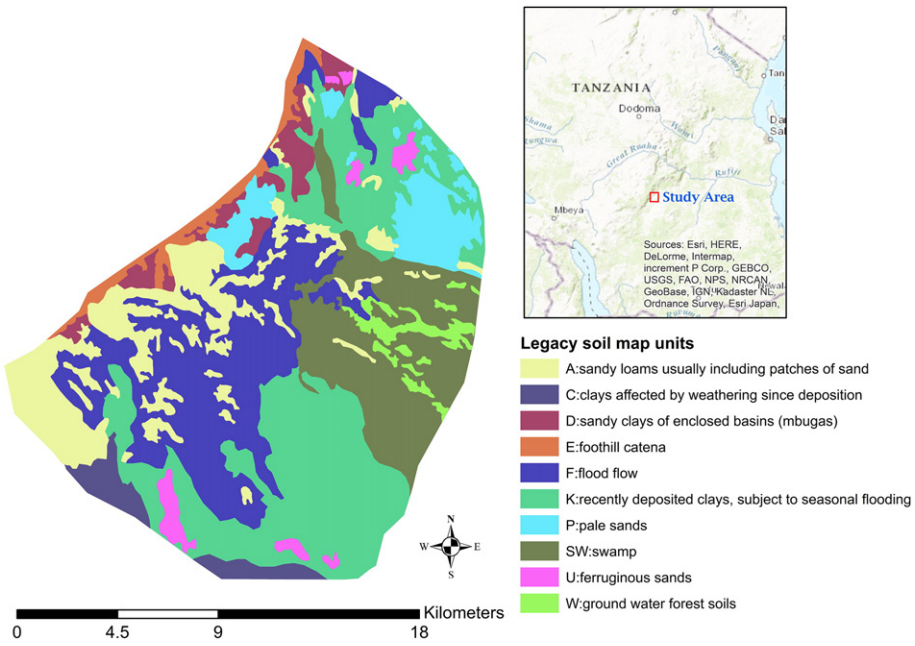

Fig. 1. Anderson's 1961 legacy soil map (Digitized from FAO, 1961). Note: Soil units' descriptions are summarized in Table 1.

Table 1
Soil Taxonomy Subgroup for the most representative (exemplar) soil profile of the soil class.

Soil class ID	No. of soil profiles identified with class	Classification (subgroup) of exemplar soil profile
S56	4	Fluvaquentic Endoaquepts
S57	2	Fluvaquentic Humaquepts
S58	5	Aquic Dystrudepts
S59	4	Aeric Umbric Endoaqualfs
S60	3	Umbric Endoaqualfs
S61	3	Fluvaquentic Humaquepts
S62	3	Fluvaquentic Hapludolls
S63	1	Mollic Fluvaquents
S64	1	Typic Endoaquepts
S65	4	Aquic Udifluvents
S66	1	Typic Endoaquerts
S67	1	Oxyaquic Eutrudepts
S68	1	Aeric Endoaquents

Note: S56 to S68 are codes representing generated soil classes.

The numerical methods have a great potential for use in Africa where soil information is still in great demand (Cook et al., 2008). Despite their potential, these methods have not been commonly applied (McBratney et al., 2010).

Numerical classification of morphological and lab data for soil horizons and profiles was performed using a web based application – OSACA (Jacobson and Carré, 2006).

In preparation for data input Munsell colour notations recorded in the field (hue, value and chroma) were converted to red, green, and blue (RGB) numerical values (0–255) using mColorConverter (He, 2013). The raw data were standardized using the equation:

$$V_s = (V_o - \mu)/\sigma \quad (1)$$

where V_s is the standardized value of the attribute, V_o is its observed value, μ is sample mean and σ is the sample standard deviation.

In the OSACA options, the Euclidean distance metric was chosen for horizon clustering, while Pedological distance metric was chosen for soil profile clustering.

2.5. Assembling environmental correlates

The following environmental correlates were used for SCORPAN (McBratney et al., 2003) prediction of soil classes.

- S- soil classes derived from a 1959 legacy soil map of the area by Anderson (FAO, 1961).
- O- effects of living organisms (vegetation), derived from a 5 m resolution RapidEye satellite image.
- R- terrain parameters derived from three DEMs.
- N- spatial location was recorded for each attribute used in the prediction of soil classes.

