Estimation of soil properties with mid-infrared soil spectroscopy across yam production landscapes in West Africa

Philipp Baumann[1], Juhwan Lee[2], Emmanuel Frossard[3], Laurie Paule Schönholzer[3], Lucien Diby[4], Valérie Kouamé Hgaza[5, 6], Delwende Innocent Kiba[3, 7], Andrew Sila[8], Keith Sheperd[8], and Johan Six[1]

[1]Group of Sustainable Agroecosystems, Institute of Agricultural Sciences, ETH Zurich, 8092 Zürich, Switzerland
[2]Department of Agronomy and Medicinal Plant Resources, Gyeongnam National University of Science and Technology, Jinju, Republic of Korea
[3]Group of Plant Nutrition, Institute of Agricultural Sciences, ETH Zurich, 8315 Lindau, Switzerland
[4]World Agroforestry Centre (ICRAF), Côte d'Ivoire Country Programme, BP 2823 Abidjan, Ivory Coast
[5]Centre Suisse de Recherches Scientifiques en Côte d'ivoire, 01 BP 1303 Abidjan, Ivory Coast
[6]Département d'Agrophysiologie des Plantes, Université Peleforo Gon Coulibaly, BP 1328 Korhogo, Ivory Coast
[7]Institut de l'Environnement et Recherches Agricoles, 01 BP 476 Ouagadougou, Burkina Faso
[8]Land Health Decisions, World Agroforestry Centre (ICRAF), Nairobi, Kenya

Correspondence: Philipp Baumann (baumann-philipp@protonmail.com)

Abstract. Low soil fertility is challenging the sustainable production of staple crops in the yam belt of West Africa. Quantitative soil measures are needed to assess soil fertility decline and to improve crop fertilization management in the region. We developed and tested a mid-infrared (mid-IR) soil spectral library to enable timely and cost-efficient assessments of soil properties. Our collection included 80 soil samples from four landscapes (10 km×10 km) and 20 fields/landscape across a

5   gradient from humid forest to savannah, and 14 additional samples from one landscape that had been sampled within the Land Health Degradation Framework. We derived partial least square regression models to spectrally estimate soil properties. The models produced accurate cross-validated estimates of total carbon, total nitrogen, total sulfur, total iron, total aluminum, total potassium, total calcium, exchangeable calcium, effective cation exchange capacity, diethylenetriaminepentaacetic acid (DTPA) extractable iron and clay content ($R^2 > 0.75$). The estimates of total zinc, pH, exchangeable magnesium, bioavail-

10   able copper and manganese were less predictable ($R^2 > 0.50$). Our results confirm that mid-IR spectroscopy is a reliable and quick method assess the regional-scale variation in most soil properties, especially the ones closely associated with soil organic matter. Although the relatively small mid-IR library shows satisfactory performance, we expect that frequent but small model updates will be needed to adapt the library to the variation of soil quality within individual fields in the regions and their temporal fluctuations.

15 **1 Introduction**

Yam (*Dioscorea* spp.) is an important food and cash crop in West Africa. The yam belt of West Africa spans across the central zone of coastal countries in West Africa, located across the humid forest zone and northern Guinean savanna. It contributes to about 92 % of total world yam production, e.g. a total yield of 73 million metric tons in 2017 (Food and Agriculture

Organization of the United Nations, 2019). The cropping area in the West African yam belt has been expanded with accelerated
20 population growth, which has in many places caused soil degradation. Furthermore, there is a trend of shortened fallow periods
in the cropping areas of West Africa over the last decades, which has further exacerbated the decline in soil fertility across
the yam belt. Traditionally, yam is grown without external input in these areas. Therefore, the production of yam and other
crops grown in the region depends on soil organic matter (SOM) status (Padwick, 1983), which serves as a main pool of
plant-available nutrients and provides cation exchange surfaces for soil nutrients (Syers et al., 1970; Soares and Alleoni, 2008).
25 Particularly, a strong positive relationship between high organic matter stocks and yam productivity is reported after fallow
and when no fertilizer is added (Diby et al., 2009; Kassi et al., 2017). Thus, maintaining or increasing SOM and available
nutrient levels is of utmost importance for sustainable production of yam and other crops in West Africa (Carsky et al., 2010).
Furthermore, linking soil properties and yam yields (Frossard et al., 2017) and accounting for soil macro- and micronutrient
status (O'Sullivan and Jenner, 2006) is fundamental to improving crop yields and soil management strategies.

30 Soil fertility is an integrative measure of soil attributes and their interactions that support the long-term agricultural pro-
duction potential. Soil fertility is commonly decomposed into three main components, the physical, chemical and biological
(Abbott and Murphy, 2007). Here, it is important to interpret soil fertility in the form of soil conditions and functions at an
adequate resolution over time and space, and in relation to the crop of interest. For yam, low tuber yields are often attributed to
an unbalanced ratio of essential nutrients (i.e. N, P, K) available in the soil (Enyi, 1972) and a fast mineralization and hence de-
35 pletion of organic matter (Carsky et al., 2010; Hgaza et al., 2011). Yet, the relationship between soil properties and tuber yield
is not fully understood (Frossard et al., 2017). The reason is that the response of yam to mineral fertilization is highly variable
because of confounding environmental and management variables, such as climate, soil type, micronutrient deficiencies, seed
tuber quality and planting density or disease pressure across the yam belt (Kang and Wilson, 1981; O'Sullivan and Jenner,
2006; Cornet et al., 2016). Further, there are no soil fertility recommendations specific for yam under West African conditions.
40 For this reason, establishing yam field trials designed with different organic and mineral fertilization strategies within different
yam growing regions is required to optimize yam fertilization targeting regional soil and environmental conditions (Frossard
et al., 2017). Despite the importance of soil fertility, it is challenging to quantify soil measures at sufficient temporal and spatial
resolution to relate them to yam productivity together with other management effects.

In order to quickly assess key soil properties, such as soil organic carbon (SOC) and cation exchange capacity (CEC),
45 we need more cost- and time-efficient methods in addition to the traditional wet chemistry laboratory analyses that are often
cost-intensive and time consuming. Proximal sensing is a method that can provide reliable soil measurements rapidly and inex-
pensively (UNEP, 2012). Soil visible and near infrared (vis-NIR), and mid-infrared (mid-IR) diffuse reflectance spectroscopy
has gained popularity over the past 30 years to assess soil properties to complement conventional laboratory analytical methods
(Nocita et al., 2015). Previous studies have shown successful spectroscopic predictions of soil properties, such as organic C,
50 texture, cation exchange capacity (CEC), and exchangeable K (Viscarra Rossel et al., 2006; Cécillon et al., 2009; Nocita et al.,
2015; Sila et al., 2016). Many soil chemical and physical properties, such as soil mineralogy, the concentration, forms and
distribution of SOM, are closely associated with IR spectral diversity. Nevertheless, soil IR spectroscopy often needs laboratory
reference analysis data for model development and calibration. Further, a library that includes a broad range of soil biophysical

conditions found in the region in which it is used needs to be established. Depending on the study scale — field (e.g., Cambou
et al., 2016), region, country (e.g., Clairotte et al., 2016), continent (e.g., Sila et al., 2016)), world (e.g., Viscarra Rossel et al.,
2016) — various statistical predictive modeling strategies are typically employed to account for regional variability in soil
properties and determine empirical relationships between spectra and soil attributes. However, particular regions in spectra
are characteristic for functional groups of soil components and thus, elucidating spectral features that are important for the
prediction of a particular soil attribute helps to understand and validate the mechanisms based on which the empirical-derived
models predict the soil properties.

