

Contents lists available at [ScienceDirect](http://ScienceDirect.com)

Geoderma Regional

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

Total elemental composition of soils in Sub-Saharan Africa and relationship with soil forming factors



Erick K. Towett^{a,b,*}, Keith D. Shepherd^b, Jerome E. Tondoh^c, Leigh A. Winowiecki^d, Tamene Lulseged^e, Mercy Nyambura^b, Andrew Sila^b, Tor-G. Vågen^b, Georg Cadisch^a

^a Institute for Plant Production and Agroecology in the Tropics and Subtropics, Universität Hohenheim, Garbenstr. 13, Stuttgart 70599, Germany

^b World Agroforestry Centre (ICRAF), P.O. Box 30677, Nairobi 00100, Kenya

^c World Agroforestry Centre (ICRAF), WCA Sahel Node BPE, 5118 Bamako, Mali

^d International Center for Tropical Agriculture (CIAT), c/o ICIPE Dudsonville Campus, P.O. Box 823, Nairobi 00621, Kenya

^e International Center for Tropical Agriculture (CIAT), c/o Chitedze Agricultural Research Station, P.O. Box 158, Lilongwe, Malawi

ARTICLE INFO

Article history:

Received 5 May 2015

Received in revised form 15 June 2015

Accepted 22 June 2015

Available online 28 June 2015

Keywords:

TXRF

Spatial variation

Soil nutrients

Total element composition

Africa soil

Soil-forming factors

Acrisols

Lixisols

Plinthosols

Vertisols

Arenosols

Nitisols

Ferralsols

Leptosols

Fluvisols

Calcisols

Cambisols

Luvissols

Alisols

Gleysols

ABSTRACT

A thorough understanding of the variation in total soil element concentrations is important especially in the Sub-Saharan Africa (SSA) soil contexts for agricultural and environmental management at large scale. Fingerprinting of soil elemental composition may form a useful basis for evaluating soils in a way that relates to soil-forming factors and inherent soil functional properties. The objectives of this paper are to quantify the proportion of variability in total elemental composition by total X-ray fluorescence (TXRF) method of 1074 soil samples from the Africa Soil Information Service (AfSIS) Project baseline and to determine the relationships with soil forming factors. The samples were from 34 sentinel sites measuring 10 × 10 km, randomized within major climate zones in SSA. Within each sentinel site there were sixteen spatially stratified 1 km² clusters, within which there were ten 100 m² plots. The within and between site patterns of variation in total element composition of 17 elements; Al, P, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, Sr, Y, Ta, and Pb, were explored. Total element concentration values were within the range reported globally for soil Cr, Mn, Zn, Ni, V, Sr, and Y and higher than reported range for Al, Cu, Ta, Pb, and Ga. There were significant variations ($P < 0.05$) in total element composition within and between the sites for all the elements analyzed with the greatest proportion of total variance and number of significant variance components occurring at the site (55–88%) followed by the cluster nested within site (10–40%) levels. The explorations of the relationships between element composition data and site factors using Random Forest regression demonstrated that soil-forming factors have important influence on total elemental composition in the soil. The fact that the soil-forming factors are related to the concentration of naturally occurring elements in the soil gives rise to the notion that they might be predicted from the soils' element composition. Results implied that >70% of variation in soil element composition patterns can be predicted using information in existing databases or readily observable features. Successful use of TXRF technique would open up possibilities for using total soil elemental composition fingerprints as a useful basis for characterizing soils in a way that relates to soil-forming factors and inherent soil functional properties.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Knowledge of the distribution and baseline concentrations of major and trace elemental contents in soils will help to assess soil, land and environmental quality especially in developing countries in Sub-Saharan Africa (SSA). Fingerprinting of soil elemental composition may form a useful basis for classifying soils in a way that relates to soil-forming factors and inherent soil functional properties (Kabata-Pendias and

Mukherjee, 2007; Rawlins et al., 2012). This is due to the fact that variations in the concentrations of soil chemical elements are derived from differences in the composition of the parent material and from fluxes of matter and energy into or from soils over time (Helmke, 2000; Rawlins et al., 2012). In addition, soil-forming factors (e.g., parent material, climate, topography, vegetation and time) are important drivers of total elemental concentrations in the soil (Jenny, 1941). The nature of the key variables explaining ecological diversity of soils can be related to the mineralogy of parent rock and although these relationships have been inferred, mineralogy of parent rock is a principal factor determining spatial patterns of land resources (Voortman, 2011). The mineralogy and elemental composition of soil are related to the nature

* Corresponding author at: World Agroforestry Centre (ICRAF), P.O. Box 30677, Nairobi 00100, Kenya.

E-mail address: e.towett@cgiar.org (E.K. Towett).

of the parent material and the degree to which this material has been weathered (Negrel et al., 2014).

Soil mineralogy is a key determinant of many soil functional properties, for example nutrient quantities and intensities, pH and buffering, anion and cation exchange capacity, aggregate stability, soil carbon protection, dispersion, and resistance to erosion. Primary associations of elements reflect mineralogy and some elements can thus be used as proxies for mineralogy including heavy minerals, carbonates and clays (Grunsky et al., 2009). While total elemental concentrations in soil reflect parent material concentrations are also impacted by weathering rates and over the longer term by land use and climate. Rawlins et al. (2009) demonstrated use of element composition for the prediction of particle size distribution and their errors.

There is potential for increasing the knowledge of the spatial extent of trace elements using recent advances in analytical chemistry and increasing applications. Total X-ray fluorescence spectroscopy (TXRF) is a relatively new technique that can provide for a rapid and simultaneous determination of the concentrations of many elements from Na to U in the periodic table with minimal sample preparation time and low matrix interferences (Stosnach, 2005; Towett et al., 2013). However, TXRF has not been widely used as an analytical technique for the total element composition of soils of SSA due to low number of well equipped soil labs. Towett et al. (2013) tested a method for the use of TXRF for direct quantification of total element concentrations in soils, demonstrating the potential utility of TXRF for rapid screening of total element concentrations in soils assuming sufficient calibration measures are followed. Successful use of the TXRF technique would open up the possibilities for using total element composition to improve global predictions of soil properties, such as cation exchange capacity and available nutrients, especially in Africa where variations in soil mineralogy and nutrient balance critically determine vegetation composition and agricultural potential (Voortman, 2011).

This paper uses data from several projects, including the first phase (2009–2012) of the Globally Integrated Africa Soil Information Service (AfSIS) project, which was a research-based project to develop a practical, timely, and cost-effective soil health surveillance service to map soil conditions, set a baseline for monitoring changes, develop global standards and methodologies, and provide options for improved soil and land management in Africa (Vågen et al., 2010). The AfSIS project area included about 17.5 million km² of continental Sub-Saharan Africa (SSA), excluding hot and cold deserts, encompassing more than 90% of Africa's human population living in 42 countries. All of the data used for this paper were taken from AfSIS sites surveyed using the Land Degradation Surveillance Framework (LDSF), (Vågen et al., 2010, 2015; Vågen and Walsh, 2012) in the period from January 2010 to March 2012. The LDSF utilizes novel data collection methodologies that are efficient, cost-effective, and vastly improve the analytical precision of the landscape level estimates by producing a suite of soil and vegetation indicators that are spatially specific and continuous across the surveyed landscape (Shepherd et al., 2015; Vågen and Walsh, 2012; Vågen and Winowiecki, 2013; Vågen et al., 2010, 2013; Winowiecki et al., 2015). These include new direct soil spectral methods using only light (visible, infrared and X-ray) that hold promise for providing rapid, low cost and reproducible soil characterization (Shepherd, 2010). The hierarchical, systematic and random sampling approach reduces sampling effort and cost. The LDSF is intended to provide a whole landscape (systems) perspective to land use planning and decision making to enable countries to both increase agricultural productivity and maintain ecosystem services (Winowiecki et al., 2015; Shepherd et al., 2015). The potential for spectroscopy as a key component of soil health surveillance systems has already been articulated (Shepherd and Walsh, 2007; Shepherd et al., 2015). This work thus set out to determine the usefulness of TXRF as an analytical technique for the total element composition of soils from 34 LDSF sentinel sites. Specific objectives of this paper were to evaluate the relationship between TXRF total element concentrations in soils and to test how well element composition is predicted by soil

forming factors as such relationships could then be used as proxies to infer soil functional properties.

2. Materials and methods

2.1. Study area and sampling

The current study was based on georeferenced samples associated with the AfSIS project taken from a set of thirty four 100-km² LDSF sentinel sites: Ghana (3 sites), Tanzania (8 sites), Ethiopia (4 sites), Mali (3 sites), Burkina Faso (1 site), Mozambique (4 sites), Nigeria (3 sites), Zambia (1 site), Kenya (3 sites), Guinea (2 sites), and Malawi (2 sites) (Fig. 1). Field sampling was conducted based on the field methods collectively referred to as LDSF protocol (Vågen et al., 2010), also described in Vågen et al., 2012, 2015, and Winowiecki et al., 2015. Within each sentinel site there were ten 1000 m² plots randomized within sixteen spatially stratified 1 km² clusters. Topsoil (0–20 cm) and subsoil (20–50 cm) samples were collected from Plot 1 from each cluster giving a total of 1074 samples (16 plots per site × 2 soil depths × 34 sentinel sites). A summarized description of the 34 sentinel sites from which the soil samples used for this study were collected, including the location, average elevation, annual total precipitation, temperature ranges, major soil classes (based on the IUSS Working Group WRB (2014) and the FAO/EC/ISRIC (2003) world soil resources map (scale 1:30,000,000 approx)), major landforms, topography, percentage of cultivated area and vegetation structure is given in Table 1.