Climate (C), lithology (P) and age (A) factors in the SCORPAN formulation were not used in this study.

ERDAS IMAGINE software was used to derive the following vegetation based parameters from the satellite image: land use/cover classes, normalized difference vegetation index (NDVI) (Gitelson et al., 1999), optimized soil adjusted vegetation index (OSAVI) (Rondeaux et al., 1996), and soil enhancement ratio (SER). Soil enhancement ratios are calculated as band3/band2, band3/band7, and band5/band7. Since RapidEye image has bands 1–5, it was only possible to calculate b3/b2.

The three available DEMs for the study area were:

i. 30 m spatial resolution ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer)
ii. 1 arc sec (approximately 30 m) spatial resolution SRTM (Space Shuttle Radar Topography Mission), and
iii. 12 m spatial resolution WorldDEM

The ASTER and SRTM DEMs are freely available on the internet. On the other hand, WorldDEM is a new data, officially released in 2014, and commercially available (Airbus Defence and Space, 2014). It is currently the finest spatial resolution terrain data available for the Kilombero Valley.

The following terrain derivatives were calculated using Whitebox Geospatial Analysis Tool 3.2 software (Lindsay, 2014): slope gradient, plan, profile, tangential, and total curvature, relative stream power index, sediment transport index, wetness index, deviation from mean elevation, difference from mean elevation, topographic ruggedness index, and flow accumulation grid. Terrain derivatives were generated after filling depressions using the Planchon and Darboux (2001) algorithm. Aspect was not used due to the flatness of the study area.

2.6. Mapping of soil classes

Machine learning was used to predict the soil taxa. Machine learning is a type of artificial intelligence that provides computers with the ability to automatically learn programs from multi-source data sets and make predictions (Witten et al., 2011). The technique has been successfully used in DSM to predict soil properties and soil classes (Stum et al., 2010; Brungard and Boettinger, 2012; Subburayalu and Slater, 2013; Subburayalu et al., 2014; Brungard et al., 2015; Vågen et al., 2016).

In this study, machine learning was performed in WEKA (Waikato Environment for Knowledge Analysis) (Witten et al., 2011). Three data sets, each with different terrain data derived from one of the three DEMs, were assembled for training and testing. Each dataset also contained satellite based derivatives and legacy soil map units. To come up with initial classifiers, training sets were developed by extracting environmental correlate information derived from the datasets listed in Section 2.5 at the x, y locations of the 33 soil profiles in Whitebox GAT 3.2 (Lindsay, 2014). Test data sets were extracted from raster centroids of 30 m spatial resolution pixels, where a total of 324,082 points were generated for the entire study area. The test datasets were then used as training data to generate new classifiers which were used for actual prediction. Two decision tree algorithms: J48 and RF were run with each training dataset to generate predictive models. The models were subsequently applied to predict soil classes across the study area and mapped using QGIS version 2.2 software (QGIS Development Team, 2014) at a spatial resolution of 30 m. Percent similarities between all combinations of DEMs and machine learning algorithms were computed by comparing the predicted soil classes at respective 30 m pixel of the predicted maps. Visual comparisons of the outputs were also done.

Table 2
Soil clusters predicted using J48 learner using ASTER, SRTM and WorldDEM datasets.

J48_ASTER	J48_SRTM	J48_WLD
S56	S56	S56
S58	S58	S58
S59	S59	S59
S60	S60	S60
S61	S61	S62
S62	S62	S64
S65	S64	S65
S68	S65	S68

Note: J48_ASTER = soil clusters predicted by using J48 learner on ASTER dataset. J48_SRTM = soil clusters predicted by using J48 learner on SRTM dataset. J48_WLD = soil clusters predicted by using J48 learner on WorldDEM dataset.