Thus the main objectives of this study are to (1) develop and evaluate mid-IR spectroscopic models to estimate soil properties
for selected landscapes representing major soil and climatic conditions in the West African yam belt, (2) to determine important
spectral features for specific soil properties, and (3) to build a new soil spectral library in four landscapes of the West African
yam belt for soil prediction and assessment.

2 Materials and methods

2.1 Landscapes and soil sampling

Our study area covered the climatic and soil biophysical conditions representative of the West African yam belt. We selected
four landscapes, two in Ivory Coast and two in Burkina Faso. Each landscape (10 km x 10 km) represents a diverse geo-
graphic ecoregion. The landscapes cover a gradient between humid forest and the northern Guinean savannah. Specifically,
the landscape Liliyo in Ivory Coast is at $5.88°N$ and in the humid forest zone. The predominant soil type is Ferralsol (FAO,
2014). The landscape Tieningboué in Ivory Coast is at $8.14°N$ and belongs to the forest savannah transitional zone. The soils
are dominated by Nitisols and Lixisols (FAO, 2014). The landscape Midebdo is at $9.97°N$ and in the sub-humid savannah of
Burkina Faso. Its dominant soil types include Lixisols, Gleysols, and Leptosols (FAO, 2014). The landscape Léo is at $11.07°N$
and in the northern Guinean savannah of Burkina Faso and has Lixisols and Vertisols as the dominant soil type (FAO, 2014).
The mean annual rainfall were approximately 1300mm in Liliyo, and 900mm in Tiéingboué, Midebdo, and Léo.

During July and August 2016, we sampled the soil from a total of 80 fields under yam cultivation across the four landscapes,
i.e. 20 yam fields in each landscape. The fields were selected in advance by taking into account visual variation in soil color
and texture across the landscape. The yam fields selected contained the maximum soil variability based on soil colour and
cropping history, taking into account both local farmers' knowledge on soil fertility and agronomic extension expertise. Yam is
typically planted on soil mounds, ranging from 5000 to 10000 mounds per hectare with a single yam plant per mound. Within
each field, we sampled the soil at four adjacent mounds in square arrangement, which were spaced between 0.5 and 2 m. At
each mound, 6 to 8 auger cores (2.5 cm in diameter) to the 30 cm depth were taken at a radius between 15 and 30 cm away
from the center of a mound, depending on the size of the mounds. Then the soils from the four mounds were combined into
one composite sample per field (around 500 to 1000 g of soil).

An additional set of 14 composite soil samples was collected by the International Center for Research in Agroforestry
(ICRAF) at Liliyo from one sentinel site called "Petit-Bouaké" (UNEP, 2012). Sampling took place between 25 and 29 August,

2015 at positions that were previously selected for the Land Degradation Surveillance Framework (LDSF) in a spatially stratified manner (Vagen et al., 2010). The soil samples received from ICRAF were within the same landscape as the sampled soils in Liliyo within YAMSYS, but sampled from different positions. All soil samples were air-dried and stored in plastic bags until

90   further analysis.

2.2   Soil reference analyses

The air-dried soil samples were crushed and sieved at 2 mm. About 60 to 70 g of the sieved soil was oven-dried at $60°C$ for 24 hours, of which 20 g were ball-milled. All chemical analyses except soil pH were conducted both on the soils sampled in yam fields ($n = 80$) and the LDSF soils obtained from ICRAF ($n = 14$).

95   The milled soils were analyzed for total C and macronutrient (N and S) concentrations using an elemental analyzer (vario PYRO cube, Elementar Analysensysteme GmbH, Germany). For each of the four landscapes, two soils were selected and analyzed based on three analytical replicates for quantifying within-sample variance of the elemental analysis. For the remaining samples, the analysis was not repeated. Sulfanilamide was used as a calibration standard for the dry combustion. For pH determination 10 g of air-dried soil per sample was placed in a 50 mL Falcon tube and 20 mL of de-ionized water was added. The

100   samples were shaken in a horizontal shaker for 1.5 hours and measured for pH using a pH electrode (Benchtop pH/ISE meter model 720A, Orion Research Inc., USA).

Resin-extractable P was used as an indicator of plant-available P, as it correlates with P uptake by plants (Nuernberg et al., 1998). Inorganic P was extracted using an anion exchange resin membrane (Kouno et al., 1995). The extraction method was slightly modified by using only one instead of two resin strips of 6×2 cm (No. 55164 2S, BDH Laboratory Supplies, Poole,

105   England) saturated with $CO3^{-2}$, and 2 g instead of 4 g of dried soil was weighted. No fumigation step to determine microbial P was performed, as the soils had been dried and had storage periods longer than one month between sampling and analysis. In the resin eluates (a mixture of 0.1 M NaCl and 0.1 M HCl), the concentrations of inorganic P were measured colorimetrically using the malachite green method (Ohno and Zibilske, 1991). Bioavailable micronutrient (Fe, Mn, Zn, and Cu) concentrations in soils were determined with the diethylenetriaminepentaacetic acid (DTPA) extraction method, as described in Lindsay and

110   Norvell (1978). The extracting solution consisted of 0.0005 M DTPA, 0.01 M $CaCl_2$, and 0.1 M triethanolamine. Briefly, 10 g of the sieved <2 mm) soils were extracted with 20 mL of DTPA solution. Micronutrient concentrations in the filtrates were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES, a Shimandzu Plasma Atomic Emission Spectrometer ICPE-9820). Final DTPA extractable concentrations of Fe, Mn, Zn, and Cu were calculated back to per kg dry soil. For each landscape, two soils were selected and analyzed in triplicates to assess analytical errors. For the remaining soils

115   the analysis was not repeated.

