Szent István University

Ph.D. School of Environmental Sciences

INNOVATIVE APPROACHES OF PREDICTING SOIL PROPERTIES AND SOIL CLASSES IN THE EASTERN SLOPES OF MT. KENYA

Thesis of Ph.D. dissertation

Evans Mutuma

Gödöllő

2017

Ph.D. School: Ph.D. School of Environmental Sciences

Discipline: Environmental sciences

Head of School: Csákiné Dr. Michéli Erika, D.Sc.
Professor
Szent István University
Faculty of Agricultural and Environmental Sciences
Institute of Environmental Sciences
Department of Soil Science and Agrochemistry

Scientific supervisor Csákiné Dr. Michéli Erika
Professor
Szent István University
Faculty of Agricultural and Environmental Sciences
Institute of Environmental Sciences
Department of Soil Science and Agrochemistry

Approval of Head of School **Approval of Scientific supervisor**

1. Introduction and objectives

The expected growth of population and the need of more food make the knowledge of soil properties essential to secure the successes of agricultural production on currently available land. Despite Kenya's economy being agricultural based, most of the existing soil inventories do not capture dynamic soil properties that are sensitive to management (Vagen *et al.*, 2013). The major existing soil inventories for Kenya like the KENSOTER database is derived from the legacy Exploratory Soil Map of Kenya scale 1:1 M (Ministry of Agriculture, Kenya Soil Survey, 1980). The limitation to most of these soil databases is the scale at which data is presented, lack of harmonized methodologies of data collection and laboratory analysis, that affect the accuracy and therefore fail to provide adequate information for soil management at farm or watershed scale. The high cost of conventional methods of soil survey and laboratory analysis (Shepherd & Walsh, 2002) may partly explain why little has been done to update the existing legacy soil inventories in Kenya.

To provide soil data for updating the existing soil inventories, rapid methods to quantify soil properties at large geographical coverage urgently need to be adopted. Over the past 30 years, visible, near and mid-infrared (Vis-NIR-MIR) reflectance spectroscopy has proved to be a fast, cost effective, environmental-friendly, non destructive, reproducible, and repeatable analytical technique of soil properties determination (Nocita *et al.,* 2015). A single spectrum may contain comprehensive information on various soil properties, and can be used to predict these simultaneously. Lack of soil monitoring networks makes it impossible to recommend and prioritise site specific soil management practices increasing vulnerability of the soils to further degradation in this study area. Soil classification help to identify differences among and between soils and their environment is an important tool to inform soil management decisions.

Based on the gaps identified with regard to the existing soil inventories in the study area, I have identified the following objectives:

1. To develop an optimized soil sampling scheme that preserves the natural distribution of soil forming factors in the study area, in the eastern slopes of Mt. Kenya.
2. To develop an ensemble model for predicting of important soil properties (i.e soil organic carbon, base cations, pH, aluminium and particle size distribution).
3. To demonstrate the usefulness of the derived database for mapping soil properties for the study area.
4. To classify the visited soils and validate the soil types in the KENSOTER soil units of the study area.
5. To compare differences of soil properties in different WRB Reference soil groups and the implications of applications for management purposes in the study area.

2. Materials and methods

2.1 Study area.

Soil sampling was conducted in Mt. Kenya region covering an area of 1200 km^2 within latitudes 37^0 36'E and 38^0 0' E and longitudes 0^0 6' N and 0^0 18' S (Figure 4.). The population density estimate according to Kenya population and housing census basic report of 31st August 2010 was 424 persons/km2. The major land use is rainfed agriculture. The altitude range was 700 m to 2000 m. The agro-climatic zone is humid in high altitudes and semi-arid in the lower altitudes (Jaetzold *et al.*, 2007). Amount of rainfall is 1500 mm in upper humid zones and 600 mm in the lower semi-arid zones. The annual average temperature is 10 ^{0}C to 35 ^{0}C. The geology is mainly volcanic rock and ash and some old metamorphic rocks (Schoeman, 1952). The volcanic rocks in the area are related to the Rift Valley development during the Pliocene time and dated from 3.5 to 2 million years and parasitic cones during the Pleistocene time. The the dominating WRB (IUSS Working Group WRB, 2015) Reference soil groups (RSG) according to KENSOTER map and database (Dijkshoorn, 2007) are: Nitisols, Ferrasols, Regosols, Vertisols and Phaeozems.

2.2 Soil sampling design

To define the soil sampling locations, Conditional Latin Hypercube Sampling (CLHS) was performed. The reason of using CLHS in sampling site selection were the foreseen constraints (inaccessibility due to poor weather roads, very steep slopes, possibility of having sampling locations coinciding with water bodies, national parks or built environment) and the need to reduce the sample size yet cover a wide geographical area with limited budget.

2.2.1 Assembly of variables for input into CHLS algorithm

Environmental variable layers and operational cost layer were generated as input variables for the CHLS algorithm. Good expressions of soil forming factors in remote sensing data have been reported (Dobos *et al.,* 2000). Jenny's (1941) state equation for soil formation: S=f (cl, o, r, p, and t) guided the choice of environmental variables.

Organisms (O) were represented using Normalized Difference Vegetation index (NDVI) derived from Landsat 8 satellite imagery with a resolution of 30 m for dry season from row/path 168/60 from 15 September 2014. NDVI is calculated as shown in equation 2.

$$NDVI = \frac{NIR_{band} - RED}{NIR_{band} + RED} \tag{1}$$

*Where NIR =Band 5, wavelength 0.64-0.67 μm and RED=Band 4 wavelength 0.85-0.88 μm and a resolution of 30*30m.*

Relief (*r*) was represented by terrain derivatives (slope and topographic wetness Index). These were calculated from Digital Elevation Model (DEM), Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) with a resolution of 30 m. SAGA GIS 2.0.6 was used to generate these terrain derivatives.

The Topographic Wetness Index (TWI) is a steady state wetness index and is a function of slope and the upstream contributing area and is calculated as below.

$$TWI = ln\left({}^{A}/_{\tan\beta}\right) \qquad (2)$$

Where A is the specific catchment area expressed as m^2 per unit width orthogonal to the flow direction, and β is the slope angle in radians.

Local slope was generated from the DEM using SAGA GIS >Spatial Analyst tools > Surfaces > slope.

The 'cost of reach layer' was generating using *r.walk* in GRASS GIS (Neteler *et al.,* 2012). Similar approach by Roudier *et al.* (2012) in Australia and Mulder *et al.* (2013) in Morocco reduced the working cost of soil survey significantly.

Dominant soils of the polygons of the KENSOTER units (Dijkshoorn, 2007) were used as categorical data to ensure sampling was done in every dominant soil type. R 'CLHS' algorithm (R Core Team, 2013), and Quantum GIS processing were used to design the sampling frame.

Box plots were used for validation of the sampling scheme based on natural and sampled distribution of environmental variables. The spread of the sampling points in the study area (Figure 1.) satisfied the initial objective to preserve the natural distribution of environmental variables that were used as input layers for the CLHS algorithm. After evaluation of the sampling scheme I embarked on field work.