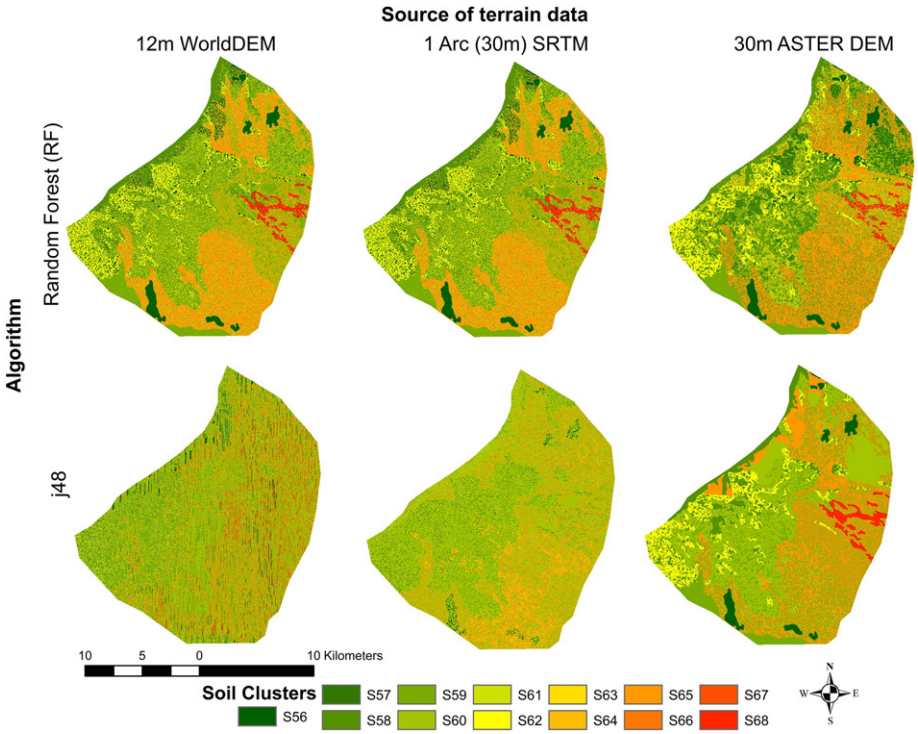

Fig. 2. Spatial distribution of soil clusters predicted by J48 and RF algorithms using 1 Arc SRTM, 30 m ASTER and 12 m WorldDEM terrain data sources. Note: Letter 'S' followed by number represent names of predicted soil clusters.

3. Results and discussions

3.1. Numerical classification

The numerical classification generated 13 soil classes (clusters). Classes grouped between one and five soil profiles. The legacy soil map identified 10 soil classes in the study area. The increased number of new soil classes identified in this study could be due to the techniques used to derive the legacy map (e.g., air photo interpretation) and the limited field data collection (due to floods and restrictive vegetation at the time) (FAO, 1961). It is also possible that some soil classes were not identified in the legacy work but were discriminated in the numerical classification. In addition, significant land use and land cover changes have occurred, which may also explain these discrepancies. For example, the legacy map describes some areas as swamps, but current field observations confirmed these areas were converted to rice cultivation. In addition, deforestation also has taken place within the Valley.

The sampled pedons with the shortest taxonomic distance to the modal pedon of each numerically synthesized soil class was classified to the Soil Taxonomy Subgroup level (Soil Survey Staff, 2014). It was generally observed that generated soil clusters were represented by different soils at Subgroup level, except for soil clusters S57 and S61 which were both classified as Fluvaquentic Humaquepts (Table 1). This suggests that clustering process was able to distinguish different soil classes in the study area.

Table 3
Comparisons of soil cluster prediction similarities between different combinations of DEM datasets and decision tree based learners' outputs.

S/N	Compared mapped predicted outputs	No. of 30 m pixels with same prediction	Similarity in soil cluster prediction (%)
1	ASTER_J48 vs ASTER_RF	175,747	54
2	ASTER_J48 vs SRM_J48	78,874	24
3	ASTER_J48 vs SRTM_RF	124,981	39
4	ASTER_J48 vs WLD_J48	84,817	26
5	ASTER_J48 vs WLD_RF	124,981	39
6	ASTER_RF vs SRTM_J48	56,153	17
7	ASTER_RF vs SRTM_RF	115,078	36
8	ASTER_RF vs WLD_J48	55,977	17
9	ASTER_RF vs WLD_RF	115,078	36
10	SRTM_J48 vs SRTM_RF	107,792	33
11	SRTM_J48 vs WLD_J48	59,468	18
12	SRTM_J48 vs WLD_RF	107,792	33
13	SRTM_RF vs WLD_J48	59,316	18
14	SRTM_RF vs WLD_RF	323,872	100
15	WLD_J48 vs WLD_RF	59,316	18