2.2. Sample preparation and analyses

Samples were analyzed at the World Agroforestry Centre's Spectral Diagnostics Laboratory in Nairobi, Kenya using TXRF and X-ray diffraction spectroscopy (XRD) spectral methods on all samples collected. Soil samples were first air-dried and passed through a 2-mm sieve before sub sampling to 10 g using coning and quartering technique to ensure homogeneity of the sample. The 10 g subsamples were oven-dried at 40 °C and then 5 g were ground to a fine powder (<200 µm) using a Retsch RM 200 mill (Retsch, Düsseldorf, Germany). Approximately 3 g of each soil sample was then further ground to <50 µm using a McCrone micronising mill (McCrone, Westmont, U.S.A.). The TXRF methodology was used to analyze total elemental concentrations in each soil sample using a S2 PICOFOX™ TXRF spectrometer (Bruker AXS Microanalysis GmbH, Germany). Based on a recently developed and tested method (Towett et al., 2013), 50 mg of the finely ground (20–50 µm) sample was mixed with 2.5 ml of Triton X100 (Fischer Scientific, UK) solution (0.1 vol.%) to form a soil suspension and spiked with 40 µl of 1000 mg l⁻¹ Selenium (Fluka Analytical, Germany) as the internal standard. The suspension was placed into an ultrasonic water bath at room temperature and sonicated in a continuous mode for 15 min, and then mixed well using a digital shaker. 10 µl of the turbid soil solution was immediately dispensed on to a clean siliconized quartz glass sample carrier and dried for 10–15 min at 52 °C on a hot plate (Staurt® SD300) in a clean laminar flow hood. Samples were analyzed in triplicate with data acquisition time of 1000 s per sample. The evaluation of spectra and element quantification were performed using the software program SPECTRA 6.3 (Bruker AXS GmbH, Germany). Mineral profiling was done using XRD spectroscopy where finely ground (<50 µm) samples were loaded into steel sample holders and analyzed using a Bruker D2 Phaser XRD spectrometer instrument that was equipped with a LYNXEYE compound silicon strip, 1-dimensional detector with Theta/Theta geometry. The instrument was integrated with the DIFFRACplus TOPAS graphics based, non-linear least squares profile analysis program. Identification of X-ray minerals was achieved by comparing the X-ray diffraction pattern (diffractogram) obtained from an unknown sample with an internationally recognized database (powder diffraction file, PDF) containing reference patterns with more than 70,000 phases. Quantitative phase analysis of crystalline powder samples was then

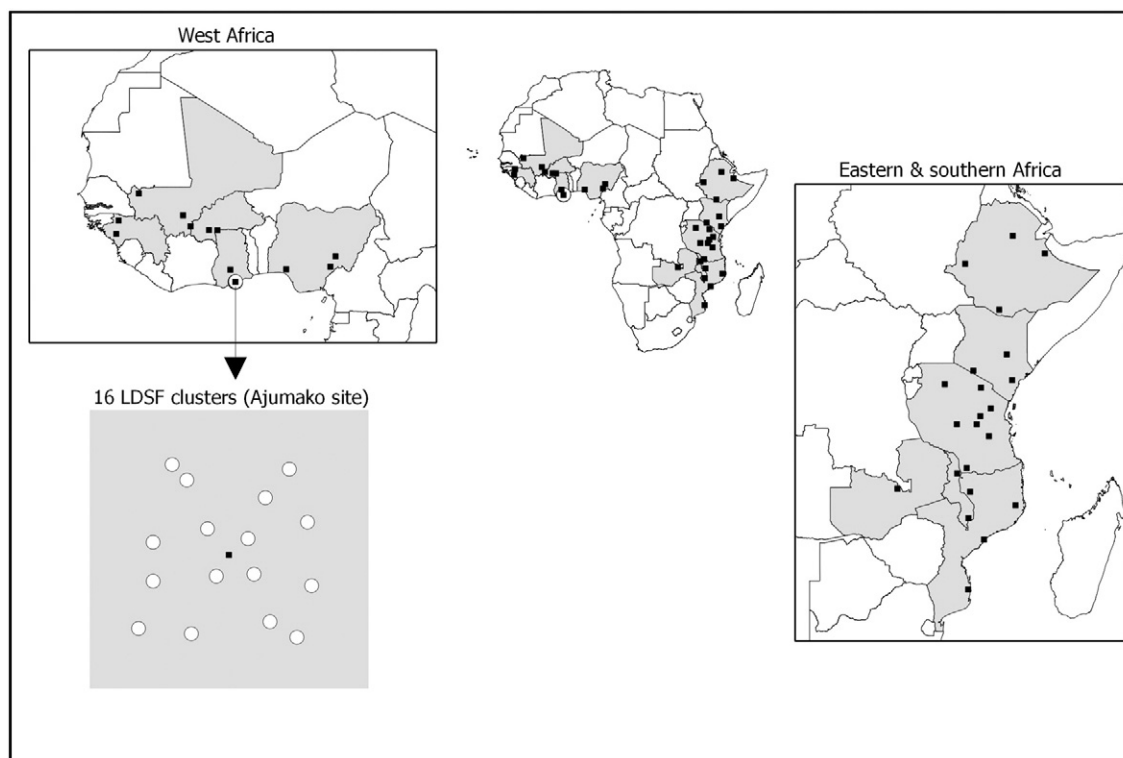


Fig. 1. Graphics illustrate the centroid locations of 34 AfSIS sentinel sites sampled and available for this study. The basic sampling unit (sentinel site) is a 10×10 km area (black squares), within which there is a spatial hierarchy of 16 randomized LDSF clusters down to 1000 m^2 plots. The layout of each sentinel site sampled for this study consisted of 16 clusters (white circles).

done using a full pattern fitting method (Rietveld method e.g., using TOPAS program).

2.3. Data analysis

In TXRF analysis an important consideration is the capability of the instrument to detect whether an element is present or not in a specimen and to be able to show with some defined statistical certainty that a given element is present if its concentration is greater than a lower limit of detection (LLD) (Rousseau, 2001). The LLD, which in this study is assumed to be the concentration equivalent to three standard counting errors of a set of measurements of the background intensity (Bruker, 2007), was calculated using the following formula (Klockenkämper, 1997):

$$LLD_i = \frac{3 \cdot C_i \cdot \sqrt{N_{BG}}}{N_i} \quad (1)$$

where LLD_i = LLD of the element i ; C_i = concentration of the element i ; N_i = area of fluorescence under peak in counts; and N_{BG} = background area subjacent the fluorescence peak.

We excluded from our dataset elements that were not accurately estimated using the TXRF technique as outlined by Towett et al. (2013), leaving us with 17 elements (Al, P, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, Sr, Y, Ta and Pb) out of 38 elements quantified by the technique. Since knowing that an element concentration below a given minimum detectable value is useful, we developed a procedure for assigning values to the LLD. In order to develop the analysis, we first of all made scatter plots of the total element concentration values against the LLD values calculated according to equation. We took the 25th percentile of the LLD values as the LLD limits for trace elements: Cu, Zn, Ga, Sr and Y $< 0.2 \text{ mg kg}^{-1}$, Al $< 208 \text{ mg kg}^{-1}$, P $< 42 \text{ mg kg}^{-1}$, K $< 8.6 \text{ mg kg}^{-1}$, Ca $< 4.1 \text{ mg kg}^{-1}$, Ti $< 1.6 \text{ mg kg}^{-1}$, V, Cr $< 1.1 \text{ mg kg}^{-1}$, Mn,

Fe $< 0.9 \text{ mg kg}^{-1}$, Ni $< 0.3 \text{ mg kg}^{-1}$, Ta $< 0.1 \text{ mg kg}^{-1}$, and Pb $< 0.4 \text{ mg kg}^{-1}$. We substituted the missing values with the 25th percentiles of average LLD values of the respective elements. The approach was similar to that used by Reeves and Smith (2009) in their examination of mid- and near-infrared diffuse reflectance spectroscopy as possible tools for the determination of major and trace elements in soils where they replaced any value below the method detection limit with one-half the detection limit prior to statistical calculations.

Concentrations of these elements showed positive skewness in frequency distributions for most of the elements. Therefore, the appropriate transformation to achieve close to statistical normality of each variable was done before proceeding with multivariate data analysis. We then established total concentrations of elements for soils occurring within particular sites sampled for this study, and documented variation in their concentration and explored the possibility of finger-printing complete element profiles. We used a principal component analysis (PCA) based on a correlation matrix using the R package "FactoMineR" (Lê et al., 2008) to explore the variation and view the interactions between different element concentrations and to detect sample patterns and variable relationships within and between the sites. To explore the proportion of variance for each element between sites, between clusters within site, depth, site and depth interaction and residual variance, we performed a mixed model analysis of variance using the following PROC MIXED statement:

$$\begin{aligned} \text{model log(element)} = /\text{ddfm} = \text{kr residual outp} = \text{resid;} \\ \text{random site; random depth; random} \\ \text{site * depth; random site * cluster} \end{aligned} \quad (2)$$

where log = natural log; ddfm = specifies the method for computing denominator degrees of freedom; kr = Kenward-Roger; outp = output; and resid = residuals.