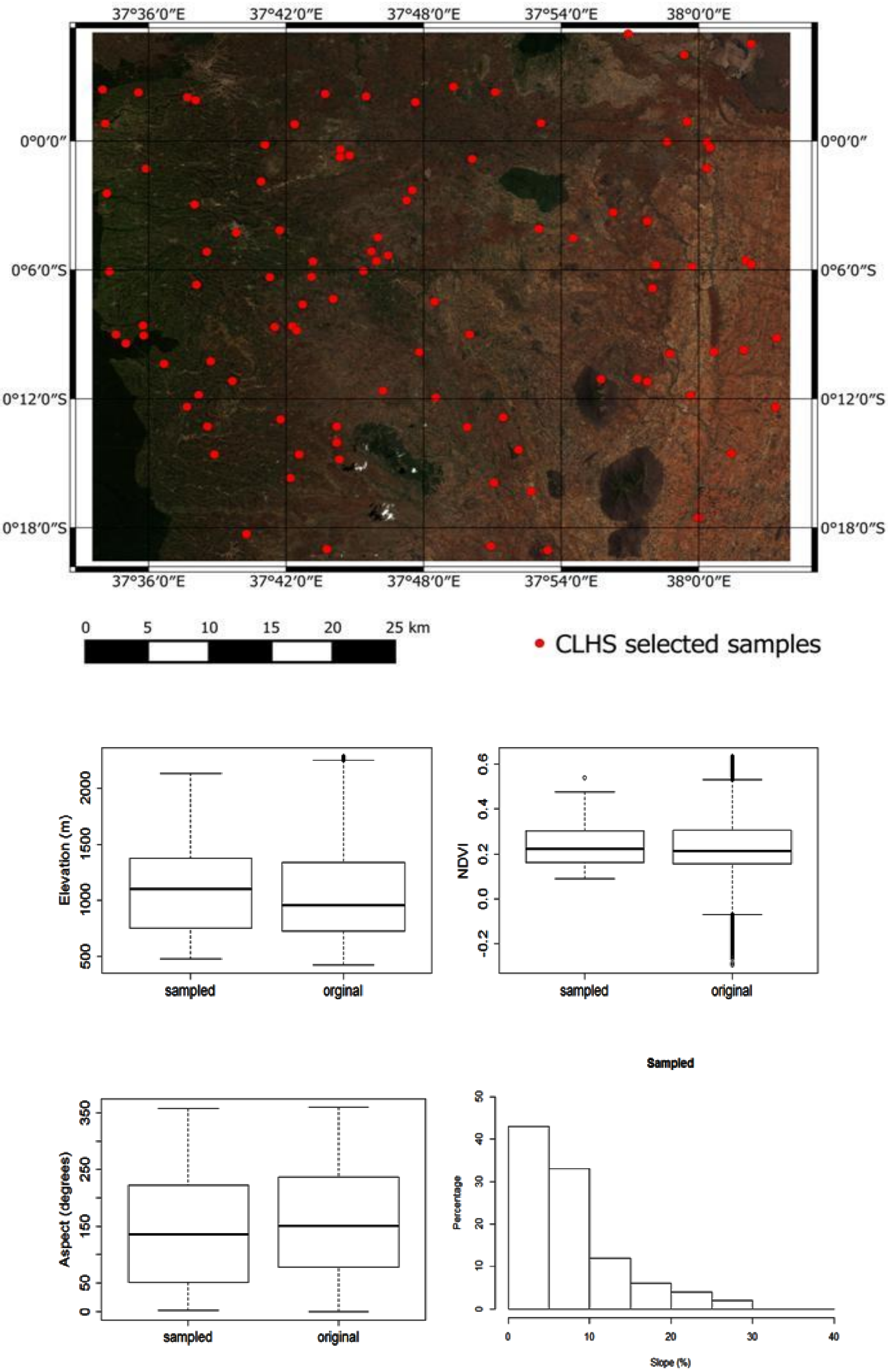

Figure 1. Comparison of the statistical distributions of the environmental covariates in the original GIS layers in selected sampling locations and sampled slope.

2.3 Field work and soil description

The developed sampling plan guided the locations to be visited. The soil sampling campaign commenced on 5th December 2015 and ended on 15th Jan 2016. FAO Soil Description Guideline (FAO, 2006) and WRB guidelines (IUSS Working Group WRB, 2015) guided soil field soil sampling and classification respectively. A total of 77 sampling sites were visited, out of this total, 28 were open profiles and 49 were augured profiles. The 28 open profiles were described based on genetic horizons. For each genetic horizon: depth, pH, Munsell color, $CaCO_3$ and structure for each master horizon, also and presence or absence of coatings (i.e clay, iron, and manganese), volcanic glass and pressure surfaces were recorded. Similar procedures were carried for the 49 augured profiles except for the depths of sampling which was based on 0-20 cm, 20-50 cm, 50-100 cm intervals. All together a total of 269 soil samples were collected.

2.4 Laboratory soil measurements

Mid infrared measurements and conventional measurements using wet chemistry methods were performed on the soil samples. The pre-processing of the total 269 samples involved air drying, crushing and sieving using a 2 mm sieve and further crushing to < 100 μm.

2.4.1 Mid-infrared (MIR) spectral-reflectance measurements

MIR soil spectra measurements for the 269 samples were performed using a Fourier-transform MIR spectrometer-FTIR, Tensor 27 with Liquid N_2–cooled HgCdTe as the detector. The first derivative of the reflectance spectra was computed based on Savitzky–Golay smoothing filters (Wand & Ripley, 2008). Figure 2. shows the absorbance spectra for all the 269 soil samples before (noisy spectra) and after pre-processing to 4000-400 cm^{-1}.

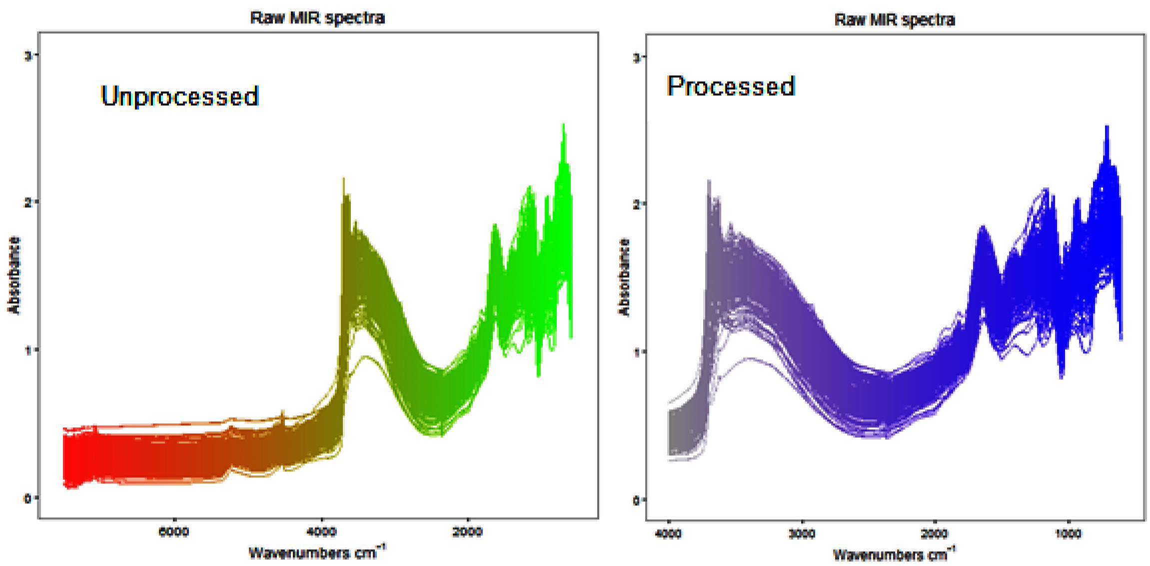

Figure 2. Unprocessed and processed MIR absorbance spectra to 4000-400 cm^{-1}

2.4.2 Calibration sample selection

Kennard-Stone algorithm (Kennard & Stone, 1969) was used for the calibration and prediction sample selection. The results of sample selection (Figure 3.) could explain 75.6% of the variations and the next step was to perform laboratory analysis of the calibration sample set.