Note: ASTER_J48 = soil clusters predicted by using J48 learner on ASTER dataset. ASTER_RF = soil clusters predicted by using RF learner on ASTER dataset. SRTM_J48 = soil clusters predicted by using J48 learner on SRTM dataset. SRTM_RF = soil clusters predicted by using RF learner on SRTM dataset. WLD_J48 = soil clusters predicted by using J48 learner on WorldDEM dataset. WLD_RF = soil clusters predicted by using RF learner on WorldDEM dataset.

3.2. Algorithm comparison

The J48 algorithm predicted 8 out of the possible 13 classes for all three DEMs, 6 of which were common across all three (Table 2). J48 predicted mainly soil clusters which comprised of higher number of sampling points. Prediction of some soil classes varied depending on the DEM source. Differences in quality between widely used and freely available DEMs (Nikolakopoulos et al., 2006; Suwandana et al., 2012) are significant for mapping specific soil classes.

The RF learner predicted all 13 possible clusters. Unlike J48 which is a single tree classifier, RF is an ensemble (forest) of bagged classification trees. The classification trees in RF are independent and the classification of samples does not depend upon previous trees in the ensemble (Kuhn and Johnson, 2013). J48 predicted mainly soil clusters which comprised of higher sampling points, while RF did not segregate.

In another study by Subburayalu and Slater (2013), RF outperformed J48 in prediction of minor soil series, in southeastern Ohio, and the authors suggested that RF has potential for digital soil mapping.

3.3. Mapping soil classes

The maps demonstrate that legacy soil class was the dominant predictor variable. There is a strong correspondence between the legacy soil map and the predicted soil map. Since most of the study area exhibits low slope gradient (1–5°), it was expected that higher resolution terrain data would be more effective in depicting relevant landscape differences in this environment, and thus terrain data would be stronger predictors. However, the RF learner produced identical prediction when used with SRTM (30 m spatial resolution) and WorldDEM (12 m spatial resolution) terrain predictors (Fig. 2 and Table 3). Terrain data with a spatial resolution finer than 12 m should also be tested with RF algorithm, when available for the study area, in order to discern the micro-topographical variation across the Valley.

Dissimilar outputs were obtained when SRTM and ASTER datasets were used on both RF and J48 algorithms, despite both DEMs having approximately 30 m spatial resolution. This agrees with some studies suggesting differences in quality between these two widely used and freely available DEMs in terms of vertical accuracy, and presence of artifacts and noise (Doumit, 2013; Forkuor and Maathuis, 2012; Nikolakopoulos et al., 2006; Suwandana et al., 2012).

4. Conclusion

This work used DSM methods to map numerically classified soil clusters of a portion of Kilombero Valley, Tanzania. In this study, terrain based predictors derived from 1 arc SRTM DEM results were similar to that of 12 m WorldDEM despite differences in resolution. It was also demonstrated that RF algorithm was less sensitive to the training set sampling intensity compared to J48.

We suggest the use of RF algorithm and SRTM DEM combination for soil class mapping for the remainder of the Kilombero Valley since RF was less sensitive to sampling intensity than J48 and a significantly lower cost of the SRTM DEM. This will help to generate spatial soil information which will enable decision-makers and farmers to make informed decisions for intensification of rice production in the Kilombero Valley. The RF and SRTM combination can be tested and applied in other areas which have relatively flat terrain like the Kilombero Valley.

Acknowledgement

This work builds on PhD research work by Boniface H.J. Massawe at the Ohio State University, USA. The authors are grateful to USAID's Innovative Agricultural Research Initiative (Cooperative Agreement 621-A-00-11-000090-00) (iAGRI) and Norman Borlaug Leadership Enhancement in Agriculture Program (2013 Borlaug LEAP Fellow, Spring) for funding this work.