Table 1
Site description: location, average elevation, annual total precipitation, monthly temperatures ranges, major soil classes, topography (toposequence), percentage of cultivated area and landuse of the 34 sentinel sites from which the soil samples used for this study were collected.

Site	Country	Lat	Long	Elev (m)	Precip (mm)	Temp range (°C)	Soil class*	PosToposeq	Cultivated area (%)	Landuse
Ajumako	Ghana	5.408	−0.745	63.1	1217.8	23.6–31.7	Acrisols, Alisols, Plinthisols	Midslope	75	–
Bondigui	Burkina Faso	10.913	−3.546	329.3	1002.3	24.4–37.4	Lixisols	Mid-footslope	44	Wooded grassland
Boumeoul	Guinea	11.935	−13.144	100.5	1692.9	24.1–38.7	Plinthosols	Upland footslope	31	Woodland
Bukwaya	Tanzania	−3.024	33.050	1197.2	876.4	18.6–30.8	Vertisols	Upland–bottomland	59	Cropland/shrubland
Chica_b	Mozambique	−14.714	39.873	275.6	1025	20.2–32.2	Arenosols	Midslope	27	Thicket/bushland
Chiculecule	Mozambique	−22.836	35.302	70.3	911.7	20.5–31.8	Arenosols	Upland	52	Cropland/woodland
Chinyanghuku	Tanzania	−6.895	36.129	902.8	600.3	18.6–31.2	Lixisols	Midslope–bottomland	20	Cropland/bushland
Dambidolo	Ethiopia	8.613	35.015	1348.8	1349.3	16.1–31.6	Nitisols, Andosols	Upland–bottomland	66	Cropland/woodland
Finnkolo	Mali	11.313	−5.502	409.1	1083	23.8–37	Lixisols	Footslope	63	–
Fisenge	Zambia	−13.094	28.477	1249.6	1239.3	16.7–31.6	Ferralsols, Acrisols, Nitisols	Upland–midslope	63	Cropland/wooded grassland
Fria	Guinea	10.507	−13.393	193.2	2619.2	23.5–37.8	Leptosols	Midslope–footslope	63	Cropland/wooded shrubland
Ibi	Nigeria	8.138	9.894	133.5	1113.2	24.1–37.1	Fluvisols, Gleysols, Cambisols	Midslope–bottomland	47	Shrubland/cropland
Ihassunge	Mozambique	−18.013	36.832	7.9	1316.3	23.3–32.7	Calcisol, Cambisol, Luvisol	Bottomland	81	Cropland
Imorun	Nigeria	6.753	4.658	113.4	1723.3	23.6–33.1	Acrisols, Alisols, Plinthosols	Upland–midslope	63	Forest
Itende	Tanzania	−6.891	34.208	1216.5	630.6	17.6–30.3	Acrisols, Alisols, Plinthosols	Upland–midslope	0	Woodland
Katsina Ala	Nigeria	7.014	9.342	175.5	1643.1	23.3–35	Lixisols	Upland–bottomland	87	–
Kiberashi	Tanzania	−5.346	37.482	1109.9	758.1	16–28.8	Lixisols	Midslope	25	Forest/woodland
Kidatu	Tanzania	−8.036	37.303	319.4	1368.1	23.4–30.5	Lixisols	Upland	0	Woodland
Kisongo	Tanzania	−3.355	36.541	1328.8	875.5	16.1–28.4	Nitisols, Andosols	Midslope–footslope	44	–
Koloko	Mali	12.483	−6.296	292.1	899.3	25.3–38.3	Lixisols	Upland–bottomland	81	Cropland
Kontela	Mali	14.809	−10.999	67.1	603.7	27–41	Lixisols	Footslope/bottomland	25	Shrubland/grassland
Kubeasi	Ghana	6.723	−1.277	229.3	1473.6	22.2–33.3	Acrisols, Alisols, Plinthosols	Upland–footslope	56	–
Kutaber	Ethiopia	11.296	39.607	2515.7	1091.4	10.3–25.5	Nitisols, Andosols	Midslope–footslope	63	Cropland/shrubland
Lambussie	Ghana	10.896	−2.649	308.1	961.5	24.5–37.3	Lixisols	Midslope	87	Cropland/shrubland
Macassangila	Mozambique	−13.407	35.495	1109.4	1228.1	17.8–29.3	Ferralsols	Midslope	45	Woodland/grassland/cropland
Marafa	Kenya	−2.634	39.546	165.5	687.2	22.6–32.9	Calcisols, Cambisols, Luvisols	Bottomland	0	Shrubland
Mbalambala	Kenya	−0.126	39.027	217.4	321.1	22.1–36.3	Calcisols, Cambisols, Luvisols	Upland	0	Shrubland
Mbinga	Tanzania	−11.091	35.161	983.4	1328.1	19.1–29.5	Lixisols	Midslope	31	Woodland
Mega	Ethiopia	4.182	38.296	1540.1	592.1	12.6–27.8	Vertisols	Bottomland	0	Bushland/shrubland/grassland
Merar	Ethiopia	9.602	42.701	1897.6	789.1	14.2–26.3	Calcisols, Cambisols, Luvisols	Upland	58	Cropland/shrubland/grassland
Morijo	Kenya	−1.719	35.811	2213.2	869.5	11.1–23.9	Calcisols, Cambisols, Luvisols	Midslope–footslope	6	Shrubland/grassland
Nkhata Bay	Malawi	−11.626	34.239	551.1	1672.8	21–30.4	Ferralsols, Acrisols, Nitisols	Midslope–bottomland	69	Cropland/woodland/grassland
Pandambili	Tanzania	−6.084	36.474	1086.8	727.3	16.7–31.4	Lixisols	Bottomland	6	Grassland
Thuchila	Malawi	−15.940	35.329	708.9	1517	19.3–30.7	Lixisols	Upland	94	Cropland

Lat – latitude; Long – longitude; °C – degree centigrade; Elev (m) – average elevation in meters; Precip (mm) – average total precipitation in mm; Temp range – monthly temperature ranges; PosTopoSeq – average topographic position at the site level. *Major soil classification in the site according to [IUSS Working Group WRB, 2014](#). Elevation, precipitation and temperature data were calculated using the WorldClim sets of climate layers (climate grids) for Africa with a spatial resolution of about 1 km² ([Hijmans et al., 2005](#)). The soil classifications are based on the [IUSS Working Group WRB \(2014\)](#) and the [FAO/EC/ISRIC \(2003\)](#) world soil resources map (Scale 1:30,000,000 approx), while the other site description data were extracted from the Land Degradation Surveillance Framework (LDSF) database used in AfSIS.

The correlations among the five parameters (sites, between clusters within site, depth, site and depth interaction and residual variance) were modeled using compound symmetry covariance matrix. The model was fitted by the Restricted Maximum Likelihood (REML) method in the SAS MIXED procedure (SAS Institute, 2011). A summary of some of the important options invoked in the SAS PROC MIXED statement by function and the class level information for three classes (site, cluster and depth), are summarized in Annex Table A2. In order to examine relationships between element fingerprints (principal components (PCs)) and site characteristics including elevation, climate, major landform, landform designation, vegetation type, management (cultivation) and mineralogy, we used the R package “randomForest” (Breiman, 2001; Liaw and Wiener, 2002) with an ‘out-of-bag’ (RF-OOB) test sample being held out and used to estimate model error and for the calculation of variable importance. We accessed the extent to which site and soil-forming factors could explain the variation in the element composition of the soil. Specific research questions addressed included:

- (i) Are patterns of elemental concentrations related to the site factors and if so, which are dominant?
- (ii) Does element concentration clustering relate to mineralogy?

For the first question, we performed a Random Forests (RF) regression (R-version 2.15.3) of the first 5 PCs of the TXRF element concentration data against site factors while for the second question, we explored the RF regression of elements PCs against mineral composition. We included and excluded different variables to test their effects on the overall model and variable importance.

3. Results and discussion

3.1. Total elemental concentrations

As expected, there was wide variation in total element concentrations (Table 2), which was presumably attributable to differences in parent materials between sites and to local pedologic and hydrological factors within sites or due to differences in management (Table 1). For

example, the total concentration of P varied from 25 to 2358 (mean 143) mg kg⁻¹ across the 34 sentinel sites, and K and Ca values ranged between 291–77,898 and 82–426,431 mg kg⁻¹, respectively (Table 2). These results of total elemental analyses of soils in the different sentinel sites selected across SSA were within the reported ranges of worldwide soils (Table 2). However, the mean values for V, Zn and Sr were below the mean worldwide contents (Kabata-Pendias and Mukherjee, 2007) by 23 (38%), 33 (53%) and 29 (20%) mg kg⁻¹ respectively, while the mean values obtained for soils of SSA exceeded the mean worldwide concentrations by 48 (11%), 12 (48%), and 22 (52%) mg kg⁻¹ for Mn, Pb, and Cr respectively.

3.2. Principal component analysis of the total elemental concentrations

The PCA of element concentrations revealed that patterns in total element concentrations between sites appeared to relate to differences in mineralogical ‘functional groups’ (Fig. 2a–b). The pattern of clustering of the individual minerals and sorting of heavy minerals (V, Pb, Ni, Cr, Cu, Ti, and Fe) along the positive Dim 1 axis is apparent (Fig. 2a). The first two PCA axes together accounted for 46.6% (Dim 1 – 32.5%, and Dim 2 – 14.1%) of the total variance within the data set (Fig. 2a). The elements K, Ca, Ti, Fe and Sr were predominantly associated with Dim 1 while Dim 2 was dominated by elements Al, Ni and Ga. Elements Cr, Mn and Zn were predominant in Dim 3 (Fig. 2b, Annex Fig. A1), while only V was predominant in Dim 4 and the elements P, Cu, Y, Ta and Pb were predominant in Dim 5 (Annex Fig. A1).