For each sample, the concentrations of total element (Fe, Si, Al, K, Ca, P, Zn, Cu, and Mn) in the soil was assessed by energy dispersive X-ray fluorescence spectrometry (ED-XRF) measurements on 4 g of the milled soil with a SPECTRO XEPHOS instrument (SPECTRO Analytical Instruments GmbH, Germany). The soil was mixed with equal amount of wax using a ball mill and pressed into pellets. Exchangeable cations (Ca^{2+}, Mg^{2+}, K^+, Na^+, and Al^{3+}) were determined with the $BaCl_2$

120   method (Hendershot and Duquette, 1986). About 2 g of the air-dried soil (<2 mm) were extracted by shaking for 2 hours

with 30 mL of 0.1 M $BaCl_2$ on a horizontal shaker (120 cycles min^{-1}). The suspension was filtered through no. 40 filter paper (Whatman, Brentford, UK). For each landscape, two soils were analyzed in analytical triplicates. The concentrations of exchangeable cations in the $BaCl_2$ extract were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES, Shimandzu Plasma Atomic Emission Spectrometer ICPE-9820). Different $BaCl_2$ extract dilutions were used in order to

125   obtain an optimal signal intensity for the quantification of specific elements across all samples. Concentration of H^+ per kg dry soil was calculated based on the pH measured in the $BaCl_2$ extractant. The $BaCl_2$ extraction does only slightly modify pH and is therefore an appropriate method to calculate effective CEC (CEC_{eff}) at native soil pH. Using the concentrations of the $BaCl_2$-extractable cations (i.e. Ca^{2+}, Mg^{2+}, K^+, Na^+, Al^{3+} and H^+), CEC_{eff} was calculated as sum of exchangeable cations in cmol of cation charge per kg dry soil. Exchangeable acidity was defined by the sum of exchangeable Al^{3+} and H^+. Base

130   saturation in % was calculated as ratio of the sum of basic cations (Ca^{2+}, Mg^{2+}, K^+, Na^+) in cmol(+) per kg soil to the CEC_{eff} multiplied by 100.

Particle size analysis was conducted by IITA in Cameroon as described in Bouyoucos (1951). Briefly, 50 g of dried 2 mm sieved soil was stirred with 50 mL 4 % sodium hexametaphosphate and 100 mL of deionized water in a mixer, for breaking down the aggregates into into individual particles. Readings with a hydrometer (ASTM 152 H, Thermco, New Jersey, USA)

135   were taken after letting it stand in the suspension for 30 minutes. The silt content was calculated by subtracting the measured proportion of sand and clay from 100 %.

Spectroscopic measurements

The milled soils ($n = 94$) were measured on a Bruker ALPHA DRIFT spectrometer (Bruker Optics GmbH, Ettingen, Germany), which was equipped with a ZnSe optics device, a KBr beamsplitter, and a DTGS (deuterated tri-glycine sulfate) detector. Mid-

140   IR spectra were recorded between 4000 cm^{-1} and 500 cm^{-1} with a spectral resolution of 4 cm^{-1} and a sampling resolution of 2 cm^{-1}. Reflectance (R) spectra were transformed to apparent absorbance (A) using $A = \log_{10}(1/R)$ and corrected for atmospheric CO_2 using macros within the OPUS spectrometer software (Bruker Corporation, US). The spectra were referenced to a IR-grade fine ground potassium bromide (KBr) powder spectrum, which was measured prior to the first soil sample and measured every hour again. All spectra were recorded by averaging 128 measurements for each of the three sample repetitions

145   per soil.

2.3   Spectroscopic modeling

2.3.1   Processing of soil spectra

Three replicates of spectra were averaged for each sample. The spectra were transformed by using a Savitzky-Golay smoothed first derivative using a third-order polynomial and a window size of 21 points (42 cm^{-1} at spectrum interval of 2 cm^{-1})

150   (Savitzky and Golay, 1964). Prior to spectral modeling, Savitzky-Golay preprocessed spectra were further mean centered and scaled (divided by standard deviation) at each wavenumber.

2.3.2 Model development and validation

The measured soil properties were modeled by applying partial least squares regression (PLSR) (Wold et al., 1983) with the preprocessed spectra as predictors. The models were fitted using the orthogonal scores PLSR algorithm. 5-times repeated 10-

155 fold cross-validation was performed to provide unbiased and precise assessment of PLSR model performance (Molinaro et al., 2005; Kim, 2009). For each individual soil property, the number of factors for the most accurate PLSR model was tuned separately. For each soil property model, the sample set was repeatedly randomly split into $k = 10$ (approximately) equally-sized subsets without replacement for all repeats $r = 1, 2, .., 5$ and all candidate values in the tuning grid with the number of PLSR factors (ncomp) $= 1, 2, ..., 10$. Within each of the $r \times \text{ncomp} = 5 \times 10 = 50$ resampling data set splits, each of the

160 10 possible held-out and model fitting set combinations (folds) was subjected to candidate model building at the respective ncomp, using $k - 1 = 9$ out of 10 subsets and remaining held-out samples were predicted based on the fitted models. The root mean square error (RMSE, eq. (1)) of the held-out samples was calculated by aggregating all repeated K-fold cross-validation predictions ($\hat{y}_i$) and corresponding observed values (y_i) grouped by ncomp, which resulted in a cross-validated performance profile RMSE vs. ncomp.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{1}$$

165

Based on this performance profile, the minimal ncomp among the models whose performance was within a single standard error ("One standard error" rule, (Breiman et al., 1984)) of the lowest numerical value of RMSE was selected.

Model assessment was done with the best factors for each property using cross-validation hold outs. We reported the cross-validated measures RMSE, R^2 (coefficient of determination) obtained via linear least-squares regression, and ratio of perfor-

170 mance to deviation (RPD), after averaging predictions across repeats. The RPD index is the ratio of the chemical reference data standard deviation to the RMSE of prediction.

$$\text{RPD} = \frac{s_y}{\text{RMSE}} \tag{2}$$

Besides calculating the above listed performance measures, the uncertainty of spectral estimates was graphically reported for each soil sample, using prediction means and 95% confidence intervals derived from cross-validation repeats ($n = r = 5$;

175 Eq. 3 and 4).

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i - \overline{\hat{y}_i}\right)^2 \tag{3}$$

$$\overline{\hat{y}_i} \pm t(n-1, 1-\alpha/2) \frac{S_n}{\sqrt{n}}; \alpha = 0.05 \tag{4}$$

In order to cover the full training data space in the models for future sample predictions, the final PLSR models were rebuilt using the entire training set and the respective values of optimal final number of PLSR components determined by the procedure

180 described above.

2.3.3 Model interpretation

The mid-IR spectra contain complex information about soil composition and properties. To establish a predictive relationship, statistical models need to find relevant spectral features for each soil property. Model interpretation requires a variable importance assessment to decide on the contribution of spectral variables to prediction and to explain spectral mechanisms.
185 Therefore, we conducted model interpretation based on the variable importance in projection (VIP) method (Wold et al., 1993; Chong and Jun, 2005), using the model at respective best number of factors (ncomp). The VIP measure v_j was calculated for each wavenumber variable j as

$$v_j = \sqrt{p \sum_{a=1}^{A} \left[SS_a \left(w_{aj} / \|w_{aj}\| \right)^2 \right] / \sum_{a=1}^{A} (SS_a)} \tag{5}$$

where w_{aj} are the PLSR weights for the ath component for each of the wavenumber variables and SS_a is the sum of squares
190 explained by the ath component:

$$SS_a = q_a^2 \, t_a^{\mathsf{T}} \, t_a \tag{6}$$

where q_a are the scores of the predicted variable y and t_a are the scores of the predictors X. These VIP scores account for multicollinearity found in spectra and are considered as robust measure to identify relevant predictors. Important wavenumbers were classified with a VIP score above 1. A variable with VIP above 1 contributes more than average to the model prediction.
195 For model interpretation, we only computed VIP at the respective finally chosen number of PLS components a_{final} for each considered model. We focused on a selection of three well performing models with $R^2 \geq 0.8$ (RPD ≥ 2.3) to illustrate model interpretation. These were total C, total N and clay content.

2.4 Statistical software

The entire analysis was performed using the R statistical computing language and environment (version 3.6.0) (R Core Team,
200 2017). We used the pls (Mevik et al., 2019) package for PLSR, as described by Martens and Naes (1989). Cross-validation resampling, model tuning, and assessment was done using the caret package (Kuhn et al., 2019). Custom functions from the simplerspec package were used for spectroscopic modeling (Baumann, 2019). All data and code to reproduce the results of this study is available online via Zenodo (Baumann, 2018).

3 Results

205 **Measured properties and mid-IR estimates of yam soils**

The distribution of soil properties at the yam fields showed a wide variation across the landscapes (Figure 1). Total C concentrations across all fields ranged from $2.4\,\mathrm{g\,C\,kg^{-1}}$ soil to $24.7\,\mathrm{g\,C\,kg^{-1}}$ soil. Total C values at the landscape scale were the

lowest (median) in Léo and the highest in Tiéningboué. Soils from yam fields in the two landscapes from Ivory Coast (13.0 ± 5.4 g C kg^{-1} soil; mean $\pm$ standard deviation) had relatively higher total C compared to the fields in the landscapes in Burkina Faso (6.1 ± 3.6 g C kg^{-1} soil). The median value and variation of CEC$_{\text{eff}}$ exhibited similar patterns across the landscapes to total C. Total N concentrations across all fields ranged from 0.18g N kg^{-1} soil to 2.48g N kg^{-1} soil. Total N within and across the four landscapes exhibited a similar pattern as total C. Generally, the landscapes in Burkina Faso were low in total N compared to those from Ivory Coast (0.44 ± 0.24 g N kg^{-1} soil vs. 1.09 ± 0.46 g N kg^{-1} soil). Median total N concentrations were almost identical for Liliyo and Tiéningboué (1.1g N kg^{-1} soil). Total S concentrations varied between 41mg S kg^{-1} soil to 242mg S kg^{-1} soil across all fields, and showed a similar pattern as total C and N. The yam fields in the landscapes of Bukina Faso had on average more than two times higher total S than the other landscapes. Total P concentrations were in a similar range for the landscapes Léo, Midebdo, and Liliyo. In Tiéningboué, total P values were on average almost two times higher than the other fields (817mg S kg^{-1} soil vs. 453mg S kg^{-1} soil), with more within-landscape variation.

Total Fe, total Al, total Ca, total Zn, and total Cu concentrations in the soil tended to be higher for the landscapes in Ivory Coast than in Burkina Faso (Figure 1). To give an example, median concentrations of total Ca were 2.16g Ca kg^{-1} soil in fields sampled from the Tiéningboué region, and similar in Liliyo ($= 1.90$g Ca kg^{-1} soil), while they were markedly lower in Léo and Midebdo ($= 0.90$ vs. 1.26g Ca kg^{-1} soil). In general, the ranges for total micronutrient contents were more variable in the landscapes of Ivory Coast (e.g., range $= 14 - 57$mg Zn kg^{-1} soil in Liliyo; lowest range in Léo $= 12.2 - 19.7$mg Zn kg^{-1} soil). Total K concentration was highly variable within and across the landscapes (overall range $= 0.5 - 34.1$g K kg^{-1} soil), and lowest in Midebdo (range $= 0.9 - 8.9$g K kg^{-1} soil), while the highest total K median was measured for yam fields in Léo (range $= 4.1 - 25.0$g K kg^{-1} soil).

Soil resin-extractable P concentrations varied between 0.8mg P kg^{-1} soil to 33.1mg P kg^{-1} soil. In Tiéningboué, resin-extractable P was on average higher than in soils of the other landscapes (Figure 1). Median extractable Fe and its interquartile ranges were comparable across the landscapes (see Figure 1). However, there were some fields where extractable Fe reached values higher than 100mg Fe kg^{-1} soil. Median extractable Zn values showed a similar pattern as total C, with the highest median values and interquartile range in Tiéningboué and had the lowest in Léo. In comparison, the highest median values and interquartile range of extractable Cu and Mn were found in Liliyo. For extractable Zn, Cu, and Mn median values and interquartile range were higher in the two landscapes in Ivory Coast than the two landscapes in Burkina Faso.

Across all samples and landscapes, soil pH varied between 4.7 and 8.4. Median pH was comparable in Tiéningboué ($= 6.4$), Liliyo ($= 6.5$), and Midebdo ($= 6.5$). Median pH of yam fields in Léo ($= 6$) was lower than in the other landscapes. Exchangeable K, Ca, and Mg concentrations showed similar patterns across the four landscapes. In Burkina Faso, each of the exchangeable cations showed relatively low median concentrations across the fields and less landscape-level variation than in Ivory Coast. In general, the highest median and variation of exchangeable cations among the landscapes were measured for the yam field soils in Tiéningboué. Median exchangeable Al values were comparable among the landscapes, although there were some outliers with exchangeable Al > 20mg kg^{-1} soil for Midebdo, Liliyo, and Tiéningboué. The CEC$_{\text{eff}}$ ranged from 0.9cmol(+) kg^{-1} soil to 14.6cmol(+) kg^{-1} soil across all fields and landscapes. Median CEC$_{\text{eff}}$ tended to decreases in the
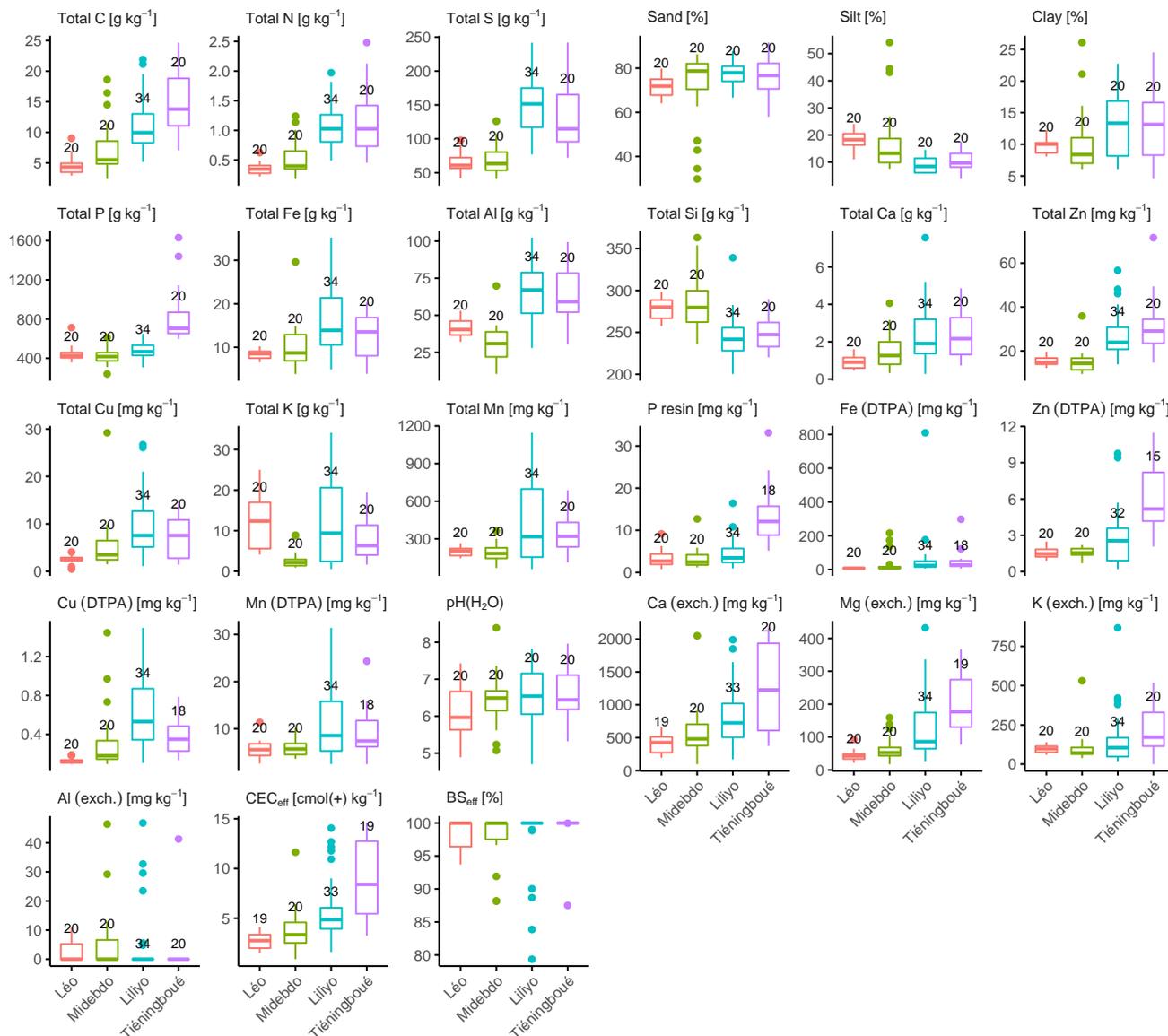
Figure 1. Reference measurements of soil chemical properties. Léo and Midebdo are two yam growing regions in Burkina Faso, and Lilyo and Tiéningboué are in Côte d'Ivoire. The chemically analyzed soils ($n = 94$) originated from 20 yam fields per landscape, and 14 additional soils from the Lilyo region were provided by the World Agroforestry Center (ICRAF). C = carbon, N = nitrogen, P = phosphorus, Fe = iron, Al = aluminum, Si = silicum, Ca = calcium, Zn = zinc, Cu = copper, K = potassium, Mn = manganese, P resin = resin-extractable P (proxy for plant-available P). Bioavailable micronutrients were measured by diethylenetriaminepentaacetic acid (DTPA extraction). Ca(exch.), Mg(exch.), K(exch.), and Al(exch.) signify exchangeable elements determined with $BaCl_2$ extraction. CEC_{eff} = effective cation exchange capacity, BS_{eff} = effective base saturation. The number of soils analyzed for each individual property is indicated above the the 75 % percentile.

following order across landscapes: Léo > Midebdo > Liliyo > Tiéningboué. The interquartile range of CEC_{eff} was also the greatest in Tiéningboué and the smallest in Léo.

Reference measurements for total N, S, exchangeable Ca, exchangeable Mg and $\text{CEC}_{\text{eff.}}$ were highly correlated to total C
245 (Figure 2; $0.71 \leq r \leq 0.92$ ($\text{CEC}_{\text{eff.}}$)). Also, total Ca, Al, and clay content correlated strongly to total C ($r > 0.70$). Clay contents were weakly related to silt ($r = 0.21$), while sand had a markedly negative relationship to silt ($r = -0.89$). Bioavailable Cu and Zn . Resin-extractable P was moderately correlated to total C ($r = 0.53$) and pH ($r = 0.38$). Bioavailable Zn (DTPA) was co-varying with both $\text{CEC}_{\text{eff.}}$ ($r = 0.58$) and total Zn ($r = 0.59$). Bioavailable Cu (DTPA) had a strongly positive association to total Cu ($r = 0.90$). Exchangeable K (BaCl_2) had the strongest relationship to total C and $\text{CEC}_{\text{eff.}}$ ($r = 0.63$, and $r = 0.64$).

250 ## 3.1 Soil mid-IR spectroscopic models

Among the measured soil properties, mid-IR PLSR models for total K ($R^2 = 0.96$) and total Al ($R^2 = 0.97$) were best performing (Table 1). Out of a total of 27 soil attributes, 11 were well quantified by the models when considering categorization judged upon on an $R^2_{\text{cv}} \geqslant 0.75$ criterion (Figure 3). The confidence intervals derived from cross-validation prediction were very narrow, showing all PLSR models were stable. Within this group of stable models, four soil attributes are directly related
255 to the mineralogy (total Fe, Al, K and Ca), three are related to soil organic matter (total C, N and S), one to texture (clay fraction), one to plant nutrition (exchangeable Fe), and two related to mineralogy and plant nutrition (exchangeable Ca and CEC_{eff}). More specifically, total C was accurately predicted, with an R^2 of 0.92 and a RMSE of 1.6g C kg^{-1} soil. The models were also able to predict total N well ($R^2 = 0.89$; RMSE $= 0.16 \text{g N kg}^{-1}$ soil). Prediction accuracy of total S was slightly lower than for total C, but its goodness-of-fit and RMSE suggest that the model was reliable for prediction. However, exchangeable
260 K ($R^2 = 0.28$), resin-extractable P ($R^2 = 0.40$) and BS_{eff} ($R^2 = 0.24$) were poorly predicted (Table 1). Predictions for percent clay were reliable ($R^2 = 0.81$; RMSE $= 2.1\%$), whereas predictions for percent sand ($R^2 = 0.45$; RMSE $= 8.1\%$) and percent silt ($R^2 = 0.41$; RMSE $= 6.5\%$) were not accurate. Finally chosen models of all soil attributes had between 1 and 9 PLSR components.
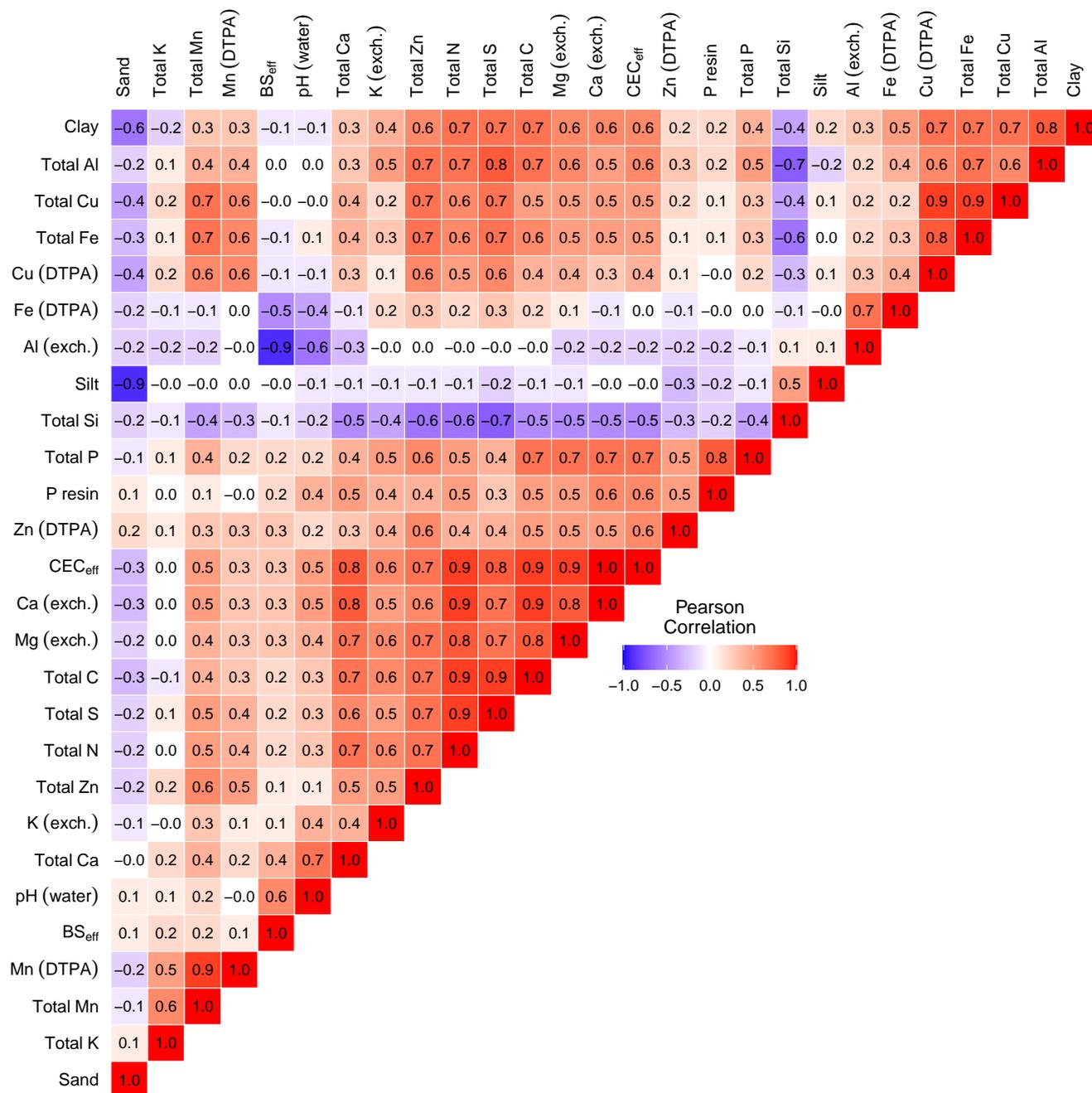
Figure 2. Correlation matrix of soil properties measured on each 20 soils sampled from individual yam fields per landscape, and 14 additional agricultural soils received from the World Agroforestry Center ($n = 94$; see Figure 1 for further details and abbreviated chemical properties). Pearson correlation coefficients (r) were rounded to 1 digit.

Table 1. Descriptive summary of measured (meas.) soil reference data (shown in Figure 1) and evaluation results of cross-validated PLSR models. All samples across the four landscapes were aggregated into a single model per respective soil property. Model evaluation was done on held-out predictions of 5 times repeated 10-fold cross-validation (abbreviated by cv, subscript) at the finally selected number of PLSR components (ncomp). CV = coefficient of variation, RMSE = root mean square error, RPD = ratio of performance to deviation. C = carbon, N = nitrogen, P = phosphorus, Fe = iron, Al = aluminum, Si = silicum, Ca = calcium, Zn = zinc, Cu = copper, K = potassium, Mn = manganese, P resin = resin-extractable P (proxy for plant-available P). Bioavailable micronutrients were measured by diethylenetriaminepentaacetic acid (DTPA) extraction. Ca(exch.), Mg(exch.), K(exch.), and Al(exch.) signify exchangeable elements determined with $BaCl_2$ extraction. CEC_{eff} = effective cation exchange capacity, BS_{eff} = effective base saturation.