References

Airbus Defence and Space, 2014. WorldDEM™: The New Standard of Global Elevation Models. Airbus DS/Infoterra2014 GmbH, Germany.
Bray, R.H., Kurtz, L.T., 1945. Determination of total, organic and available forms of phosphorus in soils. Soil Sci. 59, 39–45.
Bremner, J.M., Mulvaney, C.S., 1982. Total nitrogen. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2; Chemical and Mineralogical Properties, second ed. American Society of Agronomy, Madison, Wisconsin, pp. 595–624.
Brungard, C.B., Boettinger, J.L., 2012. Spatial prediction of biological soil crust classes; value added DSM from soil survey. In: Minasny, B., Malone, B.P., McBratney, A. (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney, pp. 57–60.
Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr., T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239–240, 68–83.
Bucheyeki, T.L., Shennkalwa, E., Kadadi, D., Lobulu, J., 2011. Assessment of rice production constraints and farmers preferences in Nzega and Igunga Districts. J. Adv. Dev. Res. 2 (1), 30–37.
Carré, F., Jacobson, M., 2009. Numerical classification of soil profile data using distance metrics. Geoderma 148, 336–345.
Carré, F., McBratney, A.B., 2005. Digital terron mapping. Geoderma 128, 340–353.
Cook, S.E., Jarvis, A., Gonzalez, J.P., 2008. A new global demand for digital soil information. In: Hartemink, A.A., McBratney, A., Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping With Limited Data. Springer, pp. 31–41.
Doumit, J.A., 2013. Comparison of SRTM DEM and ASTER GDEM Derived Digital Elevation Models with elevation points over the Lebanese territory. Lebanese J. Geogr. 27, 7–28.
FAO, 1961. The Rufiji Basin Tanganyika. FAO Exp. Techn. Ass. Progr. No. 1269 Vol. 7. Rome.
FAO, 2006. Guidelines for Soil Description. FAO, Rome.
Forkuor, G., Maathuis, B., 2012. Comparison of SRTM and ASTER derived Digital Elevation Models over two regions in Ghana - implications for hydrological and environmental modeling. In: Piacentini, T. (Ed.), Studies on Environmental and Applied Geomorphology. InTech, Rijeka, Croatia, pp. 219–240.
Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods. American Society of Agronomy, pp. 383–411.
Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1999. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298.
Hansen, M.K., Brown, D.J., Dennison, P.E., Graves, S.A., Bricklemyer, R.S., 2009. Inductively mapping expert-derived soil-landscape units within Dampo wetland catena using multispectral and topographic data. Geoderma 150 (1–2), 72–84.
Haring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schroder, B., 2012. Spatial disaggregation of complex soil map units: a decision tree based approach in Bavarian forest soils. Geoderma 185–186, 37–47.
He, Y., 2013. mColorConverter. Apple Inc., California, USA.
Jacobson, M., Carré, F., 2006. OSACA version 1.0., Land Management and Natural Hazards Unit, Institute for Environment and Sustainability. European Commission, Italy.
Kato, F., 2007. Development of a major rice cultivation area in the Kilombero Valley, Tanzania. Afr. Study Monogr. 36, 3–18 Suppl..
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.
Lindsay, J.B., 2014. The Whitebox geospatial analysis tools project and open-access GIS. Proceedings of the GIS Research UK 22nd Annual Conference16–18 April. The University of, Glasgow, UK.
Lindsay, W.L., Norvel, W.A., 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Sci. Soc. Am. J. 42, 421–428.
Malone, B., 2013. Use of R for Digital Soil Mapping. University of Sydney, Australia, Soil Security Laboratory.
McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McBratney, A., Minasny, B., Rossel, R.V., 2010. Numerical soil classification: a missed, but not a lost, opportunity. 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1–6 August, Brisbane, Australia.
McLean, E.O., 1986. Soil pH and lime requirement. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2; Chemical and Microbiological Properties, second ed. American Society of Agronomy, Madison pp, pp. 199–223.
Msanya, B.M., Magoggo, J.P., Otsuka, H., 2002. Development of soil surveys in Tanzania. Pedologist 46 (2), 79–88.
Muir, J.W., Hardie, H.G.M., Inkson, R.H.E., Anderson, A.J.B., 1970. The classification of soil profiles by traditional and numerical methods. Geordema 4, 81–90.
Munsell Color Company, 1992. Munsell Soil Color Charts, Munsell Color Co. Inc, Baltimore.
Nelson, D.W., Sommers, L.E., 1982. Total carbon, organic carbon and organic matter. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2; Chemical and Microbiological Properties, second ed. American Society of Agronomy, Madison, pp. 539–579.
Nikolakopoulos, K.G., Kamaratakis, E.K., Chrysoulakis, N., 2006. SRTM vs ASTER elevation products. Comparison for two regions in Crete. Greece. Int. J. Remote Sens. 27 (21), 4819–4838.
Odgers, N.P., McBratney, A.B., Minasny, B., 2011a. Bottom-up digital soil mapping I. Soil layer classes. Geoderma 163, 38–44.
Odgers, N.P., McBratney, A.B., Minasny, B., 2011b. Bottom-up digital soil mapping II. Soil series classes. Geoderma 163, 30–37.
Planchon, O., Darboux, F., 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. Catena 46, 159–176.
QGIS Development Team, 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org (page visited January 2015).
Rayner, J.H., 1966. Classification of soils by numerical methods. Eur. J. Soil Sci. 17, 79–92.
Rhoades, J.D., 1982. Soluble salts. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2; Chemical and Microbiological Properties, second ed. American Society of Agronomy, Madison, pp. 167–179.
Rizzo, R., Demattê, J.A.M., Terra, F.S., 2014. Using numerical classification of profiles based on VIS-NIR spectra to distinguish soils from the Piracicaba region. Brazil. Rev. Bras. Ciênc. Solo 38, 372–385.
Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil adjusted vegetation indices. Remote Sens. Environ. 55, 95–107.
Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progr. Phys. Geogr. 27, 171–197.
Soil Survey Staff, 2014. Keys to Soil Taxonomy. 12th ed. USDA, Natural Resource Conservation Service, USA.
Stum, A.K., Boettinger, J.L., White, M.A., Ramsey, R.D., 2010. Random forests applied as a soil spatial predictive model in arid Utah. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer, Dordrecht, pp. 179–190.
Subburayalu, S.K., Slater, B.K., 2013. Soil series mapping by knowledge discovery from an Ohio county soil map. Soil Sci. Soc. Am. J. 77, 1254–1268.