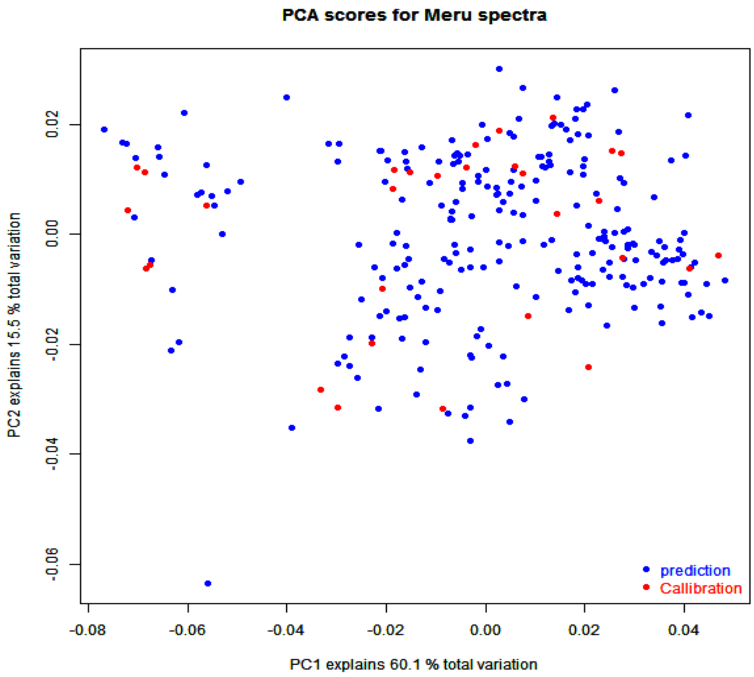

Figure 3. PCA for sample selection explains 75.6% of variations.

2.4.3 Soil analysis for the calibration samples

The samples were analysed for the following parameters with the respective methods

- pH –potentiometrically measured in the supernatant suspension of a 1:2.5 soil: liquid (H_2O) mixture (Van Reeuwijk, 2002).
- Cation Exchange Capacity (CEC)-Ammonium acetate method (Van Reeuwijk, 2002)
- Na, K, Ca, Mg and Al following Mehlich 3 extraction (Mehlich, 1984).
- Soil organic Carbon (SOC) and total nitrogen (TN) following thermal oxidation (Skjemstad & Baldock, 2008)
- Particle size distribution (PSD) using Laser Diffraction Particle Size Analyzer (LDPSA) and calibrated using pipette method (Van Reeuwijk, 2002).
- Free iron ($Fe_{dithionite}$) was analysed using the Holmgren procedure (Van Reeuwijk, 2002)
- Active iron ($Fe_{oxalate}$) using acid ammonium oxalate solution (Van Reeuwijk, 2002). The ($Fe_{dithionite}$) and ($Fe_{oxalate}$) were measured only for 5 selected samples from those that qualified the nitic horizon morphological characteristics.

- X-Ray Diffraction (XRD) was performed on 10 samples based on soil color and at the depth interval of 20-50 cm.

The ($Fe_{dithionite}$), ($Fe_{oxalate}$) and X-Ray Diffraction (XRD) data were not used for the calibration of MIR spectra but to give more details on soil samples to support classification. The results of sampling scheme, the fieldwork, and MIR spectroscopic analysis provided input data for mapping and soil classification.

2.5 Mapping of soil properties

Estimation of soil properties at the unvisited locations was performed using ordinary Kriging. The choice of this mapping method was based on my data as follows:

1. The selected soil properties were not significantly correlated with the input explanatory variables (Table 1.).

Table 1. Explanatory variables were not significantly correlated with test soil variables

Variables	TWI	SLOPE	NDVI	TN (%)	SOC (%)	P(mg/ kg)	pH H_2O
TWI	**1**						
SLOPE	**-0.492**	**1**					
NDVI	-0.117	**0.240**	**1**				
TN (%)	-0.024	0.110	-0.001	**1**			
SOC (%)	-0.021	0.117	-0.011	**0.980**	**1**		
P(mg/ kg)	-0.069	0.150	-0.107	**0.202**	**0.220**	**1**	
pH H_2O	0.060	-0.189	-0.308	**-0.413**	**-0.375**	**-0.239**	**1**

Values in bold are different from 0 with a significance level alpha=0.05

2. The variables showed spatial autocorrelation using nugget to sill ratio (Table 2.).

Table 2. Semivariogram parameters for the selected spherical model show moderate spatial dependence.

Soil property	Model	Nugget	Sill	Nugget/sill ratio	Range	Remarks
SOC (%)	Spherical	0.098	0.21	0.46	13576	M
TN (%)	Spherical	0. 556	0.9	0.61	13737	M
pH H_2O	Spherical	0.226	0.72	0.31	2142	M
Ext P (mg/kg)	Spherical	2.31	3.7	0.62	12590	M

M= Moderate spatial dependence, N= number of samples

2.6 Soil Classification

The World Reference Base of soil resources (WRB) classification system (IUSS Working Group WRB, 2015) was applied to characterize and classify soils of the visited sites. The WRB is based on a diagnostic approach defined in terms of diagnostic horizons, diagnostic properties and materials, that are measurable to the greatest extent possible and observable in the field. On the highest level 32 reference soil groups (RSG) are defined by the classification key based on the presence/absence of combination of the diagnostics. Additional qualifiers are giving are added to the RSG name to provide further information on important soil properties.

2.7 Data analysis

Kennard- Stone algorithm (Kennard & Stone 1964) and Principal Component Analysis were used in the selection of calibration and prediction sample sets. Random Forest (RF) Regression was used for the calibration of spectra data using the laboratory measurements of the calibration sample set. The utility of RF for regression and classification is described by Gislason *et al*. (2006). Cross validation was performed to evaluate the model performance based on the variations between the observed and the predicted values of soil properties. This was determined using coefficient of determination (r^2), the Mean Error (ME), Root Mean Square Error (RMSE) and the Standardized Root Mean Square Error (SRMSE). The best model is one that has ME nearest zero, the smallest RMSE and SRME nearest to 1. For good predictive model the RMSE values should be low (<0.3). Exploratory data analysis for the quantified soil properties was performed using descriptive statistics and data normality testing using Shapiro-Wilk test at 5% significant level. Discriminant analysis (Carroll *et al.* 2006) was used to evaluate contribution of each soil property in the classification of reference soil groups. Multiple pairwise comparisons using Dunn's procedure (Dunn's, 1964) was used to compare variability of soil properties in different WRB reference soil groups.

The schematic representation summarizing the methodology is given below (Figure 4.).