In the results, clay minerals in the PCA plot are represented by some elements, e.g., Al, Cr, V, Cu, Fe, and Ti, most of which fell along the positive Dim 1 axis (Fig. 2a), similar to results reported by Cannon and Horton (2009) and Grunsky et al. (2009), i.e., that the clay factor commonly includes, although not exclusively, the elements Al, Cr, Cu, Ni, V, and Fe, all elements that are common in trace amounts in clay minerals or adsorbed to them. In the PCA results, the elements typical of feldspars and carbonate minerals (Ca, K and Sr) (Fig. 2a) indicate the strong association of Ca and Sr which are reported to be highly correlated ($r = 0.98$, $p = 0.0001$) and having similar chemical properties, Sr readily substitutes for Ca in the structure of carbonate minerals (Marzecová et al., 2011). In this study, the sites with a strong Ca–Sr

Table 2

Summary statistics of the median, mean, minimum, percentiles and maximum values of the total element concentrations (mg kg⁻¹) of 1074 soil samples collected from 34 sentinel sites and the reported concentration mean ranges of background contents of elements in the world soils.

Element	Values compiled from this study (mg kg ⁻¹)								Reported mean and ranges of background contents of elements in crust and worldwide soils (mg kg ⁻¹)							
	Median	Mean	Min	2.5 th %ntile	25 th %ntile	75 th %ntile	97.5 th %ntile	Max	Crustal Average ^a	A	B	C	D	E	F	Reported worldwide ranges ^b
Al	34,187	33,927	94	251	21,372	45,719	68,242	89,068	–	–	–	–	–	–	–	10,000–40,000
P	42	143	25	42	42	44	1169	2358	–	–	–	–	–	–	–	–
K	7902	10,893	291	653	2749	16,802	34,366	77,898	–	–	–	–	–	–	–	–
Ca	1908	9780	82	173	682	8576	68,638	426,431	–	–	–	–	–	–	–	–
Ti	3193	4264	2.6	30	1921	5533	13,068	25,611	4400	–	3700	–	15,480	2900	–	200–24,000
V	19	37	0.7	1.1	7.2	44	160	393	135	60	69	180	320	80	–	5.0–500
Cr	45	64	0.7	1.4	24	78	250	598	100	42	22	58	86	54	72	1–1500
Mn	276	466	1.6	13	138	637	1598	6575	900	418	411	–	535	550	–	<7–>9000
Fe	20,789	27,954	20	234	9567	39,068	91,014	181,691	–	–	–	–	–	–	–	1000–550,000
Ni	12	19	0.3	0.9	5.0	24	72	364	20	18	13	26	25	19	39	0.2–500
Cu	13	17	0.3	1.0	6.0	22	58	114	55	14	17	48	109	25	17–29	1.0–250
Zn	20	29	0.3	1.4	10	42	91	138	70	62	65	89	73	60	45–47	10–602
Ga	7.8	8	0.2	0.2	4.3	12	20	31	15	1.2	8.9	20	31	17	–	0.4–70
Sr	47	118	1.2	3.1	20	129	807	1985	375	147	163	190	–	240	–	32–>1000
Y	9.2	13	0.2	0.5	4.8	18	44	109	33	12	27	89	73	60	–	16–33
Ta	2.7	3	0.1	0.1	1.1	4.5	9.1	16	2.0	1.1	1.1	1.7	2.3	–	–	0.8–5.3
Pb	18	37	0.3	0.4	8.4	38	189	639	14	25	18	24	22	19	18–22	2.0–16,338

Given are mean values for various soils: A – worldwide mean contents after Kabata-Pendias and Mukherjee (2007); B – Sweden soils after Eriksson (2001) (cited by Kabata-Pendias and Mukherjee (2007)); C – Japan agricultural soils, after Takeda (2004) cited by Kabata-Pendias and Mukherjee (2007); D – Median for soils in Brazil, after Licht (2005) (cited by Kabata-Pendias and Mukherjee (2007)); E – data for U.S. and Canada soils, after Burt et al., 2003 (cited by Kabata-Pendias and Mukherjee (2007)) and Haluschak et al. (1998); F – Reported median values for Africa compiled from Ghana soils, after Antwi-Agyei et al. (2009) and South African soils, after Herselman (2007).

^a Values compiled by Kabata-Pendias and Mukherjee (2007).

^b Values compiled from Kabata-Pendias and Mukherjee (2007), Hooda (2010), U.S. EPA Verification Report (2006), and Haluschak et al. (1998).

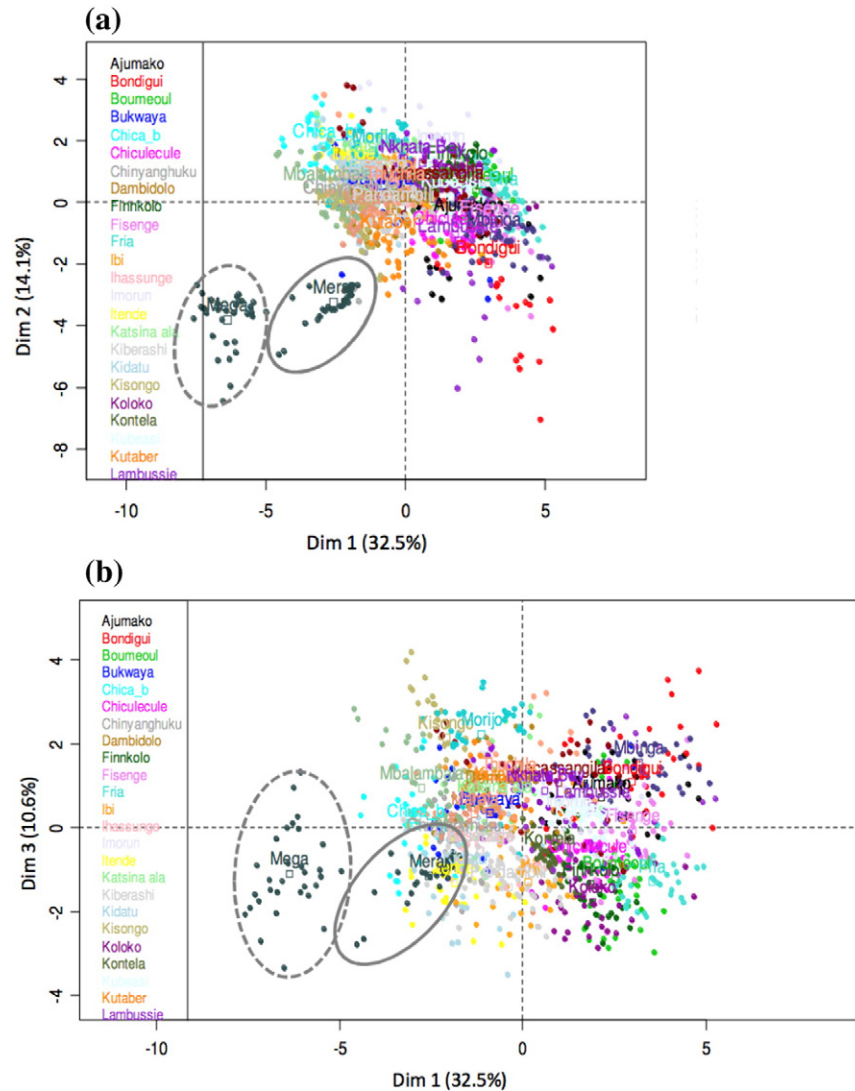


Fig. 3. Scatter plots of individual site factors, based on the principal component Dim 1 vs Dim 2 (a) and Dim 1 and Dim 3 (b), on the log transformed data of the soil total element concentration from all sites analyzed.

could also be explained by management and vegetation at the different sites (Table 1). For example, the Mega site which lies on an extensive rangeland area with Vertisols representing major management constraints brought about by the physical properties and the soil moisture regime of Vertisols and since they are used only for extensive grazing, they are consequently poor in nutrients because of no external inputs (Vågen et al., 2013). The total element concentration analysis of the soils from the Mega site revealed very low concentrations of all elements with the exception of Ca (Tables A4–A7). Fig. 3a–b acknowledges the fact that soil-forming factors are important drivers of element concentrations in soils since it shows sentinel sites with similar characteristics clustering together. The PCA patterns in total element concentrations within and between the sites, (Fig. 3, and Annex Fig. A2), could thus be used as proxies for natural nutrient status and the processes that might control the variation and differences in mineralogical “functional groupings”, and increase the wealth of knowledge of concentrations and spatial distribution of naturally occurring minerals in the soils of SSA.

3.4. Restricted maximum likelihood analysis of the proportion of variance

There were significant ($P < 0.05$) variance proportions for site, cluster nested within site, as well as depth nested within cluster and site, and the interactions between site and depth (Table 3). The greatest

proportion of total variance and number of significant variance components occurred at the level of site (55–88%) followed by the cluster nested within site levels (10–40%). This was not surprising given that we would expect lower variation within depths at a given sampling location. However, for the element P, a large proportion of the samples (75%) were at or below the 25th percentile LLD concentration of the analytical measurements. Consequently, our results concerning variation of P between and within sentinel sites should be interpreted with caution. Nevertheless, this may have implications for natural production systems and agriculture, as it may also indicate low potential P supply from minerals because TXRF is capable of detecting minerals rich in P. There were also 3 elements with a large error of variance (Al, Ti and Ga). Overall site contributed to more than 55% of the total variance of all the 17 elements analyzed (Table 3). Our results differed with those reported by Laiho et al. (2004) who examined the proportion of variance of the elements K, Zn and Mn in soils from 11 peatland forest sites in Central Finland and reported strong variation among the sites, but even for these elements, the within-site variation contributed the biggest proportion of the total variance for the depth 0–30 cm. However, the cluster nested within site and site interaction contributed to at least 10% of the total variance of all the 17 elements analyzed (Table 3). The observed strong within site as well as between site variations in the elements in our study were expected to be highly

Table 3
Covariance parameter estimates and their percentage contribution to the total variance for site, cluster nested within site and depth nested within cluster and site and their interactions.

Element	%LLD#	n	Site		Site * cluster		Site * depth		Depth		Residual	
			Estimate	%Tot var	Estimate	%Tot var	Estimate	%Tot var	Estimate	%Tot var	Estimate	%Tot var
Al	2	1068	0.966	88	0.112	10	0.004	0.4	0.005	0.45	0.016	1.4
P	75	1059	0.718	76	0.198	21	0.002	0.2	1.4×10^{-21}	<0.01	0.025	2.6
K	0	1065	0.913	71	0.354	28	0.003	0.2	6.8×10^{-21}	<0.01	0.010	0.8
Ca	0	1068	2.186	79	0.480	17	0.034	1.2	0.017	0.60	0.051	1.8
Ti	0	1067	1.398	87	0.199	12	0.001	0.1	0.001	0.04	0.014	0.9
V	7	1067	1.463	77	0.379	20	0.009	0.5	0.008	0.39	0.053	2.8
Cr	2	1068	0.808	65	0.384	31	0.005	0.4	0.006	0.46	0.039	3.2
Mn	0	1067	1.007	68	0.393	27	0.023	1.6	0.008	0.51	0.040	2.7
Fe	0	1066	1.459	80	0.335	18	0.005	0.3	0.009	0.47	0.026	1.4
Cu	0	1066	0.705	69	0.285	28	0.002	0.2	0.003	0.27	0.027	2.6
Ni	0	1067	0.745	55	0.540	40	0.014	1.1	0.008	0.56	0.049	3.6
Zn	0	1067	0.924	79	0.217	19	0.001	0.1	0.001	0.12	0.026	2.2
Ga	3	1067	0.677	77	0.164	19	0.006	0.7	0.009	1.01	0.024	2.7
Sr	0	1069	1.554	83	0.299	16	0.004	0.2	0.000	0.01	0.017	0.9
Y	1	1066	0.721	66	0.309	28	0.004	0.3	0.000	0.00	0.061	5.6
Ta	6	1069	1.015	73	0.313	23	0.005	0.4	0.004	0.28	0.050	3.6
Pb	2	1069	1.103	57	0.689	36	0.011	0.6	0.009	0.45	0.117	6.0

%LLD# = percentage of the total number of samples ($n = 1074$) with values at or below the LLD concentration; n = number of observations used after removing extreme studentized residuals; Estimate = log value estimated by the model; % Totvar = percentage contribution of the parameter estimate to the total variance. Model fitted with the SAS System Mixed Procedure with the following parameters: covariance structure = variance components; estimation method = REML; residual variance method = profile; fixed effects SE method = Kenward-Roger.

diagnostic of soil fertility potential because site and the interaction of cluster nested within site contributed more than half of the total variance for all elements (Table 3).

3.5. Relationship of total elemental composition with mineralogy and other site characteristics

In order to confirm whether site or soil-forming factors (e.g., mineralogy, climate, topography and vegetation) are important drivers of total elemental concentrations in the soil, we performed a Random Forests (RF) regression of the factors against the first 5 principal components of the TXRF element concentration. The RF regression allowed us to view the importance of the predictor variables and below we discuss only the top four variables in decreasing order of importance for the first 3 principal components or dimensions which explained as much as 57% of the variability of the total element concentrations.

3.5.1. Elemental concentration fingerprints versus site factors

A regression of 6 predictor variables namely soil class, climate zones, cluster, plot flooding, depth and topographic position showed that the RF-OOB models had an overall accuracy (R^2) of 0.59, 0.44 and 0.43 for dimensions 1, 2 and 3 respectively (Fig. 4a). The most important predictor variables explaining the variation observed in the first principal component or dimension 1 of the total element concentration data, were soil class (based on the IUSS Working Group WRB (2014) and the FAO/EC/ISRIC (2003) world soil resources map (scale 1:30,000,000 approx)), climate zone (Köppen–Geiger climatic zones) (Kottek et al., 2006), topographic position followed by whether the plot is flooded or not (Fig. 4a). Soil class is very critical in determining the total elemental compositions of soils in Sub-Saharan Africa. Table 4 gives a breakdown of the mean total elemental composition by soil class for the sample suite used in the present study. Total elemental compositions varied widely according to soil types and the specific pedological processes operating in a given soil type that influence their distribution need to be taken into account to understand and quantify of the total elemental composition. However, there is need to take into consideration the uncertainties associated with spatial predictions using e.g., FAO/EC/ISRIC soil maps, which are major in many cases and could confound the analytical data. For the dimension 2, the most important variables were soil class, climate zone, topographic position, and plot flooding, while for dimension 3 the four most important variables were similar to those in the second dimension (Fig. 4a). When we excluded the climate zone

predictor variable in the regression model and in place of this included total annual precipitation and average monthly temperature, the regression results showed an overall improved fit (R^2) of the models' overall accuracy of 0.87, 0.74 and 0.70 for dimensions 1, 2 and 3 (Fig. 4b). In addition, the order of the most important predictor variables explaining the variation observed were also changed when climate zone was removed from the model such that for dimension 1 they were soil class, precipitation, average temperature, followed by topographic position (Fig. 4b). Topography is the most commonly considered soil-forming factor and expresses the variation of total element concentration pattern due to modification of the water relationships in soils and it influences soil erosion to a considerable extent (Jenny, 1941).

Since the soil class and climate zone ranked the top most important variables explaining the variation of the total elemental concentrations' main pattern of variation in Dim 1, these were both excluded in the RF model and the predictor variables precipitation, average temperature retained and consequently the model accuracies were slightly decreased to a R^2 of 0.79, 0.69, and 0.64 for dimensions 1, 2 and 3 respectively (Fig. 4c). For dimensions 1, 2 and 3, the two most important variables were then precipitation and temperature followed by position in the topography and cluster (Fig. 4c). When a new variable 'landuse' was introduced in addition to the soil classification according to the FAO/EC/ISRIC (2003) and IUSS Working Group WRB (2014), the overall model accuracies were further marginally improved to R^2 values of 0.87, 0.77 and 0.70 for dimensions 1, 2 and 3 respectively (Fig. 4d). Excluding the soil class from the model used in Fig. 4d the fit of the models in dimensions 1–3 was 0.81, 0.70 and 0.64 R^2 respectively (Fig. 4e). The top most important variable for all the dimensions was temperature followed by precipitation, topography and landuse in dimensions 1 and 2 while in dimension 3, after temperature, the most important variables were precipitation, landuse followed by topography (Fig. 4e). The results that landuse was one of the important variables, e.g., ranking fourth, fourth, and third in dimensions 1, 2 and 3 respectively (Fig. 4e), supported results by Voortman (2011) that the presence of different vegetation types is a reliable and precise indicator of differences in soil chemical properties and that interactions among nutrients significantly explain differences in vegetation and also the distribution of vegetation types. These results are interesting in that they imply that >70% of variation in soil elemental composition patterns can be predicted using information in existing databases or readily observable features. The predictor 'soil depth' was of least importance for total

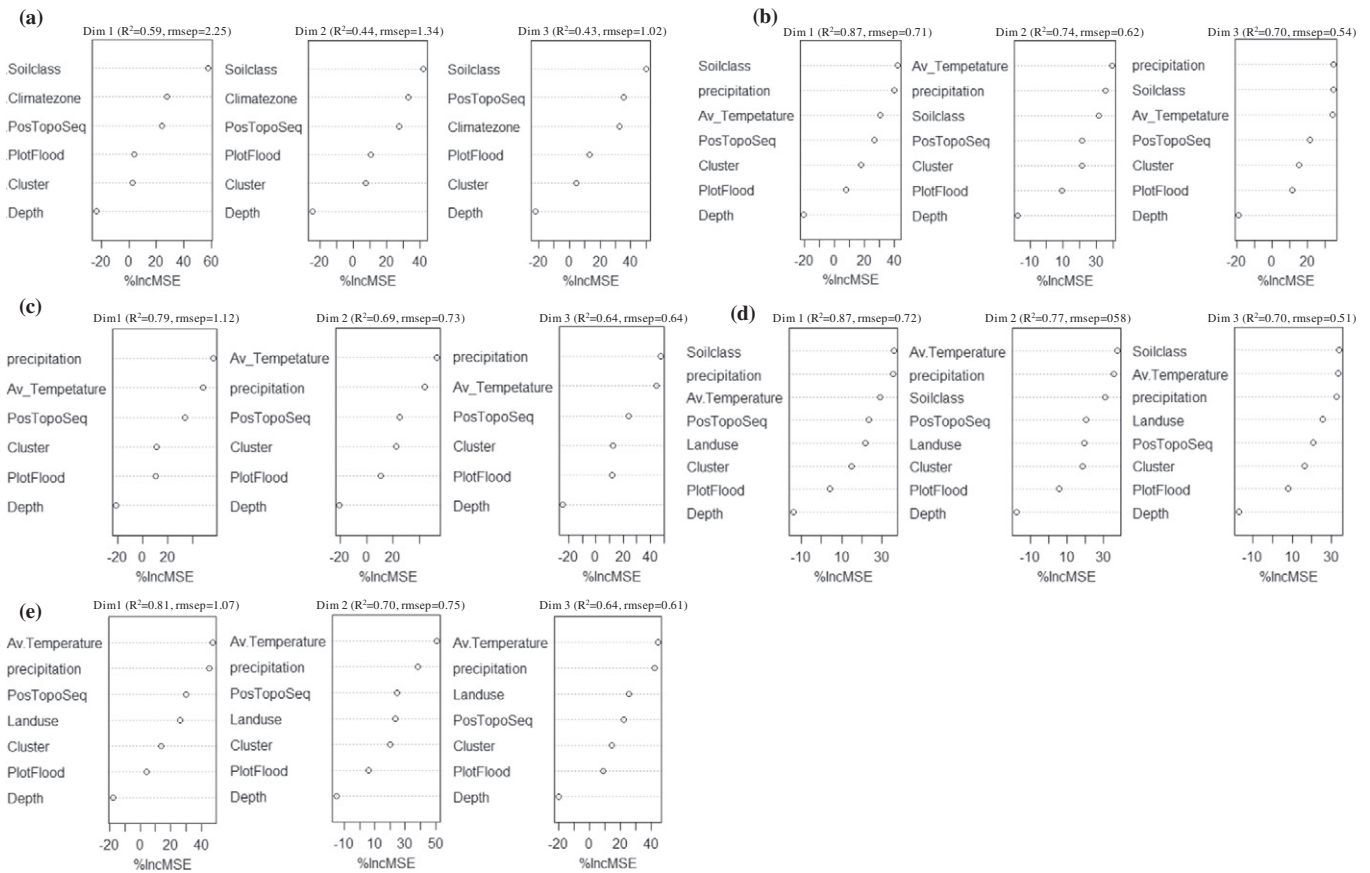


Fig. 4. Variable importance plots for the first 3 dimensions of the Random Forests regression of TXRF element concentrations against mineralogy and site/soil-forming factors showing the model accuracies and mean decrease in accuracy (%IncMSE) for site or soil-forming predictor variables (a) excluding precipitation and temperature (b) excluding climate zone but including precipitation and temperature (c) excluding climate zones and soil class but including precipitation and temperature (d) including soil classification and landuse (e) excluding climate zones and soil class but including precipitation, temperature and landuse (see Table 1 for predictor parameter description).

element concentration pattern of variation prediction, suggesting that neither top- nor subsoils influence much the variation of element concentration probably due to expected correlations of element concentrations between the two depths.

3.5.2. Element concentration versus mineralogy composition

Mineralogy data from XRD analysis were extracted for each site and for this study's analysis we used both raw semi-quantitative mineralogy data and the dominant mineralogy grouping. In general RF model

performances were acceptable with prediction accuracies for dimensions 1–3 ranging from 0.79 to 0.45 and 0.82–0.61 for raw mineralogy data and dominant mineralogy groupings respectively (Fig. 5a–b). These results suggest that patterns of variation in total element concentrations are largely due to high variability in mineralogy. The most important predictor variables explaining the variation in the dominant mineralogy grouping for dimension 1 were K-feldspars followed by kaolinite/1:1 clays, then quartz and plagioclase (Fig. 5a). Thus, since K, Ca, Ti, Fe and Sr were predominantly associated with Dim 1, the

Table 4
Breakdown of the mean total elemental composition (mg kg⁻¹) by soil class according to Table 1.

Element	Acrisols, Alisols, Plinthosols	Arenosols	Calcisol, Cambisol, Luvisol	Ferralsols	Ferralsols, Acrisols, Nitisols	Fluvisols, Gleysols, Cambisols	Leptosols	Lixisols	Nitisols, Andosols	Plinthosols	Vertisols
Al	31,993	19,136	33,338	42,542	40,940	12,312	66,925	33,678	50,440	32,143	13,305
P	42	43	246	210	47	44	44	81	620	47	48
K	9068	11,508	16,520	8717	9145	7041	3786	10,649	13,830	4114	8484
Ca	1747	3467	36,695	1440	1918	733	1055	3980	21,798	967	6312
Ti	2800	1343	3354	8564	3640	2327	8528	4340	9767	4282	871
V	19	5.3	16	30	25	18	110	49	75	45	12
Cr	71	39	56	53	45	44	244	64	68	86	13
Mn	282	192	620	578	341	114	321	451	1143	309	180
Fe	21,724	5395	23,579	34,957	23,754	11,294	68,615	29,981	54,288	30,601	8365
Ni	12	5.2	25	21	21	4.1	25	20	30	17	8.4
Cu	13	3.9	14	21	23	5.6	26	18	28	16	7.5
Zn	24	8.8	41	41	36	5.4	23	23	71	16	10
Ga	7.8	4.8	9.0	10	9.0	3.0	19	8.0	14	7.5	2.6
Sr	36	65	166	54	19	26	25	106	455	15	59
Y	7.1	15	20	10	11	11	10	11	26	12	3.4
Ta	2.7	0.6	2.5	3.5	5.7	1.7	5.0	3.2	5.6	3.3	1.1
Pb	28	10	20	36	23	17	108	46	62	41	14

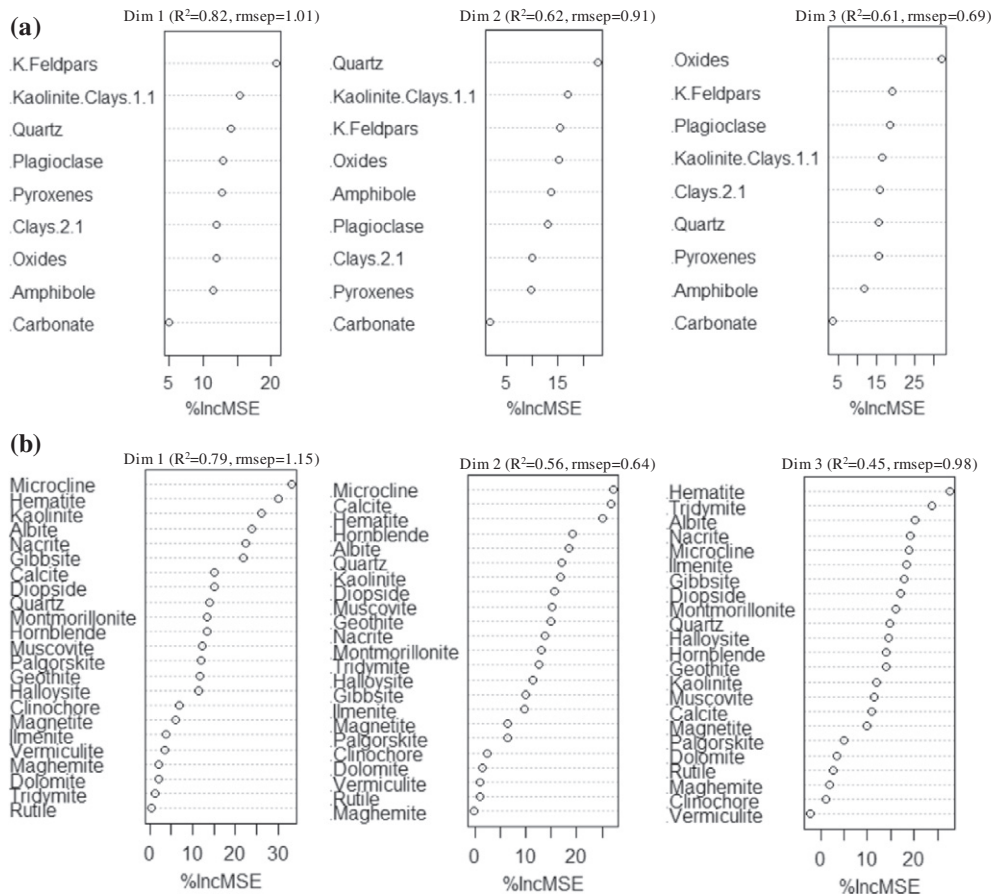


Fig. 5. Variable importance plots for the first 3 dimensions of the Random Forests regression of TXRF element concentrations against mineralogy and site/soil-forming factors showing the model accuracies and mean decrease in accuracy (%IncMSE) for (a) dominant mineralogy groupings and (b) raw semi-quant mineralogy data.

contents of these elements are presumed to be originating primarily from parent material as reported by Kabata-Pendias and Mukherjee (2007) and US.EPA (2006). Lower concentrations of certain elements, e.g., Ca and K in some soils in our study, could also be explained by a reported breakdown of primary minerals, particularly K-feldspars and plagioclase (Acosta et al., 2011) or depletion in well-drained soils over long periods of pedogenic weathering since they are either divalent or monovalent (Marques et al., 2004). The relationship between clay mineralogy composition and K forms and physico-chemical properties has also been demonstrated by several studies (Sharply, 1989; Surapaneni et al., 2002; Raheb and Heidari, 2012). Besides the clay content, K-bearing minerals and clay mineralogy may play a more important role in determining a soil's extractable K status and thus, the relationships between clay mineralogy and K can be used in evaluating potential soil K fertility, prediction of K cycling and plant uptake (Raheb and Heidari, 2012). Information on the K elemental composition along with knowledge of mineralogical clay composition can provide insights into release of K to plants and the need for K fertilizers. For dimension 2 the most important variables were quartz, kaolinite/1:1 clays, K-feldspars, and oxides (Fig. 5a). Because Al, Ni and Ga were dominant in Dim 2 (Fig. 3a, Annex Fig. A1), our results are in agreement with those reported by Acosta et al. (2011) that Al in soils has been attributed to the formation of clay minerals. Total Al in soil reflects the type of soil and the underlying geology, being present in the matrix of clays and other silicate minerals, and highly weathered soils are often high in Al concentrations (Rawlins et al., 2012). However, at the opposite extreme, low Al is a marker of organic-rich soils, which contain a smaller proportion of aluminosilicate minerals (Rawlins et al., 2012). For dimension 3 the most important variables were oxides followed by K-feldspars, plagioclase and kaolinite/1:1 clays (Fig. 5a). While our results showed

that Cr, Mn and Zn were predominant in the Dim 3 (Fig. 2b, Annex Fig. A1), Zn is reported to be generally associated with Al- and Fe-containing minerals such as feldspars, micas, pyroxenes and amphiboles (Acosta et al., 2011).

The most important predictor variables explaining the variation in the raw mineralogy for dimension 1 were microcline, followed by hematite, kaolinite and albite; for dimension 2 they were microcline, calcite, hematite and hornblende; while for dimension 3 they were hematite, tridymite, albite and nacrite (Fig. 5b). These findings were also in agreement with a report by Voortman (2011) that the nature of the key variables explaining the variability of soils are related to the mineralogy of parent rock and thus we also infer that the mineralogy of parent rock is a principal factor determining spatial patterns of soil total element concentrations. Furthermore there is ample evidence from the literature (e.g., Sharply, 1989; Surapaneni et al., 2002) that variation among soils in crop productivity and response to nutrients is primarily determined by soil mineral composition. Soil tests based on soil extracts do not characterize all the nutrient pools that determine a soil's ability to re-supply the soil solution (buffering capacity), such as readily mineralizable organic P, sorbed P, and slowly extractable K. Total X-ray fluorescence spectroscopy provides chemical fingerprinting and "functional mineral groupings" that could relate to potential nutrient supply capacity.

3.6. Element concentration versus mineralogy composition plus site and soil-forming factors

Since some site factors and mineralogy investigated in this study separately had a high explanatory power of the patterns of variations of the total elemental concentrations we tested how much variation

could be explained when they were combined. In general RF model performances were improved compared to separate models with prediction accuracies of 0.92, 0.84, and 0.79 for dimensions 1, 2 and 3 (Fig. 6a). The most important variables explaining the main patterns of variation in total element concentrations were cluster, landuse, topography, temperature and precipitation (Fig. 6a). Thus all soil-forming factors (e.g., parent material, climate, topography, management (landuse)) were demonstrated to have an important influence on total elemental concentrations in the soil. The importance of cluster can be explained by spatial correlation at distances of less than 1 km. Since cluster was the only non-readily observable variable, we tested whether predictive performance was reduced or not by leaving out cluster. In general, the model performances were slightly reduced with prediction accuracies of 0.90, 0.80 and 0.75 for dimensions 1, 2 and 3 respectively (Fig. 6b).

4. Conclusions

Information on the variation in soil element concentrations at different spatial scales is needed for, e.g., planning land use and environmental management. Knowledge of the concentration and spatial distribution of naturally occurring elements in the soils of Sub-Saharan Africa (SSA) is limited and there is a need for a better understanding of the factors that might regulate the variations. TXRF provides chemical fingerprinting for inferring soil chemical and physical functional properties which is of interest in the African soil contexts for agricultural and environmental management at large scales. The AfSIS baseline dataset of SSA provides an excellent base for studying chemical variations in soil composition at the continental scale related to factors such as mineralogy, climate, topography, vegetation and land use. Thus, this study helped to establish the baseline

concentrations of 17 elements for soils occurring within 34 sentinel sites across SSA and documented variability in total element concentrations within and between sites, which appeared to relate to differences in mineralogical ‘functional groups’. We observed strong within site and between site heterogeneity in many element compositions which were related to soil forming factors. The exploratory analyses of the relationships between element composition data and other site factors using Random Forests regressions have demonstrated that all site or soil-forming factors (e.g., mineralogy, climate, topography, vegetation and land use) have an important influence on total elemental concentrations in the soil. The fact that the soil-forming factors are related to the concentration of the naturally occurring elements in the soil gives rise to the notion that they might be predicted from the soils’ element composition. The results also implied that >70% of variation in soil element composition patterns can be predicted using information in existing databases or readily observable features. Thus, future studies should investigate the feasibility of quantitatively predicting soil functional properties from concentrations of elements and the existing databases or readily observable features e.g., for digital soil mapping.

Acknowledgments

This research work was financially supported by the World Agroforestry Centre (ICRAF) through the Africa Soil Information Service (AfSIS) project, funded by the Bill and Melinda Gates Foundation and Alliance for a Green Revolution in Africa (AGRA), and the Dr. Hermann Eiselen Ph.D. Grant from the Foundation Fiat Panis. E. Towett acknowledges the Young Excellence Scholars (YES) PhD Scholarship from the Food Security Centre (FSC) at the University of Hohenheim, Germany, and a travel grant as well as stipend within the framework of the “Global Food Security” funded by the Deutscher Akademischer Austausch

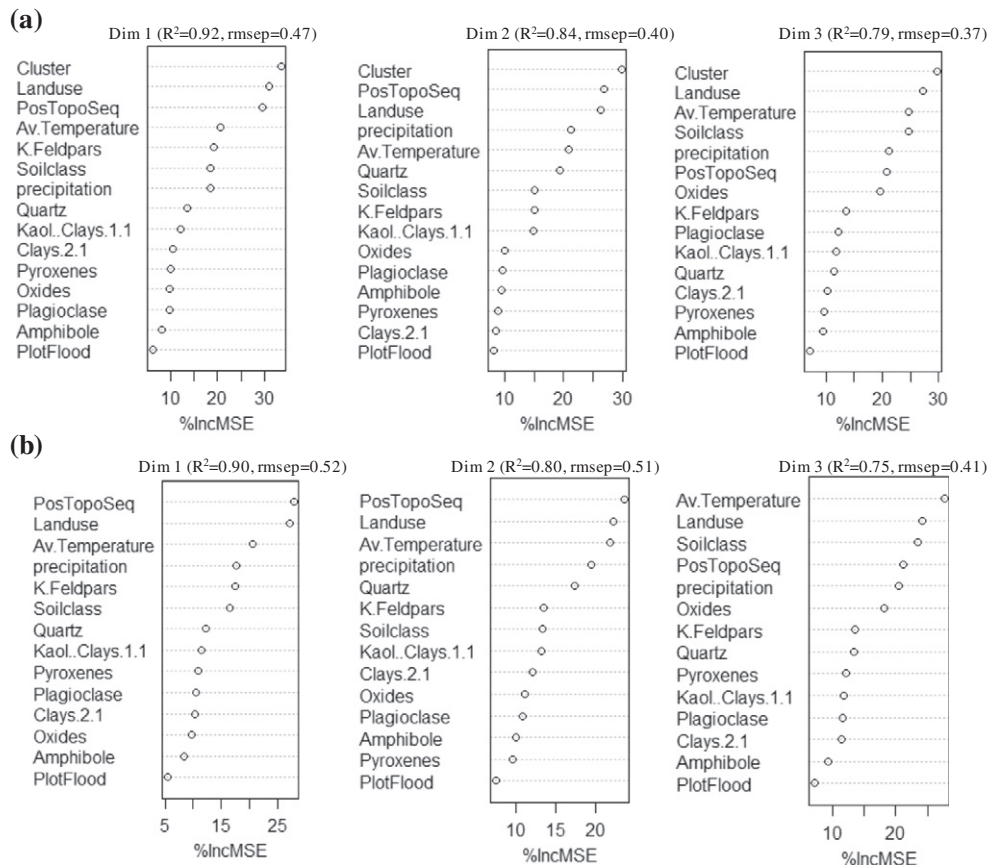


Fig. 6. Variable importance plots for the first 3 dimensions of the Random Forests regression of TXRF element concentrations against mineralogy and site/soil-forming factors showing the model accuracies and mean decrease in accuracy (%IncMSE) for combination of mineralogy and site factors (a) including cluster and (b) excluding cluster.

Dienst (DAAD). The authors acknowledge the regional AfSIS field crews who provided the soil samples, Robin Chacha and Hezekiah Nyandika for assistance with TXRF analyzing the soil samples, Beatrice Oware and Bella Kauma for assistance with XRD data semi-quantification and analysis, as well as other technical staff of the ICRAF Soil-Plant Spectral Diagnostics Laboratory for assistance with sample preparation. The authors are also grateful to Ric Coe, Anja Gassner, Hans-Peter Piepho, Joseph Ogutu, Hagen Stosnach, Michael Gatari, Ermias Aynekulu, Juan Laso, Keyfalew Sahle and Todd Rosenstock for their valuable advice and help with data analysis which helped improve this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:. These data include Google map of the most important areas described in this article.

References

- Acosta, J.A., Martínez-Martínez, S., Faz, A., Arocena, J., 2011. Accumulations of major and trace elements in particle size fractions of soils on eight different parent materials. *Geoderma* 161, 30–42.
- Adebisi, F.M., Olise, F.S., Owoade, O.K., Obiah, I.B., Ajayi, T.R., Olaniyi, H.B., Asubiojo, I.O., 2005. TXRF characterization of EDTA extractable metals from soils of bituminous sands occurrence area, Agbabu, South Western Nigeria. *J. Appl. Sci.* 5 (7), 1151–1157.
- Antwi-Agyei, P., Hogarh, J.N., Foli, G., 2009. Trace elements contamination of soils around gold mine tailings dams at Obuasi, Ghana. *Afr. J. Environ. Sci. Technol.* 3 (11), 353–359.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45 (1), 5–32.
- Bruker, 2007. S2 Picofox User Manual. Doc-M81-EXX001-V1. Bruker AXS Microanalysis GmbH, Berlin, Germany.
- Cannon, W.F., Horton, J.D., 2009. Soil geochemical signature of urbanization and industrialization — Chicago, Illinois, USA. *Appl. Geochem.* 24, 1590–1601.
- Chapman, P.M., Wang, F., Janssen, C.R., Goulet, R.R., Kamunde, C.N., 2003. Conducting ecological risk assessments of inorganic metals and metalloids: current status. *Hum. Ecol. Risk Assess.* 9 (4), 641–697.
- FAO/EC/ISRIC, 2003. World Soil Resources Map.
- Grunsky, E.C., Drew, L.J., Sutphin, D.M., 2009. Process recognition in multi-element soil and stream-sediment geochemical data. *Appl. Geochem.* 24, 1602–1616.
- Haluschak, P., Eilers, R.G., Mills, G.F., Grift, S., 1998. Status of selected trace elements in agricultural soils of southern Manitoba. Technical Report 1998-6E Land Resource Unit. Brandon Research Centre, Research Branch, Agriculture and Agri-Food Canada.
- Helmke, P.A., 2000. The chemical composition of soils. In: Sumner, M.E. (Ed.), *Handbook of Soil Science*. CRC Press, Boca Raton, FL, pp. B3–B24.
- Herselman, J.E., 2007. The concentration of selected trace metals in South African soils (PhD Thesis). University of Stellenbosch, South Africa.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Hooda, P.S., 2010. Trace Elements in Soils. John Wiley and Sons, UK.
- Jenny, H., 1941. *Factors of Soil Formation: A System of Quantitative Pedology*. McGraw-Hill, New York.
- Kabata-Pendias, A., Mukherjee, A.B., 2007. *Trace Elements from Soil to Human*. Springer-Verlag, Berlin Heidelberg.
- Klockenkämper, R., 1997. *Total Reflection X-ray Analysis*. Wiley, New York.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen–Geiger climate classification updated. *Meteorol. Z.* 15, 259–263. <http://dx.doi.org/10.1127/0941-2948/2006/0130>.
- Laiho, R., Penttilä, T., Laine, J., 2004. Variation in soil nutrient concentrations and bulk density within peatland forest sites. *Silva Fenn.* 38 (1), 29–41.
- Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* 25 (1), 1–18.
- Liaw, A., Wiener, M., 2002. Classification and regression by Random Forest. *R News* 2 (3), 18–22.
- Luo, X., Yu, S., Zhu, Y., Li, X., 2012. Trace metal contamination in urban soils of China. *Sci. Total Environ.* 421–422, 17–30.
- Marques, J.J., Schulze, D.G., Curia, N., Mertzman, S.A., 2004. Trace element geochemistry in Brazilian Cerrado soils. *Geoderma* 121, 31–43.
- Marzecová, A., Mikomägia, A., Koff, T., Martmac, T., 2011. Sedimentary geochemical response to human impact on Lake Nõmmejärv, Estonia. *Est. J. Ecol.* 60 (1), 54–69. <http://dx.doi.org/10.3176/eco.2011.1.05>.
- Negrel, P., Sadeghi, M., Ladenberger, A., Clemens, R., Birke, M., 2014. Geochemical fingerprinting and source discrimination of agricultural soils at continental scale. *Chem. Geol.* <http://dx.doi.org/10.1016/j.chemgeo.2014.12.004>.
- Qian, J., Shan, X., Wang, Z., Tu, Q., 1996. Distribution and plant availability of heavy metals in different particle-size fractions of soil. *Sci. Total Environ.* 187, 131–141.
- Raheb, A., Heidari, A., 2012. Effects of clay mineralogy and physico-chemical properties on potassium availability under soil aquic conditions. *J. Soil Sci. Plant Nutr.* 12 (4), 747–761.
- Rawlins, B.G., Webster, R., Tye, A.M., Lawley, R., O'hara, S.L., 2009. Estimating particle-size fractions of soil dominated by silicate minerals from geochemistry. *Eur. J. Soil Sci.* 60, 116–126.
- Rawlins, B.G., McGrath, S.P., Scheib, A.J., Breward, N., Cave, M., Lister, T.R., Ingham, M., Gowing, C., Carter, S., 2012. The Advanced Soil Geochemical Atlas of England and Wales. British Geological Survey, Keyworth (www.bgs.ac.uk/gbase/advsoilatlasEW.html).
- Reeves III, J.B., Smith, B.D., 2009. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Appl. Geochem.* 24 (8), 1472–1481.
- Rodríguez, J.A., Nanos, N., Grau, J.M., Gil, L., López-Arias, M., 2008. Multiscale analysis of heavy metal contents in Spanish agricultural topsoils. *Chemosphere* 70, 1085–1096.
- Rousseau, R.M., 2001. Detection limit and estimate of uncertainty of analytical XRF results. *Rigaku J.* 18 (2), 33–47.
- SAS Institute, 2011. SAS 9.3 Software. SAS Institute Inc., USA.
- Sharp, N., 1989. Relationship between soil potassium forms and mineralogy. *Soil Sci. Soc. Am. J.* 52, 1023–1028.
- Shepherd, K.D., 2010. Soil spectral diagnostics — infrared, X-ray and laser diffraction spectroscopy for rapid soil characterization in the Africa Soil Information Service. 19th World Congress of Soil Science, Soil Solutions for a Changing World 1–6 August 2010, Brisbane, Australia.
- Shepherd, K.D., Walsh, M.G., 2007. Infrared spectroscopy — enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *J. Near Infrared Spectrosc.* 15, 1–20.
- Shepherd, K.D., Shepherd, G., Walsh, M.G., 2015. Land health surveillance and response: a framework for evidence-informed land management. *Agric. Syst.* 132, 93–106.
- Stosnach, H., 2005. Environmental trace-element analysis using a benchtop total reflection X-ray fluorescence spectrometer. *Anal. Sci.* 21, 873–876.
- Surapaneni, A., Palmer, A.S., Tillman, R.W., Kirkman, J.H., Gregg, P.E.H., 2002. The mineralogy and potassium supplying power of some loessial and related soils of New Zealand. *Geoderma* 110, 191–204.
- Towett, E.K., Shepherd, K.D., Cadisch, G., 2013. Quantification of total element concentrations in soils using total X-ray fluorescence spectroscopy (TXRF). *Sci. Total Environ.* 463–464, 374–388. <http://dx.doi.org/10.1016/j.scitotenv.2013.05.068>.
- US EPA, 2006. Innovative Technology Verification Report. XRF Technologies for Measuring Trace Elements in Soil and Sediment, Rontec Picotax XRF Analyzer. United States Environmental Protection Agency, Office of Research and Development, Washington DC (EPA/540/R-06/005).
- Vägen, T.-G., Walsh, M.G., 2012. Sentinel site surveillance in Segou Region, Mali: an evidence-based approach to assessing land degradation and targeting sustainable land management interventions. *Land Health Surveillance — An Evidence-Based Approach to Land Ecosystem Management*. United Nations Environment Programme (UNEP), pp. 115–171.
- Vägen, T.-G., Winowiecki, L.A., 2013. Mapping of soil organic carbon stocks for spatially explicit assessments of climate change mitigation potential. *Environ. Res. Lett.* 8 (2013), 015011. <http://dx.doi.org/10.1088/1748-9326/8/1/015011> (9 pp.).
- Vägen, T.-G., Shepherd, K.D., Walsh, M.G., Winowiecki, L.A., Desta, L.T., Tondoh, J.E., 2010. Africa Soil Information Service (AfSIS) Technical Specifications — Soil Health Surveillance [Version 1.0].
- Vägen, T.-G., Davey, F., Shepherd, K.D., 2012. Land health surveillance: mapping soil carbon in Kenyan rangelands. In: Nair, P.K.R., Garrity, D. (Eds.), *Agroforestry — The Future of Global Land Use*. Springer, pp. 455–462.
- Vägen, 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.
- Vägen, T.-G., Winowiecki, L.A., Tamene Desta, L., Tondoh, J.E., 2015. The Land Degradation Surveillance Framework (LDSF) — Field Guide v4_1. World Agroforestry Centre, Nairobi, Kenya.
- Voortman, R.L., 2011. Explorations into African land resource ecology on the chemistry between soils, plants and fertilizers (PhD dissertation). Faculty of Economics and Business, Vrije Universiteit, p. 264.
- Winowiecki, L., Vägen, T.-G., Huisling, J., 2015. Effects of land cover on ecosystem services in Tanzania: a spatial assessment of soil organic carbon. *Geoderma* <http://dx.doi.org/10.1016/j.geoderma.2015.03.010>.
- IUSS Working Group WRB, 2014. World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.