Soil attribute	n	Min$_{meas.}$	Max$_{meas.}$	Med$_{meas.}$	Mean$_{meas.}$	CV$_{meas.}$	ncomp	RMSE$_{cv}$	R^2_{cv}	RPD$_{cv}$
Total C [g kg^{-1}]	94	2.4	24.7	8.5	9.9	58	6	1.6	0.92	3.6
Total N [g kg^{-1}]	94	0.18	2.48	0.72	0.81	61	6	0.16	0.89	3.0
Total S [mg kg^{-1}]	94	41	242	99	111	46	2	20	0.85	2.6
Sand [%]	80	29.8	91.6	75.6	74.2	14	2	8.1	0.42	1.3
Silt [%]	80	3.9	54.1	12.0	14.1	60	2	6.5	0.41	1.3
Clay [%]	80	4.5	26.1	10.1	11.6	42	2	2.1	0.81	2.3
Total P [mg kg^{-1}]	94	240	1631	467	530	40	3	131	0.61	1.6
Total Fe [g kg^{-1}]	94	4	35	10	12	54	5	3	0.81	2.3
Total Al [g kg^{-1}]	94	10	102	48	53	42	5	4	0.97	6.0
Total Si [g kg^{-1}]	94	200	363	262	262	12	3	20	0.59	1.6
Total Ca [g kg^{-1}]	94	0.3	7.6	1.4	1.9	70	5	0.6	0.78	2.2
Total Zn [mg kg^{-1}]	94	9.5	71.6	19.1	22.6	49	1	6.7	0.63	1.7
Total Cu [mg kg^{-1}]	94	0.5	29.2	4.7	6.8	87	7	3.2	0.71	1.9
Total K [g kg^{-1}]	94	0.5	34.1	5.8	9.5	91	7	1.7	0.96	5.1
Total Mn [mg kg^{-1}]	94	59.2	1146.0	221.5	308.0	74	5	116.4	0.74	2.0
log(P resin) [mg kg^{-1}]	92	-0.2	3.5	1.4	1.4	57	2	0.6	0.40	1.3
log(Fe(DTPA)) [mg kg^{-1}]	92	1.0	6.7	2.7	2.9	38	9	0.5	0.77	2.0
Zn (DTPA) [mg kg^{-1}]	87	0.2	11.5	1.9	2.8	89	3	2.1	0.25	1.1
Cu (DTPA) [mg kg^{-1}]	92	0.1	1.5	0.2	0.4	89	6	0.2	0.74	2.0
Mn (DTPA) [mg kg^{-1}]	92	2.5	31.4	6.5	8.6	69	3	4.0	0.55	1.5
pH$_{H_2O}$	80	4.7	8.4	6.4	6.4	11	8	0.5	0.61	1.6
Ca (exch.) [mg kg^{-1}]	92	98	2170	604	774	70	5	237	0.81	2.3
Mg (exch.) [mg kg^{-1}]	93	18	432	76	113	84	3	58	0.62	1.6
K (exch.) [mg kg^{-1}]	94	0	868	104	145	95	1	120	0.28	1.2
Al (exch.) [mg kg^{-1}]	94	0	47	0	4	258	2	9	0.21	1.1
CEC$_{eff}$ [cmol(+) kg^{-1}]	91	0.9	14.6	4.2	5.3	67	6	1.4	0.84	2.5
BS$_{eff}$ [%]	91	79	100	100	98	4	2	3	0.24	1.1
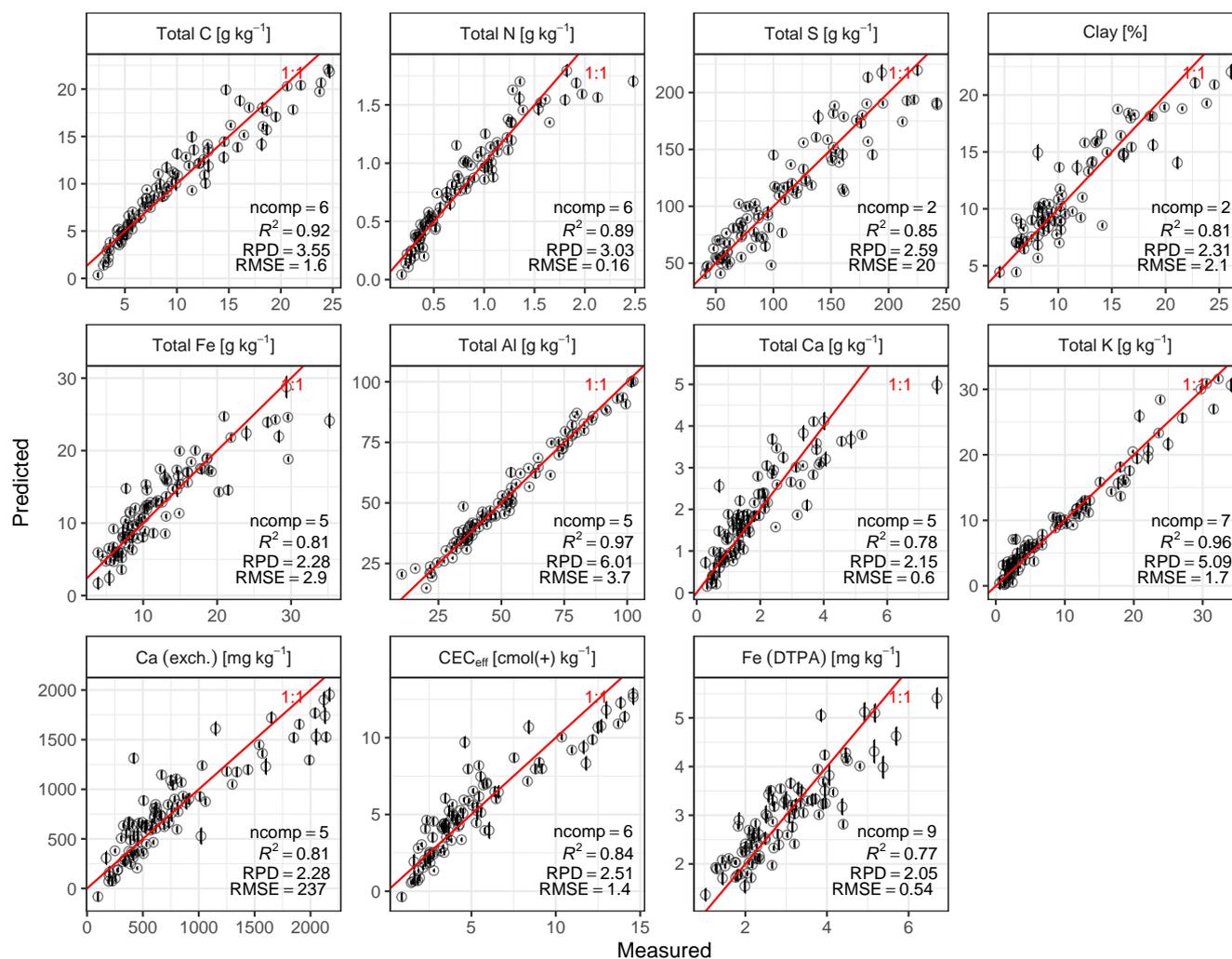
Figure 3. Cross-validated predictions of soil properties (y-axis) derived from best mid-IR partial least squares regression (PLSR) models vs. laboratory reference measurements (x-axis; see Figure 1). Average estimates, their confidence intervals (error bars), and evaluation metrics were derived with $5 \times$ repeated 10-fold cross-validation. ncomp = number of PLSR components of most accurate final models, RSME = root mean square error, RPD = ratio of performance to deviation. Only soil properties modeled with $R^2 > 0.75$ are shown. CEC_{eff} = effective cation exchange capacity. Exchangeable (exch.) elements were determined with $BaCl_2$. Bioavailable Fe was determined diethylenetriamine-pentaacetic acid (DTPA) extraction.