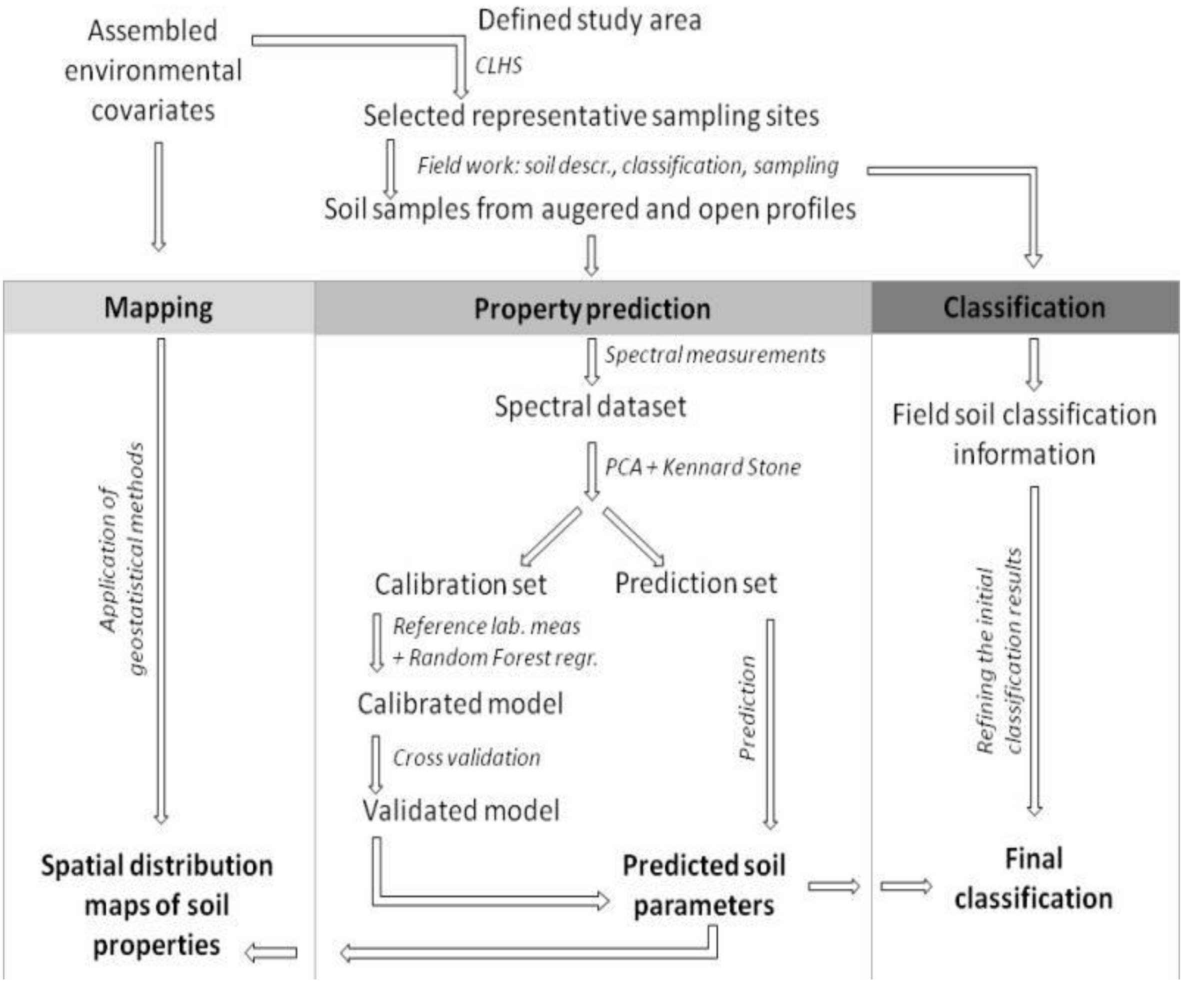

Figure 4. Methodology flowchart summarising this research work

3. Results and discussion

3.1 Accuracy of soil property predictions

The accuracy of the soil properties predictions is presented (Table 3.). The model performance was excellent when r^2>0.90, good when 0.81<r^2<0.90, moderately successful when 0.66<r^2<0.80 and unsuccessful when 0.50<r^2<0.65. This classification was adopted from Saeys *et al.* (2005).

Table 3: Validation results of soil properties predictions

Soil property	Coefficient of determination (r^2)	Root mean Square error (RMSE)
SOC (%)	0.78	1.64
pH H_2O	0.88	0.63
Exc Na cmol/kg	0.93	1.45
Exc Mg cmol/kg	0.88	1.60
Exc K cmol/kg	0.92	2.32
Exc Ca cmol/kg	0.92	189
Exc Al cmol/kg	0.85	71.25
Ext P mg/kg	0.78	0.51
Clay (%)	0.59	9.9
Silt (%)	0.63	7.3

Moderately successful predictions were achieved for SOC % at r^2=0.78 and RMSE=1.64 (Table 3.). Comparable accuracy (r^2=0.77 and RMSE=1.64) for SOC has been reported by Terhoeven-Urselmans *et al.* (2010). But our coefficient of determination (r^2) was better. This can be associated to the use of different calibration statistics (Terhoeven-Urselmans *et al.* (2010) used partial least squares regression) and I used random forest regression (RF) as the multivariate statistics to calibrate the soil MIR spectra. Higher accuracy of SOC prediction (r^2 = 0.96) was achieved by McDowell *et al.* (2012), this may be attributed to high and wide range of SOC (%) of the calibration set (0.24 to 55.29%) compared to narrower SOC range (0.56 to 10.83%) in my soil samples. The accuracy for the calibration of pH was good (r^2 = 0.88 and RMSE = 0.48) and better in terms of r^2 and RMSE compared to those achieved by Terhoeven-Urselmans *et al.* (2010) (r^2 = 0.81, RMSE = 0.63) and Shepherd & Walsh (2002) (r^2 = 0.83, RMSE = 0.54) while analysing a spectral library of soils from Africa. Calibrations for Mehlich extraction of Na, Mg, K and Ca were excellent (r^2=0.93 and RMSE=1.45; r^2=0.88 and RMSE=1.60; r^2=0.92 and RMSE=2.32 and r^2=0.92 and RMSE=189.16 respectively. The prediction of Ca was better than that of Sila *et al.* (2016) who reported r^2=0.91 for Ca. Again this could be associated with the efficiency of RF, a multivariate statistic that was used for the calibration of soil MIR spectra. The high RMSE values especially for Ca can be attributed to under prediction by MIR due to low calcium content in all my sampling locations. Calibration of Mehlich extraction of Al was good (r^2=0.85 and RMSE=0.51).Calibration of P was satisfactory (r^2=0.78 and RMSE=71.25). Calibration for clay (r^2 = 0.59 and RMSE = 9.9); Silt (r^2 = 0.63 and RMSE = 7.3) were low compared to other soil properties. The low calibration accuracy for sand (r^2= 0.30 and RMSE=5) was because the soil samples were mainly clay dominated due to extreme weathering resulting in poor MIR absorbance. The percentage sand was calculated from of the cumulative percentages of clay and silt.

3.2 Spatial interpolation of soil properties

SOC, TN, pH and P were selected for spatial interpolation because they present important agronomic soil properties that can be restored through good soil management practices. Moderate spatial dependence was exhibited by all the soil properties (Table 2.). Spherical model and OK resulted in spatial distribution maps (Figure 5.).