Subburayalu, S.K., Jenhani, I., Slater, B.K., 2014. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. Geoderma 213, 334–345.
Suwandana, E., Kawamura, K., Sakuno, Y., Kustiyanto, E., Raharjo, B., 2012. Evaluation of ASTER GDEM2 in comparison with GDEM1, SRTM DEM and Topographic-Map-Derived DEM using Inundation Area Analysis and RTK-dGPS Data. Remote Sens. 4, 2419–2431.
Thomas, G.W., 1982. Exchangeable cations. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2; Chemical and Microbiological Properties, second ed. American Society of Agronomy, Madison pp, pp. 595–624.
TIC, 2013. Tanzania Investment Guide 2013–14. Tanzania Investment Center, Dar es Salaam.
Vågen, T.-G., Winowiecki, L.A., Abegaz, A., Hadgu, K.M., 2013. Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia. Remote Sens. Environ. 134, 266–275.
Vågen, T.-G., Winowiecki, L.A., Tondoh, J.E., Desta, L.T., Gumbricht, T., 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. Geoderma 263, 216–225.
Watanabe, F.S., Olsen, S.R., 1965. Test of ascorbic acid method for determining phosphorus in water and NaHCO3 extracts from soil. Soil Sci. Soc. Am. 29, 677–678.
Winowiecki, L., Vågen, T.-G., Massawe, B., Jelinski, N.A., Lyamchai, C., Sayula, G., Msoka, E., 2016a. Landscape-scale variability of soil health indicators: effects of cultivation on soil organic carbon in the Usambara Mountains of Tanzania. Nutr. Cycl. Agroecosystems 105, 263–274.
Winowiecki, L., Vågen, T.-G., Huising, J., 2016b. Effects of land cover on ecosystem services in Tanzania: a spatial assessment of soil organic carbon. Geoderma 263, 274–283.
Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington.