

Rothamsted Repository Download

A - Papers appearing in refereed journals

Winowiecki, L. A., Vagen, T. G., Boeckx, P. and Dungait, J. A. J. 2017. Landscape-scale assessments of stable carbon isotopes in soil under diverse vegetation classes in East Africa: application of near-infrared spectroscopy. *Plant and Soil*. 421 (1-2), pp. 259-272.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1007/s11104-017-3418-3>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/8v5yv>.

© 16 October 2017, Rothamsted Research. Licensed under the Creative Commons CC BY.

Landscape-scale assessments of stable carbon isotopes in soil under diverse vegetation classes in East Africa: application of near-infrared spectroscopy

Leigh Ann Winowiecki · Tor-Gunnar Vågen ·
Pascal Boeckx · Jennifer A. J. Dungait

Received: 26 October 2016 / Accepted: 8 September 2017 / Published online: 16 October 2017
© The Author(s) 2017. This article is an open access publication

Abstract

Aims Stable carbon isotopes are important tracers used to understand ecological food web processes and vegetation shifts over time. However, gaps exist in understanding soil and plant processes that influence $\delta^{13}\text{C}$ values, particularly across smallholder farming systems in sub-Saharan Africa. This study aimed to develop predictive models for $\delta^{13}\text{C}$ values in soil using near infrared spectroscopy (NIRS) to increase overall sample size. In addition, this study aimed to assess the $\delta^{13}\text{C}$ values between five vegetation classes.

Methods The Land Degradation Surveillance Framework (LDSF) was used to collect a stratified random set of soil samples and to classify vegetation. A total of 154 topsoil and 186 subsoil samples were collected and analyzed using NIRS, organic carbon (OC) and stable carbon isotopes.

Results Forested plots had the most negative average $\delta^{13}\text{C}$ values, -26.1‰ ; followed by woodland, -21.9‰ ; cropland, -19.0‰ ; shrubland, -16.5‰ ; and grassland,

-13.9‰ . Prediction models were developed for $\delta^{13}\text{C}$ using partial least squares (PLS) regression and random forest (RF) models. Model performance was acceptable and similar with both models. The root mean square error of prediction (RMSEP) values for the three independent validation runs for $\delta^{13}\text{C}$ using PLS ranged from 1.91 to 2.03 compared to 1.52 to 1.98 using RF.

Conclusions This model performance indicates that NIR can be used to predict $\delta^{13}\text{C}$ in soil, which will allow for landscape-scale assessments to better understand carbon dynamics.

Keywords Carbon cycling · Landscape scale assessments · Random forest modeling

Abbreviations

$\delta^{13}\text{C}$	Delta carbon 13
BT	Boosted regression trees
ICRAF	World Agroforestry Centre
IRMS	isotope ratio mass spectrometry
LCCS	Land Cover Classification System
LDSF	Land Degradation Surveillance Framework
NIRS	Near infrared spectroscopy
MARS	Multivariate adaptive regression splines
MIRS	Mid-infrared spectroscopy
MPLS	Modified partial least squares
PLS	Partial least squares regression
RF	And random forest
RMSEP	Root mean square error of prediction
SOC	Soil organic carbon
SSA	Sub-Saharan Africa

Responsible Editor: Elizabeth M Baggs

L. A. Winowiecki (✉) · T.-G. Vågen
World Agroforestry Centre (ICRAF), Nairobi, Kenya
e-mail: l.a.winowiecki@cgiar.org

P. Boeckx
Isotope Bioscience Laboratory – ISOFYS, Ghent University,
Coupure Links 653, 9000 Gent, Belgium

J. A. J. Dungait
Rothamsted Research, Sustainable Soils and Grassland Systems
Department, North Wyke, Okehampton, Devon EX20 2SB, UK

UNCCD	The United Nations Convention to Combat Desertification
UNFCCC	United Nations Framework Convention on Climate Change

Introduction

Aboveground vegetation influences belowground carbon dynamics. Optimizing soil organic carbon (SOC) content is recognized as an essential component of ecosystem functioning (Lal 2010; Palm et al. 2007; Vågen et al. 2012). The United Nations Convention to Combat Desertification (UNCCD) and the United Nations Framework Convention on Climate Change (UNFCCC) both recognize that reduced SOC content is a consequence of, and can lead to further land degradation, and ultimately poor land and agricultural productivity. However, understanding the influence of vegetation classes on SOC is still needed, especially in light of progressive degradation of soil and water resources (Vågen et al. 2013a; Vågen and Gumbritsch 2012; Verchot et al. 2005). Although SOC is almost universally proposed as the most important soil quality indicator (Amundson et al. 2015; Gregorich et al. 1994), the complexity and extent of SOC dynamics at the landscape scale is still poorly understood. This includes, but is not limited to, understanding the influence of inherent soil properties (e.g. geochemistry, aggregation, texture, etc.) on SOC content as well as the effects of aboveground vegetation types, land management and climate. Furthermore, the impacts of land-use change on SOC dynamics in sub-Saharan African (SSA) ecosystems are still understudied, especially across diverse landscapes, but essential if food production is to keep pace with predicted population growth in the region (Rosegrant and Cline 2003). Assessing the impact of vegetation shifts on SOC dynamics and quantifying SOC turnover rates can improve our understanding of the effects of land-use change from native vegetation to agricultural food production (Schlesinger 1997), as well as the impacts of management shifts introduced by climate smart agriculture (Lipper 2014; Rwehumbiza 2014) and sustainable agricultural intensification (Vanlauwe et al. 2014) on soil health in smallholder farming systems.

Stable carbon isotopes in soil

Stable carbon isotopes are important tracers used to understand ecological food web processes and vegetation shifts over time. This is because, the majority of plants (trees and broad-leaved crops) use the C3 photosynthetic pathway and have $\delta^{13}\text{C}$ values between -22 and -30‰ , while about 15% of plants use the C4 photosynthetic pathway and have less negative $\delta^{13}\text{C}$ values, generally ranging from -10 to -14‰ (Farquhar 1989; Loomis and Connor 1992). The latter includes the majority of tropical herbs and grasses, including maize which is the major crop grown in many of our study areas.

Understanding differences in photosynthetic pathways is important in the assessment of SOC dynamics, including SOM turnover rates and carbon cycling (Accoe et al. 2002; Bernoux et al. 1998; Ehleringer et al. 2000; Six and Jastrow 2002); to identify vegetative sources of organic matter in the soil (Boutton et al. 1998; Von Fischer and Tieszen 1995; Kindscher and Tieszen 1998; Krull et al. 2006; Puttock et al. 2014, 2012; Roscoe et al. 2001); to address the impact of land conversion on soil condition (Awiti et al. 2008; Schulp and Veldkamp 2008; Vågen et al. 2006b) and to improve the overall understanding of ecosystem function (Staddon 2004). In addition to vegetative shifts, there are several other factors that influence $\delta^{13}\text{C}$ values in soil, including (i) microbial decomposition, (ii) the Suess effect (Balesdent and Mariotti 1996; Roscoe et al. 2001), and (iii) inherent factors such as soil texture and geochemistry (e.g., quantity of iron and aluminum oxides) (Krull et al. 2002; Krull and Skjemstad 2003; Powers and Schlesinger 2002). There are still gaps in our knowledge regarding stable carbon isotope signatures in soils under in diverse systems, due in part to the associated costs, infrastructure requirements and sample preparation time required for stable carbon isotope analysis, which inhibit landscape-scale assessments. However, the interest in stable isotopes is increasing, as is its utility across disciplines. For example, the spatial patterns of vegetative signatures in soil have been mapped using stable carbon isotopes (Boeckx et al. 2006; Wynn and Bird 2007), $\delta^{13}\text{C}$ values in the soil profile have been used as a proxy for SOC stability (Oelbermann and Voroney 2006; Salomé et al. 2010), to trace and quantify erosion (Häring et al. 2013), and $\delta^{13}\text{C}$ values in sediments have been used to determine vegetative sources in depositional environments (Puttock et al. 2012).

Compound-specific stable carbon isotope values are increasingly applied to understand different soil processes using a range of organic compounds with contrasting chemistries used as biomarkers (Dungait et al. 2008, 2010). The increase in the use of stable carbon isotopes has created a demand for new, more rapid analysis of plants, soils and sediments in order to obtain the larger datasets needed for landscape-scale ecological assessments.

Use of near-infrared spectroscopy (NIRS) to predict soil properties

Infrared (IR) is now a well established methodology for the prediction of soil properties such as SOC, pH, base cations and texture (Brown 2007; Brown et al. 2006; Genot et al. 2011; Nocita et al. 2014; Stenberg et al. 2010; Vasques et al. 2009; Viscarra Rossel et al. 2006). In addition to being cost effective, IR spectroscopy allows for estimation of several soil characteristics simultaneously, with minimal sample preparation and no use of chemicals (Brown et al. 2006; Vågen et al. 2006a, b; Terhoeven-Urselmans et al. 2010; Genot et al. 2011). The ability to distinguish between different properties of a wide range of materials using NIR has resulted in a growing field of research across several disciplines, including chemometrics, forage science, soil science and plant science. In soil science, the application of NIR has resulted in a significant lowering of costs associated with measurements of soil properties, which has in turn resulted in significant advances when it comes to landscape-scale assessments of soil.

Soil spectroscopy is the “reflectance part of the electromagnetic radiation that interacts with the soil matter across the VIS-NIR-SWIR spectral region” (Ben-Dor and Banin 1995). Specifically, spectra in the near infrared (NIR) range (wavelengths 8000–4000 cm^{-1}), can be analyzed to characterize the chemical, physical and mineralogical composition of the soil (Ben-Dor and Banin 1995; Stoner and Baumgardner 1982; Viscarra Rossel et al. 2006). NIRS are influenced not only by the chemistry of the soil but also by its physical structure, making individual (well-defined and narrow) absorption bands at specific wavelengths less pronounced. As NIR absorbance features occur due to both overtones and combination bands of fundamental vibrations of OH, CH, NH, CO, CN and NO bonds in the mid

infrared region, the absorbance of light is directly related to frequency or wavelength and corresponds to the difference in energy between two vibrational states (quantum numbers) in molecular bonds. These energy levels are also influenced by surrounding molecules and functional groups, for example, but fundamentally various substances and molecules can be identified due to different absorption patterns in the spectra. For example, spectral regions around 7000 cm^{-1} , 5200 cm^{-1} and 4460 cm^{-1} are particularly important for the prediction of SOC (Ben-Dor and Banin 1995). Iron oxides are often represented in the adsorption bands at less than 1000 nm, hydroxyl bonds near 1400 and 1900 nm, clay mineral absorb near 2200, and organic matter absorbs at various wavelengths throughout NIR spectrum (Soriano-Disla et al. 2014; Viscarra Rossel et al. 2016). Few studies have reported important spectral regions for the prediction of other soil properties than SOC. The difference between the application of spectroscopy in soil science versus for example in food sciences, is that that soil properties do not exhibit strong peaks at particular wavelengths, calibration models with reference datasets containing wet chemistry analysis is needed to develop robust predictive models.

The World Agroforestry Centre (ICRAF), with headquarters in Nairobi, Kenya has established a soil IR spectral database that currently has soil spectra from soil samples from across a wide range of landscapes that represent agricultural soils, forested landscapes, wetlands, and savannas (Brown 2007; Towett et al. 2015). These landscapes also represent a wide range of climatic conditions, from sub-humid and humid ecosystems to semi-arid and arid ecosystems in the global tropics. This soil IR database combined with associated reference wet chemistry enables the application of data mining and analysis techniques to explore the potential of IR to predict soil properties and even indices of soil condition (Vågen et al. 2006a, b).

Other studies have assessed the potential for NIR to predict stable carbon isotopes in soil, as it is plausible that NIR spectra should be able to detect the differences in the atomic mass of the carbon isotope (Kleinebecker et al. 2009). For example, Fuentes et al. (2009, 2012) explored the possibility to predict $\delta^{13}\text{C}$ values in soil using NIR spectra from soils collected from a well documented experimental station in Mexico ($n = 100$ soil samples). Using modified partial least squares (MPLS) regression,

they developed calibration models with an R^2 of 0.81 (Fuentes et al. 2012, 2009). Fuentes et al. (2009) used Mahalanobis distance to characterize and discard spectra from the population of soil samples, hence attempting to reduce variability in the soil samples used in the study. While the results presented in Fuentes et al. (2009) were promising, and demonstrated the potential for using NIR to predict $\delta^{13}\text{C}$, the study represented a dataset with very limited variations in $\delta^{13}\text{C}$ and separate models were developed for soils with plant residues ($n = 50$ range -16.6 to -23.3‰) and without plant residues ($n = 50$ range -19.1 to -22.1‰), which limits the application of these models beyond the specific case study they were developed for. Our study builds on this work to explore the potential for NIR to predict $\delta^{13}\text{C}$ values across a varied set of soils that represent a wide range of chemical and physical characteristics, including different vegetation classes. Kleinebecker et al. (2009) used NIR spectra to predict $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in plant tissues. Applying partial least squares (PLS) regression they developed calibration models obtaining positive results with R^2 values of 0.89 and 0.99 for $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$, respectively (Kleinebecker et al. 2009). While NIR has long been used to determine protein quality in forages (Marten et al. 1983), Clark et al. (1995) assessed the potential to determine carbon isotope composition using NIR in various genotypes and cultivars of forage species in the USA. Their study used PLS regression models for each cultivar and obtained R^2 values between 0.69 to 0.93 for $\delta^{13}\text{C}$ calibration models, further highlighting the utility of NIR to predict stable carbon isotopic content (Clark et al. 1995) in plant material.

Recent advances in big data analytics and ensemble learning methods allow for the development of predictive models that are stable across a range of soil functional operating ranges. In this paper we present a case study where we use NIR spectroscopy of soils to develop predictive models for SOC and $\delta^{13}\text{C}$, exploring the application of this approach in the scaling of soil analysis to landscape level assessments of soil health. Specific objectives of this study include: 1) Assess the potential for NIRS to predict $\delta^{13}\text{C}$ using a diverse set of soil samples; 2) Compare different predictive models (e.g., PLS and RF) and 3) Better understand stable carbon isotope variation across various vegetation classes and smallholder farming systems in SSA.

Materials and methods

Soil sampling

Biophysical field surveys and soil sampling were conducted in nine-100 km² sites using the Land Degradation Surveillance Framework (LDSF). Top ($n = 156$) and sub ($n = 184$) soil samples from nine LDSF sites across Ethiopia (Mega), Kenya (Ollentille, Mpala, Kipsing), Democratic Republic of Congo (DRC) (Luhhi, Burhale), Uganda (Hoima), Madagascar (Didy) and Tanzania (Mbola) were included in the study (Fig. 1). The LDSF uses a spatially stratified random sampling design (Vågen et al. 2013b) with 160 sampling plots, each 1000 m², across 16 spatially stratified clusters (10 plots in each cluster), with 4 subplots (100 m²) within each sampling plot. Measurements and observations were made at the subplot and plot levels, respectively. Land use was classified at each plot using a simplified version of the FAO Land Cover Classification System (LCCS) (Di Gregorio and Jansen 1998), into three distinct vegetation classes: 1) primarily vegetated, forest; (2) primarily vegetated, woodland; (3) primarily vegetated, shrubland; (4) primarily vegetated, grassland; and (5) primarily vegetated, cropland (Di Gregorio and Jansen 1998). Altitude was recorded at the plot level. Tree counts were made at the subplot level and then averaged at the plot level. Soil samples were collected from each of the 160 plots by compositing soil samples from the four subplots within each plot for topsoil (0–20 cm) and subsoil (20–50 cm). For most of the sites, we analyzed soil samples from one reference plot from each cluster, for a total of 16 topsoil samples and 16 subsoil samples (unless there were depth restrictions), providing a total of 30 soil samples from Luhhi, Burhale, and Ollentille, 31 soil samples from Mpala, 32 soil samples from Kipsing, Mega and Hoima. However, two reference plots per cluster were used for Didy and Mbola, providing 64 and 59 soil samples from these sites, respectively.

Laboratory analysis

Soil samples were air-dried and sieved to 2 mm. Air-dried soil samples were scanned in duplicate in near-infrared spectral range (wavelengths between 8000 to 4000 cm⁻¹) with a resolution of 4 cm⁻¹ using a Bruker Multipurpose Analyzer (MPA) at the World Agroforestry Centre (ICRAF) Plant and Soil Spectroscopy

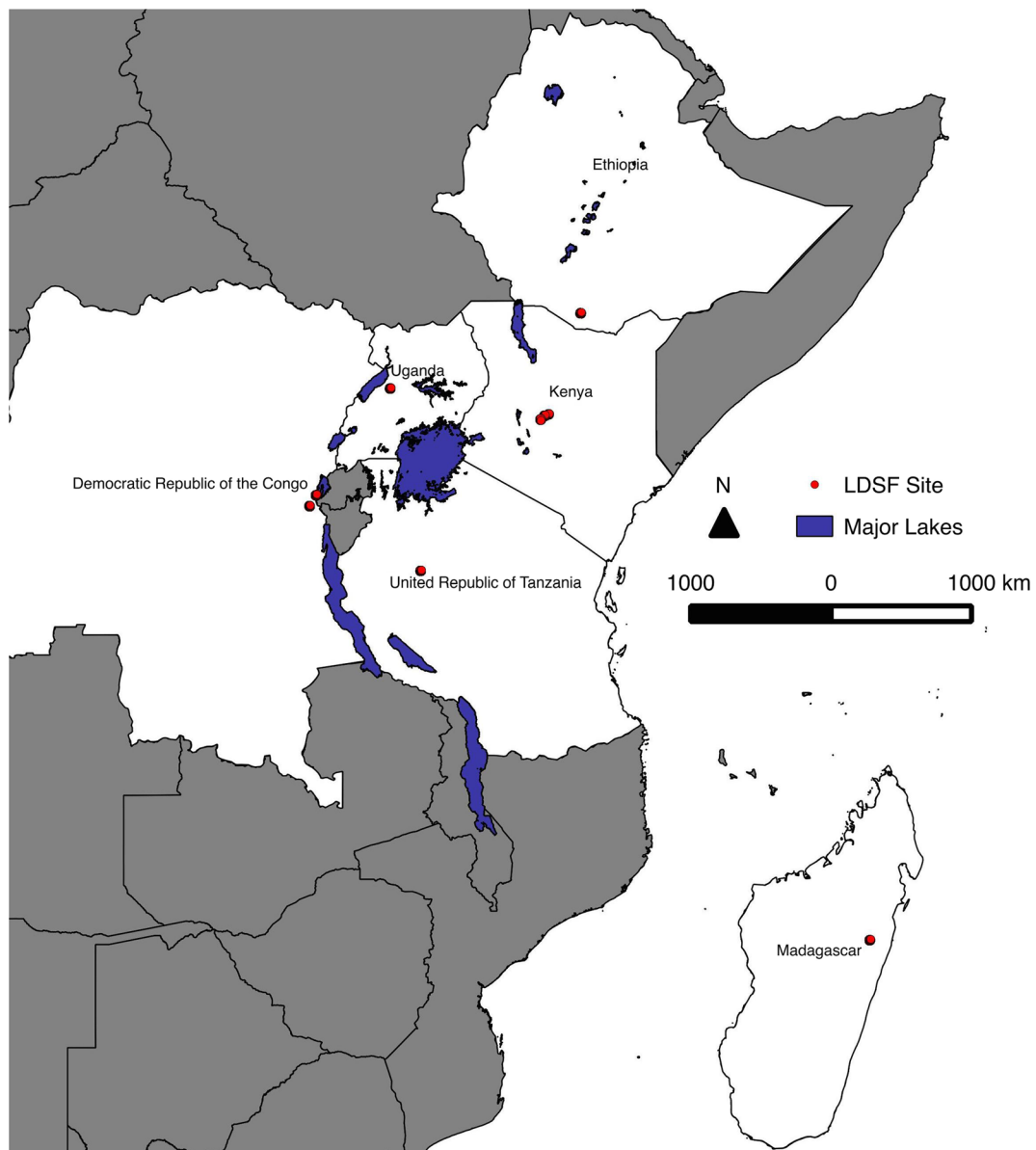


Fig. 1 Location of the nine LDSF sites used in the study (red circles)

Laboratory in Nairobi, Kenya (<http://worldagroforestry.org/research/land-health/spectraldiagnostics-laboratory>).

Soil samples from all locations except DRC were analyzed for carbon concentration (% dry mass) using dry combustion on acidified samples and for stable carbon isotopes with an elemental analyzer isotope ratio mass spectrometer (EA-IRMS) at IsoAnalytics Laboratory (<http://www.iso-analytical.co.uk>). Soil samples from DRC were analyzed at Isotope Bioscience Laboratory (ISOFYS, www.ISOFYS.be) of Ghent

University, Belgium using EA-IRMS (ANCA-SL (SerCon, Crew, UK), coupled to a 2020 IRMS (SerCon, Crew, UK)). Stable carbon isotopes were expressed as $\delta^{13}\text{C}$ in parts per mile (‰) relative to the V-PDB (Pee Dee Belemnite) standard (Loomis and Connor 1992).

NIR processing and prediction of SOC and $\delta^{13}\text{C}$

All calculations and statistical analysis were performed using R statistics (R Core Team 2015) and KNIME (Berthold et al. 2007). Manipulation of the NIR spectra

included computing the first derivatives of the spectra using a Savitsky-Golay polynomial smoothing filter implemented in the *locpoly* function of the *KernSmooth* R package (Wand 2015). Partial Least Squares (PLS) regression analysis was conducted in R using the *pls* package and the *mvr* function (Mevik et al. 2015). A Random Forest (RF) model (Breiman 2001) was computed in R using the *randomForest* package (Liaw and Wiener 2002), while statistical analysis of between site differences was conducted using linear mixed effects models using the *nlme* package in R (Pinheiro et al. 2013).

Prediction models for $\delta^{13}\text{C}$ were developed using PLS regression (Martens and Naes 1989), which is commonly used as a standard tool in chemometrics (Wold et al. 2001). In brief, PLS is a dimension reduction technique similar to classical canonical correlation analysis (CCA), but where covariance is maximized rather than correlation (Boulesteix and Strimmer 2007). The X- and Y-scores are chosen so that the relationship between successive pairs of scores is as strong as possible, similar to a robust form of redundancy analysis. Directions are sought in the factor space that are associated with high variation in the responses but biasing them toward directions that are accurately predicted (Tobias 1995). The PLS model was compared to a RF (Breiman 2001) prediction model. Random forests have a wide range of applications, both in classification and regression, and are increasingly used for multivariate calibration, including in soil science (Vågen et al. 2013a; Winowiecki et al. 2016a). An ensemble of 500 regression trees was built, where each tree was learned on a different set of observations in the input data and different combinations of NIR spectral wavebands.

Results

Description of sites

Basic site characteristics for the nine sites included in the study are shown in Table 1. The sites ranged in elevation from about 970 m for Didy in Madagascar to about 1800 m for Ol Lentille, which is located in Laikipia County in central Kenya (Table 1). The sites ranged from wet tropical forests in Madagascar to dry-land savanna in Kenya (Mpala, Ol Lentille and Kipsing), but also included agriculturally dominated areas such as Burhale, Hoima, Luhihi, and Mbola. There were large variations in land use between the sites, with 69% of the sampled plots under cultivation in Burhale, compared to Mpala, Ol Lentille and Kipsing where there was no cultivation (Table 1). The latter represent shrubland/rangeland systems. In addition, we calculated the average tree densities for each site. Didy had the highest tree density (3910 trees ha^{-1}) compared with Mega, which had the lowest (13 trees ha^{-1}) (Table 1).

NIR spectra, soil organic carbon (SOC) and carbon isotopic signatures

There was substantial variation in soil NIR spectra both between and within the sites (Fig. 2). For example, Luhihi topsoil samples had high variation in absorbance within the site, especially compared to Didy, Mbola and Kipsing. The spectra of both top- and subsoil samples were used to develop calibration and validation models for $\delta^{13}\text{C}$ in soil. Given the diversity of the spectra, the potential for applying these models across diverse landscapes is high.