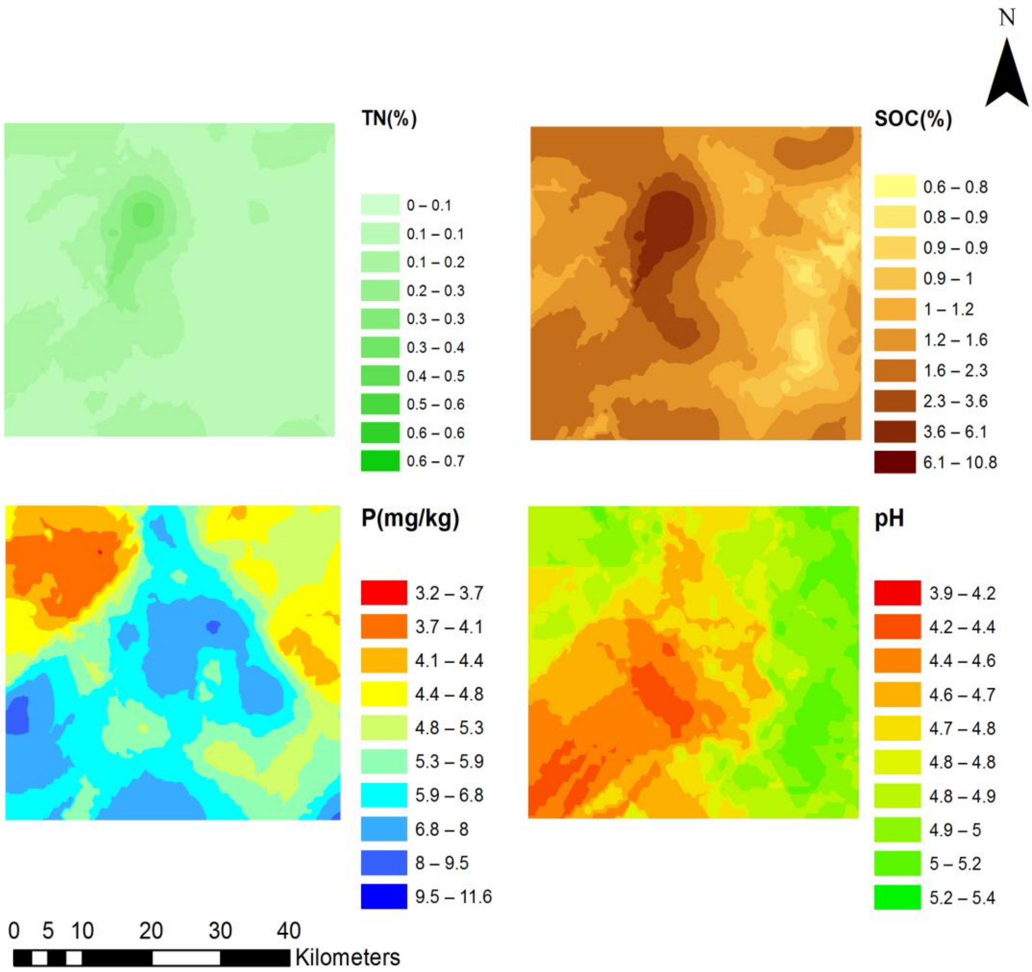

Figure 5. Maps of predicted spatial distribution of soil properties

The performance of spatial prediction for TN and pH was satisfactory based on the cross validation results (Table 4.). The low accuracy SOC (RMSE=0.81) and P (RMSE=1.1) compared to TN and pH predictions. SOC was mainly influenced by land management and this input layer into the Conditional Latin Hypercube Sampling Algorithm was missing due to lack of data. P was influenced by soil type and land management as well. Inventories for land management practices were not available and soil type data layers were only available in the KENSOTER database with limitations highlighted earlier in the introduction section.

Table 4. Cross validation results of the prediction model.

Soil property	**ME**	**RMSE**	**SRMSE**
SOC (%)	0.001	0.81	0.76
TN (%)	0.0002	0.06	0.96
pH	-0.0006	0.13	0.98
P (mg/kg)	0.006	1.11	0.77

ME=Mean error, RMSE=Root Mean Square Error, SRMSE=Standardized Root Mean Square Error.

Important trends are visible from the spatial distribution of soil properties. The region with high SOC contents coincides with the region with very low pH values. The low pH values are a result of intense leaching caused by heavy rainfall in this humid high altitude part of the study area. The soil type in humid, high altitude is dominated by Umbrisols (by default Umbrisols are Dystric and have high organic matter). Low pH slows down the activity of microbes responsible for the mineralization of SOC resulting in its stabilization and accumulation. Litter from the tea plantations in this part of the study area provided input material for humification.

The trend for TN is similar to that of SOC because this is mainly organic nitrogen from soil organic matter. Therefore areas with high TN coincided with areas with high SOC. Fertilization using NPK resulted in high nitrogen in tea plantation zones.

The trends for P are determined by soil types, rainfall intensity and management. High P values in the middle zone of the study area can be associated with tea plantations due to addition of NPK fertilizers, mineralization of SOC resulting in addition of organic phosphorus and less P fixation because of the nature of the dominant soil types (i.e Umbrisols and Nitisols). Towards the far east of the study P increases as the pH increase due to less leaching occasioned by decreasing rainfall amounts towards the east.

3.3 Results of soil classification.

My classification results at the RSG level were not fully matching the soil types of the associations of the relevant soil mapping units of the KENSOTER map units (Table 5). The KENSOTER humic Nitisols polygon had 28 sampled profiles. Matching associations with classified profiles in this polygon were identified. The 10 profiles classified as Andosols matched the humic Andosol association and 1 profile classified as Cambisol matched the humic Cambisol association in the humic Nitisol polygon. Umbrisols were represented by 10 profiles, Alisols 5 profiles and Regosols by 1 profile.

Table 5. Matrix representation of soil classes in KENSOTER polygons.

KENSOTER Polygons	**The classification of the profiles in the map unit (polygon)**								**Profiles**
	AN	**UM**	**Al**	**RG**	**CM**	**LP**	**PH**	**PT**	
Humic Nitisols (NTu70:ANu16:CMu14)	10	10	5	1	2	0	0	0	28
Rhodic Nitisols (NTr60:NTu30:FRr10)	0	19	2	0	5	1	0	0	27
Luvic Phaeozems (PHl40:NTh30:CMx30)	0	8	0	0	3	0	5	0	16
Eutric Vertisols (VRe70:CMe30)	0	0	0	0	0	0	1	1	2
Lithic Leptosols (LPq60:PHl40)	0	1	0	0	0	0	0	0	1
Rhodic Ferrasols (FRr90:LXh5:ACh5)	0	0	1	0	1	0	0	0	2
Total RSGs	10	38	8	1	11	1	6	1	76
Share of the RSG in the study area (%)	13.3	50	10.5	1.3	14.4	1.3	7,9	1,3	100

Where PH=Phaeozems, LP=Leptosols, AN=Andosols, UM=Umbrisols, CM=Cambisols, AL=Alisols, PT=Plinthosols, RG= Regosols, HS=Histosols, NT=Nitisols, VR=Vertisols, FR=Ferrasols, CMu=Humic Cambisols, NTu=Humic Nitisols, NTr=Rhodic Nitisols, FRr=Rhodic Ferralsols, VRe=Eutric Vertisols, LXh=Humic Lixisols, Ach=Humic Acrisols, LPq=Lithic Leptosols.

The Rhodic Nitisol polygon had 27 profiles. No profile was classified as a Nitisol even with dense sampling and the high chance (90%) based on Nitisol association ratio in this polygon. Out of the 27, 19 profiles were classified as Umbrisols, 2 as Alisols, 5 Cambisols and 1 Leptosol. Matching associations with 5 profiles classified as Phaeozems were identified in the Luvic Phaeozem polygon. Umbrisols were represented by 8 profiles and Cambisols by 3 profiles out of the total 16 profiles. One profile was classified as Phaoezems and another as Plinthosols in the Eutric Vertisols polygon. Only one profile was visited and classified as Umbrisols in the Lithic Leptosol polygon. Two profiles were visited in the Rhodic Ferralsols polygon and classified as Alisols and Cambisols.

Based on the KENSOTER map units, this is a Nitisol rich region (Figure 6.). Soil profiles of my study that satisfied the morphological characteristics of ‘nitic horizon’ failed the silt/clay ratio of <0.4 diagnostic criterion in the WRB classification system. The higher silt ratio might be related to rejuvenation of the Nitisols during Pleistocene eruption in the Mt. Kenya region. Similar suggestions were made by De Wispelaere *et al.* (2015) based on observations in Ethiopia. Regardless of the cause, the current criterion makes soil with nitic horizon morphology excluded from the Nitisols, and also creates confusion in legacy databases. Based on these conclusion suggestions have been made to the IUSS WRB Working Group to skip this criteria from the nitic horizon diagnostic criteria.

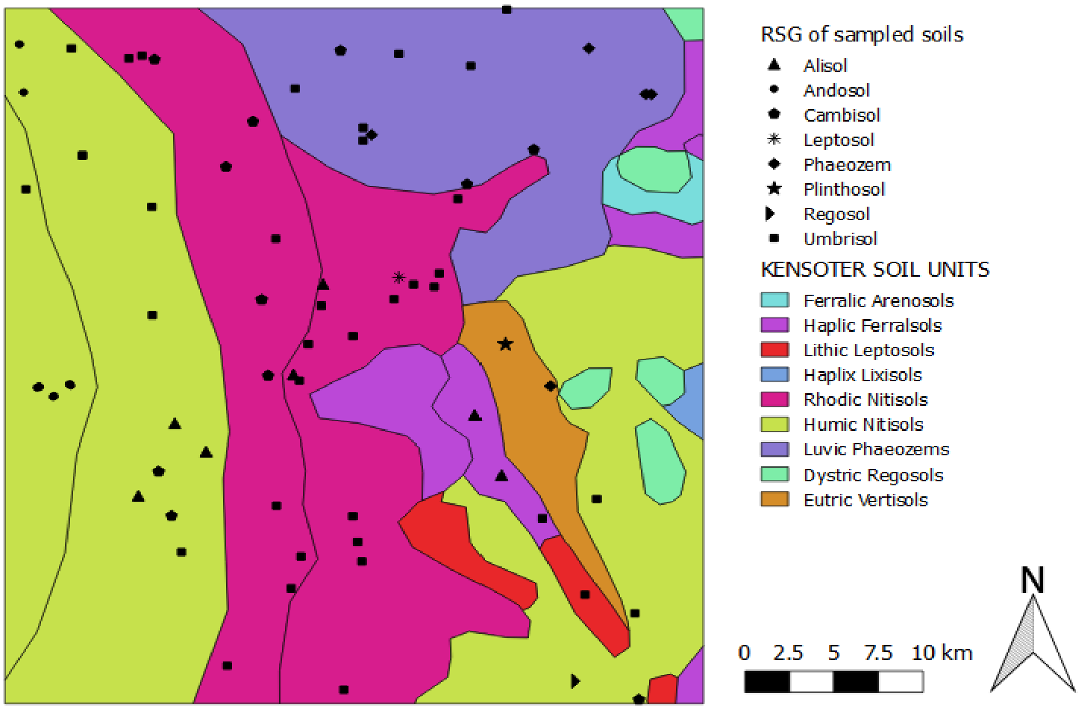

Figure 6. KENSOTER soil units and classified RSG of sampled soils

3.4 Inferences for management from predicted soil properties

The results of this study show the soils were acidic with pH range of 3.9 - 5.3 (Table 6.). The causes and effects of acidification on soils are well explained by Goulding (2016). The soil pH influence the bioavailability of plant nutrients and so indirectly affect crop plant growth. The range for P (Table 6.) was 3.15 mg/kg and 11.63 mg/kg, where 43.7% of total samples did not meet the threshold of 5mg/kg (Okalebo *et al.*, 2006). The high phosphate retention may be due to the existence of andic material occasioned by presence of Andosol reference soil groups in higher altitudes. The minimum and maximum values were 3.6 g/kg and 57.2 g/kg (Table 6.) and 93.07% did not meet the threshold of 20 g/kg using the assessment of Loveland & Webb (2003). TN results show a minimum 0.4 g/kg and Max 6.8 g/kg but only 15% of the total sampling area met the threshold of 2.0 g/kg based on suggestions from Ndakidemi & Semoka (2006). The textural classes were mainly clay, silt clay and silty clay loam. The ratio of sand in all the observations was less than 20 % while clay and silt were 36-58 % and 32-52% respectively.

Table 6. Statistics of predicted soil properties

Variable	**Min**	**Max**	**Mean**	**STD**	**CV**	**Skewness**	**Shapiro-Wilk test**
TN (g/kg)	0.41	6.8	1.64	1.28	0.78	1.88	< 0.0001
SOC (g/kg)	3.60	57.2	10.34	7.57	0.73	3.49	< 0.0001
Clay (%)	36.37	58.71	46.26	3.04	0.07	0.43	< 0.0001
Silt (%)	32.64	52.42	42.35	2.69	0.06	0.18	0.002
Sand (%)	7.66	16.17	12.09	1.54	0.12	0.34	0.002
P (mg/kg)	3.15	11.63	5.82	1.88	0.32	0.87	< 0.0001
K (cmol/kg)	1.19	1.88	1.59	0.14	0.09	-0.56	0.000
Ca (cmol/kg)	5.32	18.09	8.03	1.44	0.18	1.98	< 0.0001
Mg (cmol/kg)	2.70	5.89	4.34	0.51	0.117	-0.265	0.174
Al (cmol/kg)	3.17	5.39	4.08	0.51	0.12	0.27	< 0.0001
Al (%)	42.9	73.1	55.0	6.87	0.12	0.26	< 0.0001
Na (cmol/kg)	0.07	0.17	0.13	0.02	0.13	-0.65	< 0.0001
pH H_2O	3.9	5.3	4.7	0.2	0.05	-0.5	

3.5 Summary of the suggested management options

Soil erosion control and addition of soil organic matter into the soil

Colluvic qualifier (Figure 22.) is an indication of movement of soil material through erosion. Measures of erosion control like terraces, increasing cover crops, mulching, improving the soil structure through addition of soil organic matter need to be emphasized in this study area. Farm yard manure (FYM) is a good source of organic matter. A major challenge remains as these FYM are usually insufficiently available to provide the nutrients needed to maintain agricultural production at a desirable level. Composting and use of biochar technologies may also increase availability of OMs for soil amendments.

Liming and use of P fertilizers

The Dystric qualifier (Figure 22.) show that the soils are mainly acidic with base saturation <50%. Andic and Vitric qualifiers (refer Appendix1 and Appendix 2) which were present in some of the sampling sites are good agents of P fixation. Application of agricultural lime containing Ca and/or Mg compounds increase Ca^{2+} and/or Mg^{2+} ions and reduces Al^{3+}, H^{+}, Mn^{4+}, and Fe^{3+} ions. This leads to increase in soil pH and available P due to reduction in P fixation (Connor *et al.*, 2011). Liming also increases earthworm activity and therefore macro porosity (Bolan *et al.,* 2003). Bennett *et al.,* (2014) found that lime applied at 5t/ha was improving aggregate stability, hydraulic conductivity, vegetation cover, total carbon and Nitrogen and respiration 12 years after application. However, high cost and inaccessibility of liming and fertilization inputs have been reported as key constraints in Kenya (Okalebo *et al.,* 2006). Due to these challenges, alternative management options need be explored to provide a bigger matrix of options to the famers.

Use of acid tolerant crops

Planting crops that are tolerant to low pH and Al toxicity like tea and sorghum. Although the use of tolerant plant germplasm does not reverse soil acidity conditions, it minimizes the problems experienced by farmers, especially those who do not use lime.

4. CONCLUSION

Combining Conditional Latin Hypercube Sampling, Mid Infrared spectroscopy, and Random Forest Regression (RF), soil properties were satisfactorily predicted. The results of the linear regression show a strong relationship between the measured values and the predicted. Geostatistical analysis of my data revealed spherical model as the best fit for the semivariogram. Cross validations result show that TN and pH had excellent results compared to SOC and P. The low accuracy for SOC and P can be explained is because important explanatory variables like land management were missing during the prediction procedure due to lack of data. Also the input variable KENSOTER database that represented soil types had its share of limitations that were discussed in the introduction and affected the accuracy of the model. The maps of the predicted spatial distribution of the soil properties are a good demonstration of how a rapid and an accurate methodology can provide data to update maps and legacy inventories in the study area. The scale for these maps can inform land management at watershed and farm level (30m*30m resolution). Moreover, they form a good monitoring network that has not been there before for this study area.

The second part of my research involved soil classification using the predicted soil properties and morphological field observations. The results show that the difficulties of Nitisols classification are primarily caused by the failure of the profiles data to qualify the nitic diagnostic criterion of silt to clay ratio of < 0.4. Modification of the criteria was suggested by the author during the European Geosciences Union 2017 general assembly in Vienna on 28th April 2014. This was through an oral presentation that I delivered '*Classification problems of Mount Kenya soils with reference to Nitisols'* in the presence of IUSS WRB working Group. The diagnostic elements, principal and supplementary qualifiers made it possible not to loose soil information even when the reference soil groups were different. The building blocks of the system capture specific information important for soil management.

5. Key scientific findings and important output of this research

1. In my doctoral research I have applied Conditional Latin Hypercube Sampling for optimising the sample size; Mid Infrared Spectroscopy for rapid scanning of soil samples to generate a soil spectra library and Random Forest Regression to calibrate the soil MIR spectra using reference soil properties. This combination was 1st time and successfully used in the eastern slopes of Mt. Kenya to generate soil properties data.

2. Based on the results of the Conditional Latin Hypercube Sampling I have developed the first spectral library representing the spectral variability of the Eastern Mt. Kenya soils. This database contributes to future soil sampling campaigns, and enables the efficient soil property prediction.

3. The spatial density of available soils data in my study area is very low. The 77 georeferenced and fully described and analyzed profiles contribute to the understating of soil resources of the area. The investigated representative locations can serve for future monitoring of the determined soil parameters. Beside the point data I have compiled the spatial prediction map of the properties, which can support farmers in making soil management decisions in the area.

4. In earlier surveys extended areas were mapped as Nitisols on Mt. Kenya. In my study I have not classified any of my profiles as Nitisols, although they satisfied the important morphological criteria of Nitisols, the silt/clay ratio < 0.4 diagnostic criterion for nitic horizon (introduced in 1998) was not satisfied. Nitisols that undergo rejuvenation, may lose this criterion but still fulfil all other requirements and the concept of the Nitisols. As shown in the thesis, the introduction of this criterion is creating confusion in legacy databases as well. My finding is that this criteria is not useful and should be removed from the diagnostic criteria of the nitic horizon in updated versions of the WRB. (This proposal was accepted by the IUSS WG WRB)

5. The WRB proved to be a proper system to be applied for soil classification and soil characterization in the study area. It was well demonstrated that the building blocks (diagnostic horizons properties, materials and the qualifiers) do carry the important information for management purposes on soils of the studied area. Although my profiles did not closely match the soil associations of the KENSOTER polygons on the highest level, the information content derived from the associations provided a better match with the classification building blocks of my profiles in the relevant polygons. This underlines the importance of the diagnostic approach in soil classification, mapping and validation processes.

6. RELATED PUBLICATIONS

Mutuma E, Csorba Á, Dobos E, Reka B, Michéli E. (2017): Classification problems of Mount Kenya soils. *Geophysical Research Abstracts Vol. 19, EGU2017-16230,* 2017. http://meetingorganizer.copernicus.org/EGU2017/EGU2017-16230.pdf

Mutuma E, Csorba Á, Michéli E. (2017): Rapid assessment of soil properties for timely soil fertility management in the eastern slopes of Mt. Kenya. *Land degradation and development*. **Minor revisions resubmitted**.

Mutuma E, Csorba A, Michéli E. (2016): Prediction of soil properties using Mid-Infrared Spectroscopy and Random Forest regression in the Eastern slopes of Mt. Kenya Region. *Agricultural Science Research Journal*. 6: 253 – 262. http://resjournals.com/journals/agricultural-science-research-journal/OCTOBER%202016/Mutuma,%20%20Csorba%20and%20Mich%C3%A9li.pdf

Mutuma E, Mahiri I, Murimi S, Njeru P: Chapter 14: Adoption of water resource conservation under fluctuating rainfall regimes in Ngaciuma/Kinyaritha watershed, Meru County, Kenya. In: Walter LF, Esilaba AO, Rao KPC, Sridhar G. Adapting African Agriculture to climate change, climate change Management: **Springer, 2015**. pp 159-169. ISBN 978-3-319-12999-0. DOI: 10.1007/978-3-319-13000-2_14.
http://www.springer.com/kr/book/9783319129990

Njeru PNM, Mugwe J, Maina I, Mucheru-Muna M, Mugendi D, Lekasi JK, Kimani SK, Miriti J, Oeba VO, Esilaba AO, **Mutuma E,** Rao KPC, Murithi F: Chapter 16: Integrating farmers and scientific methods for evaluating climate change adaptation options in Embu County. In: Walter LF, Esilaba AO, Rao KPC, Sridhar G. Adapting African Agriculture to climate change, climate change Management: **Springer, 2015**. pp 185-197. ISBN 978-3-319-12999-0. DOI: 10.1007/978-3-319-13000-2_16.

http://www.springer.com/kr/book/9783319129990

Mutuma E, Láng V, Csorba A, Dobos E, Michéli, E: Classification problem of soils of Eastern slope of Mount Kenya. Proceedings of the 5th International Soil Classification Congress held in Bloemfontein, South Africa from 5-7 December 2016. p 28.

Mutuma E, Csorba Á, Michéli E: Conditional Latin Hypercube Sampling (CLHS) for selecting soil sampling locations for prediction and mapping of soil properties in Mt. Kenya Region. 28th Bi-annual conference of Soil Science Society of East Africa in Morogoro Tanzania 2015.

7. REFERENCES

Barré P, Fernandez-Ugalde O, Virto I, Velde B, Chenu C. (2014): Impact of phyllosilicate mineralogy on organic carbon stabilization in soils: incomplete knowledge and exciting prospects. Geoderma. 235: 382–395.

Batjes NH. (2009a): Harmonized soil profiles data for applications at global and continental scales: updates to the WISE database. Soil use and management. 25: 124-127. DOI: 10.1111/j.1475-2743.2009.00202.x

Carroll S, Goonetilleke A, Khalil WA, Ray F. (2006): Assessment via discriminant analysis of soil suitability for effluent renovation using undisturbed soil columns. Geoderma. 131: 201–217.

De Wispelaere L, Marcelino V, Alemayehu R, De Grave E, Dumon M, Mees F, Van Ranst E. (2015): Revisiting nitic horizon properties of Nitisols in SW Ethiopia. Geoderma. 243-244: 69–79.

Dijkshoorn JA. (2007): Soil and terrain database for Kenya (ver. 2.0) (KENSOTER). http://library.wur.nl/WebQuery/wurpubs/452214. Accessed 2014.03.17.

Dobos E, Michéli E, Baumgardner MF, Biehl L, Helt T. (2000): Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma. 97: 367–391. DOI: 10.1016/S0016-7061(00)00046-X

Dunn OJ.(1964): Multiple comparisons using rank sums. Technometrics 6: 241–252.

FAO. (2006): Guidelines for soil description, Fourth edition. ISBN 92-5-105521-1.

Gislason PO, Benediktsson JA, Sveinsson JR. (2006): Random forests for land cover classification. Pattern Recognition Letters. 27**:** 294–300.DOI: 10.1016/j.patrec.2005.08.011.

Goulding KWT. (2016): Soil acidification and importance of liming agricultural soils with particular reference to United Kingdom. Soil Use & Management. 32**:** 390-399. DOI: 10.1111/sum.12270.

IUSS Working Group WRB. (2015): World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.

Jaetzold R, Schmidt H, Hornetz B, Shisanya C. (2007): Farm management handbook of Kenya (2nd edition), Vol.II: natural conditions and farm management information, part C East Kenya. Ministry of Agriculture, Kenya.

Jenny H. (1941): Factors of soil formation: a system of quantitative pedology. McGraw-Hill, New York. http://soilandhealth.org/wp-content/uploads/01aglibrary/010159.Jenny.pdf (Accessed 2015.06.20)

Kennard R, Stone L. (1969): Computer aided design of experiments. Technometrics. 11:137–148. DOI: 10.1080/00401706.1969.10490666.

Loveland P, Webb J. (2003): Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil Tillage Research. 70: 1–18. DOI: 10.1016/S0167-1987 (02)00139-3.

McDowell ML, Bruland GL, Deenik JL, Grunwald S, Knox NM. (2012): Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma. 189: 312–320. Doi:10.1016/j.geoderma.2012.06.009.

Mehlich A. (1984): Mehlich-3 soil test extractant: a modification of Mehlich-2 extractant. Communications in Soil Science and Plant Analysis. 15:1409-1416. http://dx.doi.org/10.1080/00103628409367568.

Michéli E, Owens PR, Lang V, Fuchs M, Hempel J. (2014): Organic Carbon as a Major Differentiation Criterion in Soil Classification Systems. http://www.springer.com/us/book/9783319040837(Accessed online 21.02.2016)

Ministry of Agriculture. (1980): Exploratory Soil Map and Agro-ClimaticZone Map of Kenya, 1980 (Scale 1: 1M). Exploratory Soil Survey Report No. E1, Kenya Soil Survey, National Agricultural Laboratories, Ministry of Agriculture, Nairobi.

Mulder VL, de- Bruin S, Schaepman ME. (2013): Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. International Journal of Applied Earth Observation and Geo-information. 21: 301–310. DOI:10.1016/j.jag.2012.07.004

Ndakidemi PA, Semoka JMR. (2006): Soil fertility survey in Western Usambara Mountains, Northern Tanzania. Pedosphere. 16**:** 237–244. http://www.academia.edu/26332765.

Neteler M, Bowman MH, Landa M, Metz M. (2012): GRASS GIS: A multi-purpose open source GIS. Environmental Modelling and Software 31: 124 – 130.

Nocita M, Stevens A, van Wesemael B, Aitkenhead M, Bachmann M, Barthès B, Dor EB, Brow DJ, Clairotte M, Csorba A, Dardenne P, Demattê JAM, Genot V, Guerrero C, Knadel M, Montanarella L, Noon C, Ramirez-Lopez L, Robertson J, Sakai H, Soriano-Disla JM, Shepherd KD, Stenberg B, Towett EK, Vargas R, Wetterlind J. (2015): Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Advances in Agronomy. 132: 139–159. DOI:10.1016/bs.agron.2015.02.002

Okalebo JR, Othieno CO, Woomer PL, Karanja NK, Semoka JRM, Bekunda MA, Mugendi DN, Muasya RM, Bationo A, Mukhwana E.(2006): Available technologies to replenish soil fertility in East Africa. Nutrient Cycling and Agroecosystems. 76: 153-170. DOI: 10.1007/s10705-005-7126-7

R Core Team. (2013): R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Roudier P, Hewitt AE. (2012). A conditioned Latin hypercube sampling algorithm incorporating operational constraints. *Digital Soil Assessments and Beyond.*

Saeys W, Mouazen AM, Ramon H. (2005): Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. Biosystems Engineering. 91: 393–402.DOI: 10.1016/j.biosystemseng.2005.05.001.

Schoeman JJ. (1952): A geological reconnaissance of the country between Embu and Meru.Rep. geology Survey Kenya17.http://library.wur.nl/isric/fulltext/isricu_i15820_001.pdf (Accessed online 1.10.2015)

Shephed KD, Walsh MG. (2002): Development of reflectance libraries for characterization of soil properties. Soil Science Society of America Journal. 66: 988–998. DOI: 10.2136/sssaj2002.9880

Sila AM, Shepherd KD, Pokhariyal GP. (2016): Evaluating the utility of Mid Infrared spectral subspaces for predicting soil properties. Chemometrics and Intelligent Laboratory Systems. 153: 92-105. DOI: 10.1016/j.chemolab.2016.02.013

Skjemstad J, Baldock JA. (2008): Total and organic carbon. In ‘Soil Sampling and Methods of Analysis. 2nd Edition’ (MR Carter and EG Gregorich, eds.), pp. 225-238. Soil Science Society of Canada.

Terhoeven-Urselmans T, Vagen TG, Spaargaren O, Shepherd KD. (2010): Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. Soil Science Society of America Journal. 74: 1792-1799. DOI: 10.2136/sssaj2009.0218

Vågen TG, Winowiecki, LA, Abegaz A, Hadgu K.M. (2013): Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia. Remote Sensing of Environment. 134: 266–275. DOI: S0034425713000850

Van Reeuwijk P. (2002): Procedures for soil analysis. 6th ed. Tech. Pap. 9. Int. Soil Ref. and Inf. Ctr., Wageningen, the Netherlands.

Wand M, Ripley B. (2008): KernSmooth: Functions for kernel smoothing R *package version* 2: 22–22. org/web/packages/KernSmooth/KernSmooth.pdf.