Table 1 Basic biophysical characteristics of the nine sites included in the study

Site	Average elevation (m)	% of the plots that were cultivated	Average tree density (tree ha^{-1})	Dominant Vegetation Classes
Burhale, DRC	1595	69	23	Cropland/ Woodland
Didy, Madagascar	973	5	3910	Forest
Hoima, Uganda	1168	41	41	Cropland/Shrubland
Kipsing, Kenya	1219	0	111	Shrubland
Luhihi, DRC	1653	68	85	Cropland/ Woodland
Mbola, Tanzania	1196	39	611	Cropland/Woodland
Mega, Ethiopia	1548	0.6	13	Shrubland/Grassland
Mpala, Kenya	1714	0	87	Shrubland/Grassland
Ol Lentille	1807	0	65	Shrubland

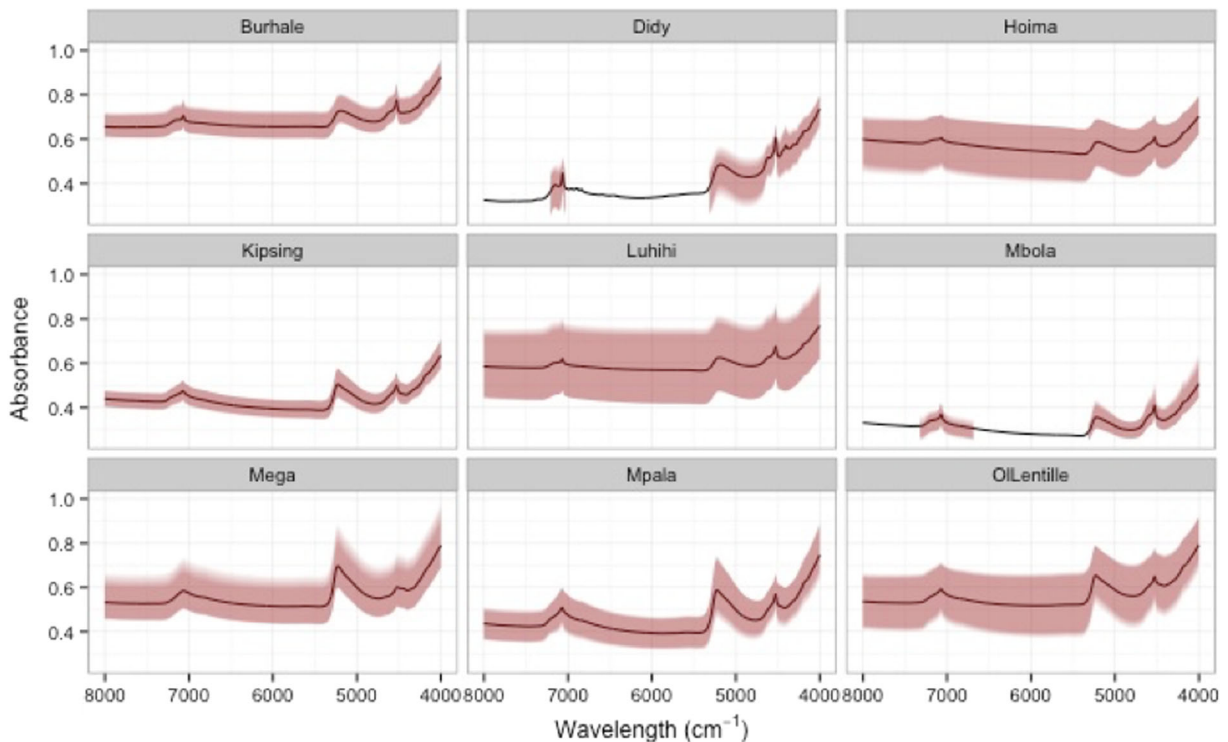


Fig. 2 NIR spectra for the 156 topsoil samples from each of the nine LDSF sites. The black line is the mean spectra for the site and the shaded area is the standard deviation

Soil organic carbon varied across the nine sites from 0.9 to 98.3 g kg⁻¹. Mean topsoil OC across the sites was 18.9 g kg⁻¹ and mean subsoil OC was 12.5 g kg⁻¹. Figure 3 shows boxplots for topsoil and subsoil for each site (e.g., the median, 25th and 75th percentiles). Kipsing had the lowest median topsoil OC (3.8 g kg⁻¹), followed by Mbola (4.1 g kg⁻¹), OilLentille (8.3 g kg⁻¹), Mpala (11.0 g kg⁻¹), Mega (17.9 g kg⁻¹), Burhale (25.0 g kg⁻¹), Hoima and Didy (29.0 g kg⁻¹), and then Luhihi (33.5 g kg⁻¹) (Fig. 3). A comparison of the boxplots per site, not only indicate with sites had the highest SOC content, but also, which sites had the greatest variation within the site. For example, Didy had the greatest difference between top- and subsoil OC. Luhihi, Hoima, and Burhale had the highest variability in both top and subsoil SOC values within the site. Kipsing and Mbola had the smallest variation of SOC within the site, despite the variation in vegetation classes represented in Mbola.

Average $\delta^{13}\text{C}$ in topsoil was -18.8‰ , and average $\delta^{13}\text{C}$ in subsoil was -19.4‰ , which indicates that these are mixed C3-C4 ecosystems, with the exception of Didy, which is dominated by C3 vegetation (forest). Didy topsoil had the most negative $\delta^{13}\text{C}$ values with a

median of -27.04‰ , followed by Mbola (-21.77‰), Kipsing (-19.00‰), Luhihi (-18.68‰), Hoima (18.27‰), Mpala (-16.50‰), Burhale (-16.42‰), OilLentille (-15.70‰), and Mega (-13.60‰) (Fig. 4). Furthermore, there was very little difference between top and sub soil $\delta^{13}\text{C}$ values (with the exception of Didy), which may have implications for whether or not the organic matter is at steady state. However, the lack of a strong shift in isotopic signature with depth at most sites, could be due to the sampling intervals (0–20 and 20–50 cm), compared to sampling strategies that use smaller increments, or it could indicate that their have not been recent vegetative shifts.

Vegetation structure class and $\delta^{13}\text{C}$ values

Fifty-six plots were classified as forest, 41 plots as woodland, 135 plots as shrubland, 30 plots as grassland and 78 plots as cropland. Forested plots had the most negative average $\delta^{13}\text{C}$ values (combining top and sub soil values), indicating a dominance of C3 vegetation, -26.1‰ ; followed by woodland, -21.9‰ ; cropland, -19.0‰ ; shrubland, -16.5‰ ; and finally grassland, -13.9‰ (Fig. 5). In general, the $\delta^{13}\text{C}$ values for forested

Fig. 3 Boxplots of the top- and subsoil SOC variation for each of the nine LDSF sites. The black dotted vertical line is the mean topsoil OC across the sites, 18.9 g kg^{-1} and the gray dotted vertical line is the mean subsoil OC, 12.5 g kg^{-1}

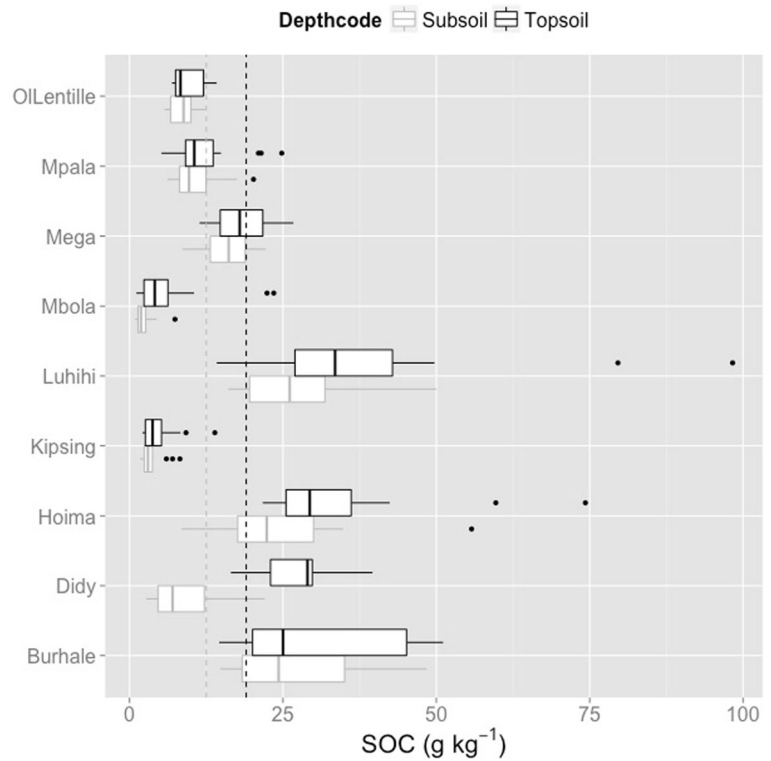


Fig. 4 Boxplot of $\delta^{13}\text{C}$ values for the top- and subsoil samples for each of the LDSF nine sites. The black dotted vertical line is the mean $\delta^{13}\text{C}$ across the sites, -18.8 ‰ and the gray dotted vertical line is the mean subsoil $\delta^{13}\text{C}$, -19.4 ‰

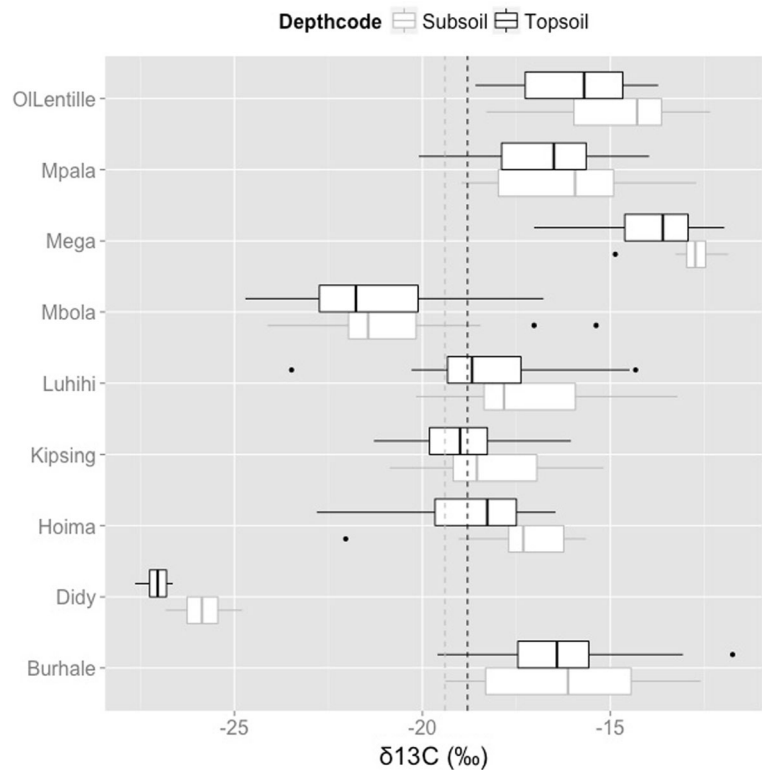
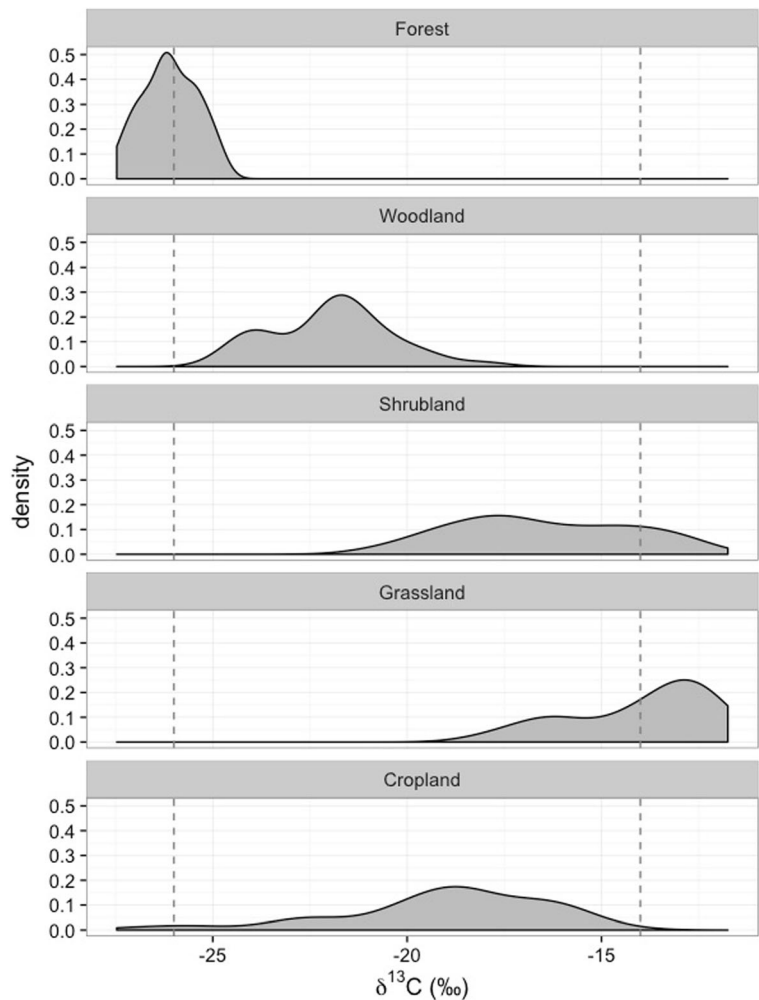


Fig. 5 Density plots of $\delta^{13}\text{C}$ values for the top- and subsoil samples for each of the five Land Cover Classifications (Forest ($n = 56$), Woodland ($n = 41$), Shrubland ($n = 135$), Grassland ($n = 30$), Cropland ($n = 78$)). The black dotted vertical lines are average $\delta^{13}\text{C}$ for C3 vegetation, -14‰ and for C4 vegetation, -26‰



plots exhibited a C3 signature, while grassland plots exhibited a C4 signature. In contrast, the remaining vegetation classes exhibited a mixed C3-C4 signature. Results of the linear mixed effects models demonstrated that compared to the forest $\delta^{13}\text{C}$ values, only shrubland ($p < 0.05$) and grassland ($p < 0.001$) plots showed significant difference. These data have important implications for the use of stable carbon isotopes in East Africa, e.g., since both semi-natural and cropland systems have mixed C3-C4 signatures large sample sizes will be needed to develop robust models to assess vegetation shifts and soil organic matter turnover rates.

NIR prediction results for $\delta^{13}\text{C}$ values

Prediction performance for $\delta^{13}\text{C}$ was similar for the PLS and RF models when tested on three different validation datasets. Root Mean Square Error of Prediction

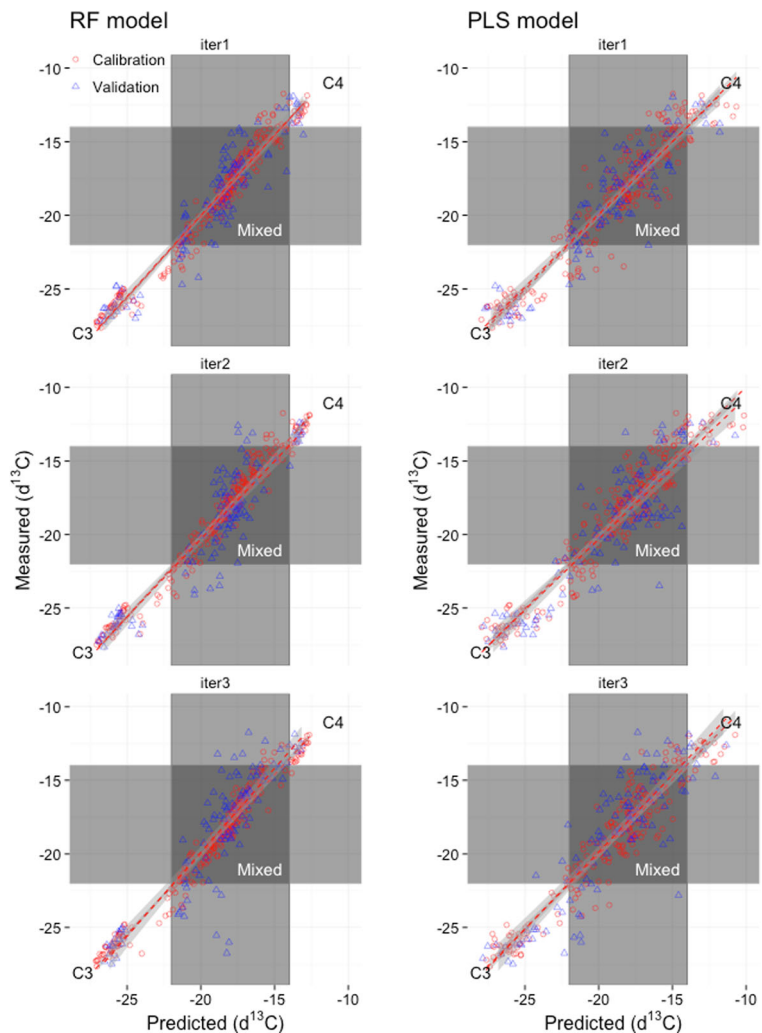
(RMSEP), which is a useful measure of accuracy reflecting the overall difference between measured and predicted values, was low for both PLS and RF models (1.95 for PLS and 1.77 for RF). The results further show average R^2 for the validation runs using PLS of 0.80 compared to a slightly higher average R^2 using RF of 0.84. Overall, R^2 values were slightly higher for calibration runs compared to validation runs, which is expected, with average R^2 for calibration using PLS of 0.91 compared to 0.97 for RF. It is important to test prediction models such as the ones presented here on datasets that are independent of that used for developing the model in order to generate some measure of model stability. In our case, validation samples were drawn randomly in each iteration and the calibration model was fitted to this dataset with low RMSEP in both cases indicating good performance for both models (Table 2). Model stability was also good for both PLS and RF

Table 2 Prediction performance for $\delta^{13}\text{C}$ for three cross-validation (CV) runs for the calibration and validation datasets for the partial least squares (PLS) and random forest (RF) models, expressed as the root mean squared error of prediction (RMSEP) and R^2

CV run	Dataset	n	$RMSEP_{\delta^{13}\text{C}}$ RF	$R^2_{\delta^{13}\text{C}}$ RF	$RMSEP_{\delta^{13}\text{C}}$ PLS	$R^2_{\delta^{13}\text{C}}$ PLS
1	Calibration	226	0.92	0.97	1.33	0.90
	Validation	114	1.52	0.88	1.91	0.83
2	Calibration	227	0.86	0.97	1.24	0.91
	Validation	113	1.98	0.85	2.03	0.81
3	Calibration	227	0.87	0.97	1.34	0.91
	Validation	113	1.83	0.78	1.92	0.76

models (Fig. 6). Given the stability of the models, indicated by the similarity between the slopes of the regression lines in Fig. 6 and low RMSEP for the

calibration and validation runs, respectively, there is significant potential for the application of RF models to large spectral libraries.

Fig. 6 Measured vs. predicted $\delta^{13}\text{C}$ values for the three calibration (red open circles) and validation (blue open triangles) runs using the Random Forest (RF) model (left panel) and the Partial Least Squares (PLS) regression model (right panel)

Discussion

There is a need to better understand the global distribution of C3 and C4 plants for a number of different reasons, including for improving global circulation models for CO₂, better assessing soil organic matter dynamics, as well as estimating water and energy cycles. Further, there is a need for better estimates of the spatial distribution of and temporal changes in plant communities with different photosynthetic pathways as plant communities respond differently to rising CO₂ levels in the atmosphere, as well as to land degradation status. The nine sites used in this study had diverse vegetation classes and cropping systems, which is reflected in the wide range of SOC values (e.g., from 0.9 to 98.3 g kg⁻¹). We found the highest SOC concentrations in Didy (eastern Madagascar), Burhale and Luhihi (eastern DRC) and Hoima (western Uganda), which all represent sub-humid and humid environments. The lowest SOC values were found in semi-arid environments (Ol Lentille, Mpala, Mega) and in Miombo woodlands (Mbola).

There are many factors influencing SOC concentrations, including inherent soil properties such as sand content, land degradation status such as erosion prevalence (Winowiecki et al. 2016a, 2016b), as well as land management practices (including burning in semi-natural and cropland systems, fertilization, residue retention, among others) (Vanlauwe et al. 2015). The use of stable carbon isotopes can aid in better understanding SOC dynamics and the influence of vegetation shifts over time. Our results show that semi-natural woodland and shrubland systems, as well as cropland systems in East Africa had mixed C3-C4 signatures, while wet tropical forest plots in Madagascar exhibited a strong C3 signature and tropical grassland systems exhibited a C4 signature, as expected. Given that woodlands and shrublands contain both woody vegetation and grasses, such as in Miombo woodland systems with deciduous trees and an understory of tall perennial grasses, explains the mixed carbon isotopic signatures across most of the sites. Furthermore, woodland fragmentation is often driven by several competing activities including agricultural expansion, demand for forest resources, grazing, charcoal production, firewood collection and shifting cultivation practices (King and Campbell 1994; Sauer and Abdallah 2007; Syampungani et al. 2009). In croplands where farmers predominantly plant C4 species such as maize, we also saw a stronger C4 signature,

depending on factors such as time since conversion in the case of areas that have been converted from natural forest. Examples of such sites in our study included Burhale in DRC and Hoima in Uganda. However, in cropland systems where tobacco, rice, soybean and maize are cultivated, the carbon signature is mixed, which increases the complexity in assessing the impacts on vegetation shifts using stable carbon isotopes.

In a chronosequence study from Madagascar, Vågen et al. (2006a) found a strong relationship between SOC and $\delta^{13}\text{C}$, with decreasing carbon going from C3-dominated systems (e.g., tropical forest) to C4-dominated vegetation (e.g., degraded grasslands and croplands) along a conversion gradient (Vågen et al. 2006a). However, in the current study we observe more mixed results and a low level of correlation between SOC and $\delta^{13}\text{C}$ overall, partly due to the inclusion of more diverse (mixed) systems. In addition to conversion studies, Billings and Richter (2006) highlight the need for decadal studies in somewhat stable systems to better identify discrimination processes, again illustrating the complexity of stable isotopic pathways (Billings and Richter 2006).

Based on the validation predictions, model performances for predicting $\delta^{13}\text{C}$ from NIR spectra were good both for RF (average RMSEP = 1.78) and PLS (average RMSEP = 1.95) models. Average R² values for the three model validation runs were 0.84 and 0.80 for the RF and PLS models, respectively. Our results show that both PLS, which is a data-reduction technique, and RF, which is generally good for feature selection, can be successfully applied for the prediction of $\delta^{13}\text{C}$ in soils. Further, the high performance of the RF model applied in our study indicates that this technique is suitable for detecting spectral features that are important for determining the relative abundance of ¹²C and ¹³C in soils. Viscarra Rossel and Behrens (2010) also evaluated a number of data mining techniques with NIR spectra, however they reported that PLS and support vector machines (SVM) outperformed RF, multivariate adaptive regression splines (MARS) and boosted regression trees (BT) (Viscarra Rossel and Behrens 2010). Exploratory application of remote sensing to map carbon isotopes was used in southern Africa (mixed C3-C4 systems), highlighting the need and potential solution to better understand ecosystem processes at larger spatial scales (Wang et al. 2010). Our results are similar to those reported by Fuentes et al. (2012), and given that we develop models across such a wide range of sites and

spectral variation shows the potential of NIR spectroscopy for routine prediction of $\delta^{13}\text{C}$, including for the direct classification of soil samples into C3, Mixed or C4 carbon. Furthermore, we assessed the relationship between SOC and $\delta^{13}\text{C}$, and found that it was not linear. This confirms that the prediction accuracy of both the PLS and RF models are because we are able to predict $\delta^{13}\text{C}$ (e.g., the weight of the isotope) and not merely predicting SOC. This further highlights the promise of spectroscopy for predicting stable isotopes in soil and shows the potential of new technological advances such as soil spectroscopic techniques for prediction of $\delta^{13}\text{C}$, which can lower the costs of analytical procedures and hence enable larger sample sizes and landscape-scale assessments of vegetation dynamics and SOC.

Building on these results, the potential use of mid-infrared spectroscopy (MIRS) should also be explored and is recommended by the authors for future studies. These results have important implications for the use of stable carbon isotopes in East Africa since both semi-natural and cropland systems have mixed C3-C4 signatures. Hence, large sample sizes will be needed to develop robust models to assess vegetation shifts and soil organic matter turnover rates. In conclusion, application of spectroscopic techniques would allow for more cost-effective analysis and increased sample sizes, which are needed for landscape-scale ecological studies.

Acknowledgements This work was supported by the Bill and Melinda Gates Foundation (BMGF) grant numbers 51353 and 5949 MDG MGATE3.02 254SAR, the CGIAR research program on Forests Trees and Agroforestry (FTA), the CGIAR research program on Climate Change, Agriculture and Food Security (CCAFS), International Fund for Agricultural Development (IFAD), grant number 2000000520 and Wajibu MS.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Accoe F, Boeckx P, Van Cleemput O et al (2002) Evolution of the $\delta^{13}\text{C}$ signature related to total carbon contents and carbon decomposition rate constants in a soil profile under grassland. *Rapid Commun Mass Spectrom* 16:2184–2189
- Amundson R, Berhe AA, Hopmans JW, Olson C, Sztein AE, Sparks DL (2015) Soil and human security in the 21st century. *Science* 348: 1–6. <https://doi.org/10.1126/science.1261071>
- Awiti AO, Walsh MG, Shepherd KD, Kinyamario J (2008) Soil condition classification using infrared spectroscopy : a proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma* 143:73–84. <https://doi.org/10.1016/j.geoderma.2007.08.021>
- Ben-Dor E, Banin A (1995) Near infrared analysis (NIRA) as a rapid method to simultaneously evaluate several soil properties. *Soil Sci Soc Am J* 59:364–372
- Bernoux M, Cerri C, Arrouays D et al (1998) Bulk densities of Brazilian Amazon soils related to other soil properties. *Soil Sci Soc Am J* 62:743. <https://doi.org/10.2136/sssaj1998.03615995006200030029x>
- Berthold MR, Cebon N, Dill F et al (2007) KNIME: The {K}onstanz {I}nformation {M}iner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds) *Data Anal. Mach. Learn. Appl.* 31st Annu. Conf. Gesellschaft für Klassif. e.V., Albert-Ludwigs-Universität Freiburgand. Springer, Berlin, pp 319–326
- Billings SA, Richter D (2006) Changes in stable isotopic signatures of soil nitrogen and carbon during 40 years of forest development. *Ecology* 148:325–333. <https://doi.org/10.1007/s00442-006-0366-7>
- Boeckx P, Van Meirvenne M, Raulo F, Van Cleemput O (2006) Spatial patterns of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in the urban topsoil of Gent, Belgium. *Org Geochem* 37:1383–1393
- Boulesteix A, Strimmer K (2007) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform* 8:32–44. <https://doi.org/10.1093/bib/bbl016>
- Boutton TW, Archer SR, Midwood AJ et al (1998) $\delta^{13}\text{C}$ values of soil organic carbon their use in documenting vegetation change in a subtropical savanna ecosystem. *Geoderma* 82: 5–41
- Breiman L (2001) Random forests. *Mach Learn* 45:35. <https://doi.org/10.1023/A:1010933404324>
- Brown DJ (2007) Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma* 140:444–453. <https://doi.org/10.1016/j.geoderma.2007.04.021>
- Brown DJ, Shepherd KD, Walsh MG et al (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132:273–290. <https://doi.org/10.1016/j.geoderma.2005.04.025>
- Clark DH, Johnson DA, Kephart KD, Jackson NA (1995) Near infrared reflectance spectroscopy estimation of $\delta^{13}\text{C}$ discrimination in forages. *J Range Manag* 48:132–136
- Di Gregorio A, Jansen LJM (1998) *Land Cover Classification System (LCCS): Classification concepts and user manual*. Environment and Natural Resources Service, GCP/RAF/287/ITA Africover - East Africa Project and Soil Resources, Management and Conservation Service. FAO, Rome, p 157
- Dungait JAJ, Docherty G, Straker V, Evershed RP (2008) Interspecific variation in bulk tissue, fatty acid and monosaccharide $\delta^{13}\text{C}$ values of leaves from a mesotrophic grassland plant community. *Phytochemistry* 69:2041–2051. <https://doi.org/10.1016/j.phytochem.2008.03.009>
- Dungait JAJ, Bol R, Lopez-Capel E et al (2010) Applications of stable isotope ratio mass spectrometry in cattle dung carbon cycling studies. *Rapid Commun Mass Spectrom* 24:495–500

- Ehleringer JR, Buchmann N, Flanagan LB (2000) Carbon isotope ratios in belowground carbon cycle processes.pdf. *Ecol Appl* 10:412–422
- Farquhar G (1989) Carbon isotope discrimination and photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol* 40:503–537. <https://doi.org/10.1146/annurev.arplant.40.1.503>
- Farquhar GD, Ehleringer JR, Hubick KT (1989) Carbon isotope discrimination and photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol* 40:503–537. <https://doi.org/10.1146/annurev.pp.40.060189.002443>
- Fuentes M, González-Martín I, Hernández-Hierro JM et al (2009) The natural abundance of ^{13}C with different agricultural management by NIRS with fibre optic probe technology. *Talanta* 79:32–37. <https://doi.org/10.1016/j.talanta.2009.03.002>
- Fuentes M, Hidalgo C, Gonzalez-Martin I et al (2012) NIR spectroscopy: an alternative for soil analysis. *Commun Soil Sci Plant Anal* 43:346–356. <https://doi.org/10.1080/00103624.2012.641471>
- Genot V, Colinet G, Bock L et al (2011) Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. *J Near Infrared Spectrosc* 19: 117. <https://doi.org/10.1255/jnirs.923>
- Gregorich EG, Carter MR, Angers DA et al (1994) Towards a minimum data set to assess soil organic matter quality in agricultural soils. *Can J Soil Sci* 74:367–385
- Häring V, Fischer H, Cadisch G, Stahr K (2013) Improved $\delta^{13}\text{C}$ method to assess soil organic carbon dynamics on sites affected by soil erosion. *Eur J Soil Sci* 64:639–650. <https://doi.org/10.1111/ejss.12060>
- Kindscher K, Tieszen LL (1998) Floristic and soil organic matter changes after five and thirty-five years of native tallgrass prairie restoration. *Restor Ecol* 6:181–196. <https://doi.org/10.1046/j.1526-100X.1998.06210.x>
- King JA, Campbell BM (1994) Soil organic matter relations in five land cover types in the Miombo region (Zimbabwe). *For Ecol Manag* 67:225–239
- Kleinebecker T, Schmidt SR, Fritz C et al (2009) Prediction of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ in plant tissues with near-infrared reflectance spectroscopy. *New Phytol* 184:732–739. <https://doi.org/10.1111/j.1469-8137.2009.02995.x>
- Krull ES, Skjemstad JO (2003) ^{13}C and ^{15}N profiles in 14 C-dated Oxisol and Vertisols as a function of soil chemistry and mineralogy. *Geoderma* 112:1–29
- Krull ES, Bestland EA, Gates WP (2002) Soil organic matter decomposition and turnover in a tropical Ultisol: evidence from $\delta^{13}\text{C}$, $\delta^{15}\text{N}$ and geochemistry. *Radiocarbon* 44:93–112
- Krull ES, Bestland EA, Skjemstad JO, Parr JF (2006) Geochemistry ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$, ^{13}C NMR) and residence times (^{14}C and OSL) of soil organic matter from red-brown earths of South Australia: Implications for soil genesis. *Geoderma* 132:344–360. <https://doi.org/10.1016/j.geoderma.2005.06.001>
- Lal R (2010) Enhancing eco-efficiency in agro-ecosystems through soil carbon sequestration. *Crop Sci* 50:S-120–S-131. <https://doi.org/10.2135/cropsci2010.01.0012>
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2:18–22
- Lipper L, Thornton P, Campbell BM, Baedeker T, Braimoh A, Bwalya M, Caron P, Cattaneo A, Garrity D, Henry K, Hottle R, Jackson L, Jarvis A, Kossam F, Mann W, McCarthy N, Meybeck A, Neufeldt H, Remington T, Sen PT, Sessa R, Shula R, Tibu A, Torquebiau EF (2014). Climate-smart agriculture for food security. *Nat Clim Chang* 4. <https://doi.org/10.1038/nclimate2437>
- Loomis RS, Connor DJ (1992). Crop ecology: productivity and management in agricultural systems. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Marten GC, Halgerson J, Cherney J (1983) Quality prediction of small grain forages by near infrared reflectance spectroscopy. *Crop Sci* 23:94–96
- Martens H, Naes T (1989). Multivariate calibration. John Wiley & Sons, Chichester, p 438
- Mevik BH, Wehrens R, Liland KH (2015) Rpackage: pls, Partial Least Squares and Principal Component Regression. <https://cran.r-project.org/package=pls>. Accessed on Feb 5, 2016.
- Nocita M, Stevens A, van Wesemael B et al (2014) Soil spectroscopy: an opportunity to be seized. *Glob Chang Biol*:1–2. <https://doi.org/10.1111/gcb.12632>
- Oelbermann M, Voroney RP (2007) Carbon and nitrogen in a temperate agroforestry system: Using stable isotopes as a tool to understand soil dynamics. *Ecol Eng* 29:342–349. <https://doi.org/10.1016/j.ecoleng.2006.09.014>
- Palm C, Sanchez P, Ahamed S, Awiti A (2007) Soils: a contemporary perspective. *Annu Rev Environ Resour* 32:99–129. <https://doi.org/10.1146/annurev.energy.31.020105.100307>
- Pinheiro J, Bates D, DebRoy S, Sarkar D (2017) nlme: Linear and nonlinear mixed effects models. R package version 3.1–131. <https://CRAN.R-project.org/package=nlme>
- Powers JS, Schlesinger WH (2002) Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica. *Atlantica* 109:165–190
- Puttock A, Dungait JAJ, Bol R et al (2012) Stable carbon isotope analysis of fluvial sediment fluxes over two contrasting C4–C3 semi-arid vegetation transitions. *Rapid Commun Mass Spectrom* 26:2386–2392. <https://doi.org/10.1002/rcm.6257>
- Puttock A, Dungait JAJ, Macleod CJA et al (2014) Woody plant encroachment into grasslands leads to accelerated erosion of previously stable organic carbon from dryland soils. *J Geophys Res Biogeosci* 119:2345–2357. <https://doi.org/10.1002/2014JG002635>. Received
- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing
- Roscoe R, Buurman P, Velthorst EJ, Vasconcellos CA (2001) Soil organic matter dynamics in density and particle size fractions as revealed by the $^{13}\text{C}/^{12}\text{C}$ isotopic ratio in a Cerrado's oxisol. *Geoderma* 104:185–202.
- Rosegrant MW, Cline SA (2003) Global Food Security: Challenges and Policies. *Science* 302, 5652:1917–1919
- Rwehumbiza FBR (2014) A comprehensive scoping and assessment study of climate smart agriculture (CSA) policies in Tanzania. Morogoro
- Salomé C, Nunan N, Pouteau V et al (2010) Carbon dynamics in topsoil and in subsoil may be controlled by different regulatory mechanisms. *Glob Chang Biol* 16:416–426. <https://doi.org/10.1111/j.1365-2486.2009.01884.x>
- Sauer J, Abdallah JM (2007) Forest diversity, tobacco production and resource management in Tanzania. *Forest Policy Econ* 9: 421–439. <https://doi.org/10.1016/j.forpol.2005.10.007>

- Schlesinger WH (1997) Biogeochemistry: an analysis of global change, 2nd edn. Academic Press, San Diego
- Schulp CJE, Veldkamp A (2008) Geoderma long-term landscape – land use interactions as explaining factor for soil organic matter variability in Dutch agricultural landscapes. *GSA Today* 146:457–465. <https://doi.org/10.1016/j.geoderma.2008.06.016>
- Six J, Jastrow JD (2002) Organic matter turnover. *Encycl Soil Sci*: 936–942
- Soriano-Disla JM, Janik LJ, Viscarra-Rossel RA et al (2014) The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl Spectrosc Rev* 49:139–186
- Staddon PL (2004) Carbon isotopes in functional soil ecology. *Trends Ecol Evol* 19:148–154
- Stenberg B, Viscarra Rossel RA, Mouazen AM, Wetterlind J (2010) Visible and near infrared spectroscopy in soil science. *Adv. Agron.*, 1st ed. Elsevier Inc, pp 163–215
- Stoner E, Baumgardner MF (1982) Characteristic variations in reflectance of surface soils. *Soil Sci Soc Am J* 45:1161–1165
- Sympungani S, Chirwa PW, Akinnifesi FK et al (2009) The miombo woodlands at the cross roads : potential threats, sustainable livelihoods, policy gaps and challenges. *Nat Res Forum* 33:150–159
- Terhoeven-Urselmans T, Vagen T-G, Spaargaren O, Shepherd KD (2010) Prediction of Soil Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library. *Soil Sci Soc Am J* 74(5):1792
- Tobias RD (1995) An introduction to partial least squares regression. *Proc. Ann. SAS users gr. Int. Conf.*, 20th, Orlando, FL, pp 2–5
- Towett EK, Shepherd KD, Sila A et al (2015) Mid-infrared and Total X-ray fluorescence spectroscopy complementarity for assessment of soil properties. *Soil Sci Soc Am J*. <https://doi.org/10.2136/sssaj2014.11.0458>
- Vågen T-G, Gumbritch T (2012) Sahel atlas of changing landscapes: tracing trends and variations in vegetation cover and soil condition, 1st edn. UNEP, Nairobi
- Vågen T-G, Shepherd KD, Walsh MG, Va T (2006a) Sensing landscape level change in soil fertility following deforestation and conversion in the highlands of Madagascar using Vis-NIR spectroscopy. *Geoderma* 133:281–294. <https://doi.org/10.1016/j.geoderma.2005.07.014>
- Vågen T, Walsh MG, Shepherd KD (2006b) Stable isotopes for characterisation of trends in soil carbon following deforestation and land use change in the highlands of Madagascar. *Geoderma* 135:133–139. <https://doi.org/10.1016/j.geoderma.2005.11.012>
- Vågen T-G, Davey FA, Shepherd KD (2012) Land health surveillance: mapping soil carbon in Kenyan rangelands. In: Nair PK, Garrity D (eds) *Futur. Glob. L. Use Adv. Agrofor.* Springer Netherlands, Dordrecht, pp 455–462
- Vågen T-G, Winowiecki LA, Abegaz A, Hadgu KM (2013a) Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia. *Remote Sens Environ* 134:266–275
- Vågen T-G, Winowiecki LA, Tamene Desta L, Tondoh JE (2013b) The land degradation surveillance framework (LDSF) - field guide v3. World Agroforestry Centre, Nairobi
- Vanlauwe B, Coyne D, Gockowski J et al (2014) Sustainable intensification and the African smallholder farmer. *Curr Opin Environ Sustain* 8:15–22. <https://doi.org/10.1016/j.cosust.2014.06.001>
- Vanlauwe B, Six J, Sangina N, Adesina AA (2015) Soil fertility decline at the base of rural poverty in sub-Saharan Africa. *Nat Plants* 1:15101
- Vasques GM, Grunwald S, Harris WG (2009) Spectroscopic models of soil organic carbon in Florida, USA. *J Environ Qual* 39:923–934. <https://doi.org/10.2134/jeq2009.0314>
- Verchot L, Mackensen J, Kandji S et al (2005) Opportunities for linking adaptation and mitigation in agroforestry systems. In: Robledo C, Kanninen M, Pedroni L (eds) *Trop. For. Adapt. to Clim. Chang. search Synerg.* Center for International Forestry Research (CIFOR), Bogor, pp 103–121
- Viscarra Rossel RA, Behrens T (2010) Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158: 46–54. <https://doi.org/10.1016/j.geoderma.2009.12.025>
- Viscarra Rossel RA, Walvoort DJJ, Mcbratney AB et al (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131:59–75
- Viscarra Rossel RA, Behrens T, Ben-Dor E et al (2016) A global spectral library to characterize the world's soil. *Earth-Sci Rev* 155:198–230. <https://doi.org/10.1016/j.earscirev.2016.01.012>
- Von Fischer JC, Tieszen LL (1995) Carbon isotope characterization of vegetation and soil organic matter in subtropical Forest in Luquillo, Puerto Rico. *Biotropica* 27:138–148. <https://doi.org/10.2307/2388989>
- Wand M (2015) KernSmooth: functions for kernel smoothing supporting Wand & Jones (1995) R package version 2.23–15. <https://CRAN.R-project.org/package=KernSmooth>
- Wang L, Macko SA, Okin GS (2010) Remote sensing of nitrogen and carbon isotope compositions in terrestrial ecosystems. In: West JB, Bowen GJ, Dawson TE, Tu KP (eds) *Isoscapes Underst. Movement, Pattern, Process. Earth through Isot. Mapp*, 1st edn. Springer, London, p 19
- Winowiecki L, Vågen T-G, Huising J (2016a) Effects of land cover on ecosystem services in Tanzania: a spatial assessment of soil organic carbon. *Geoderma* 263:274–283. <https://doi.org/10.1016/j.geoderma.2015.03.010>
- Winowiecki L, Vågen T-G, Massawe B et al (2016b) Landscape-scale variability of soil health indicators: effects of cultivation on soil organic carbon in the Usambara Mountains of Tanzania. *Nutr Cycl Agroecosyst* 105:263–274. <https://doi.org/10.1007/s10705-015-9750-1>
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- Wynn JG, Bird MI (2007) C4-derived soil organic carbon decomposes faster than its C3 counterpart in mixed C3/C4 soils. *Glob Chang Biol* 13:2206–2217. <https://doi.org/10.1111/j.1365-2486.2007.01435.x>