3.1.1 Model interpretation

265 A large proportion of absorptions had $\textsc{vip} > 1$ for each the total C, total N and clay models (Figure 4). Important wavenumbers ($\textsc{vip} > 1$) for total C were mostly between 3140cm^{-1} and 1230cm^{-1}. Besides clear absorption peaks, there were relatively continuous spectral features that were important to the models. For example, the relatively continuous and smooth spectral region between the alkyl $\mathrm{C-H}$ vibrations at 2855cm^{-1} and 2362cm^{-1} had comparable contribution to the model as peak regions associated with total C prediction. The VIP patterns across wavenumbers were almost identical for total C and N

270 models, and its reference measurements were strongly correlated ($r = 0.94$; Figure 2). In contrast, the clay content model deviated from the total C model in particular regions, for example around the kaolinite OH− feature at 3620cm^{-1} or at kaolinite $\mathrm{Al-O-H}$ vibrations at 934cm^{-1} and 914cm^{-1}.
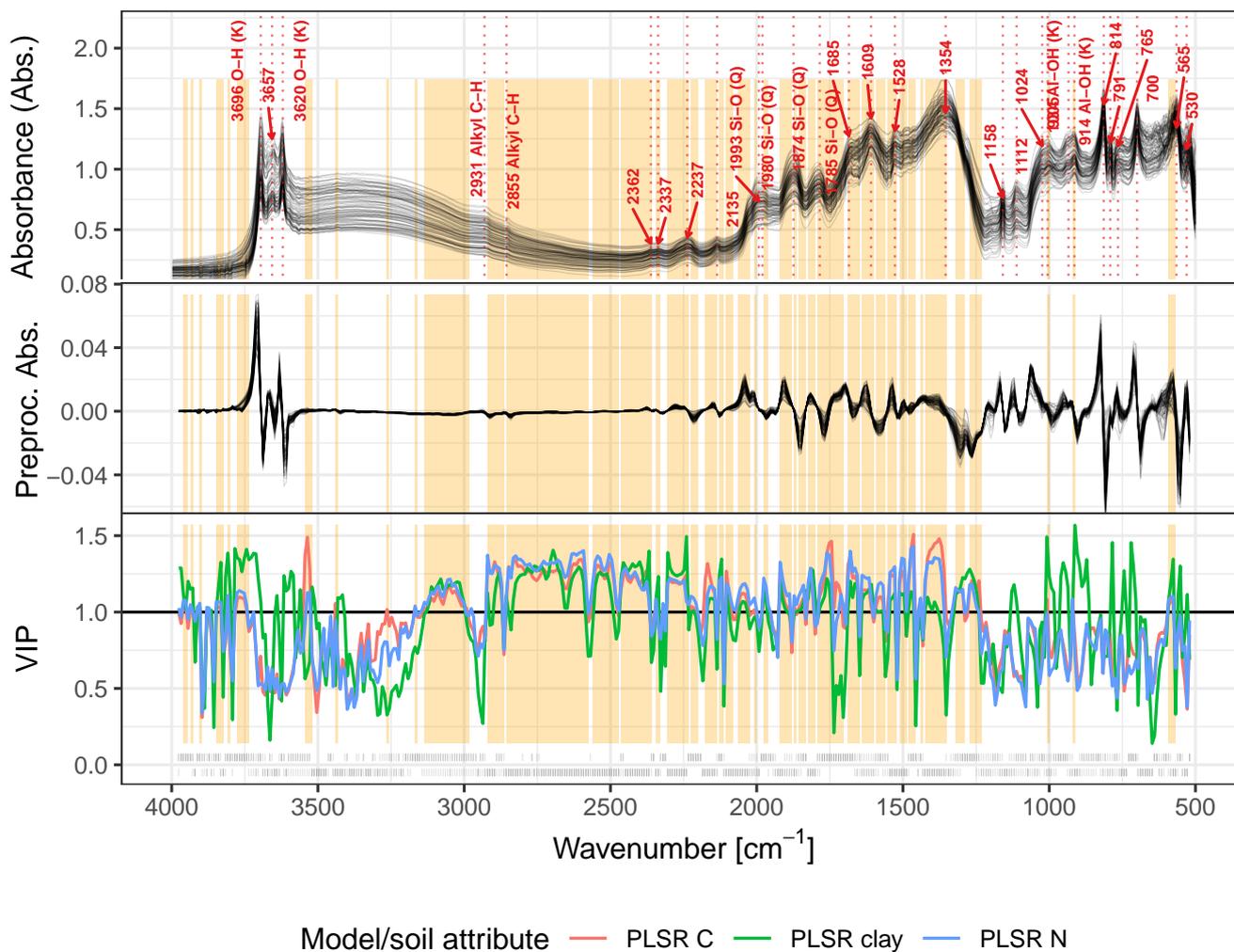
Figure 4. Variable importance analysis of partial least squares regression (PLSR) models for total soil C, total N and % clay , including over-laid raw and preprocessed spectra. Top panel shows resampled mean sample absorbance spectra ($n = 94$). Prominent peaks were identified as local maxima with a span of 10 points 20 cm^{-1}) for the selected wavenumbers. Fundamental mid-IR vibrations that are well described in the literature (e.g., Madejová et al., 2002; Rossel and Behrens, 2010; Stevens et al., 2013) were added as labels when identified peaks matched literature assignments. (Q) stands for quartz and (K) for kaolinite. The middle panel depicts preprocessed spectra (Savitzky-Golay first derivative with a window size of 21 points (42 cm^{-1}); 3rd order polynomial fit). The bottom panel shows variable importance in the projection (VIP) for three selected well performing PLSR models (total C, total N and % clay; $R^2 > 0.81$). The black horizontal line at VIP $= 1$ indicates the threshold above where absorbance at the wavenumbers explain more than average to the prediction of a certain soil property. Dashed points closely below the $y = 0$ line of the VIP graph visualize positive (above $y = 0$) and negative (below $y = 0$) PLSR β coefficients.

Discussion

3.2 Accuracy of mid-IR spectroscopy for agronomic diagnostics

275 Timely and accurate estimates of multiple soil properties are required to better understand and predict soil constraints across the yam belt in West Africa. The soil spectral library from our study, which includes four landscapes of the yam belt, can be practical to diagnose and monitor (and eventually manage) soil fertility that is considered to be low and therefore being a major constraint for yam production in West Africa. Specifically, our results show that properties closely related to organic matter — total amount of C, (micro)-nutrients, and exchangeable cations — can be accurately estimated using mid-IR spectra

280 and in the selected yam growing landscapes (Figure 3). Soil organic matter plays a crucial role during vegetative growth and tuber formation phases of yam, as it guarantees among many other functions the storage and availability of essential nutrients and water needed for yam and tuber growth throughout the season, and as well prevents soil erosion due to it's structural stabilization capacity. Fertilizers are becoming more essential to replenish mineral nutrients for prolonged cropping; however, since the soils in the regions are likely to continue undergoing more and more land use pressures with less frequent fallows to

285 restore the soil organic C pools and as slash-and-burn agriculture is still being practiced, careful soil quality monitoring will probably become more important, too. Thus, assessing the chemical environment such as pH and improving farming practices (i.e., considering nutrient (re)cycling) while closing the yield gap and maintaining soil quality attributes will be paramount to sustain soils' ecosystem functions over time. Quick and reasonably accurate soil estimates derived from mid-IR spectra and empiric models as for example outlined in this study can inform site-adapted timing, placing and form of nutrient supply

290 based on local soil conditions. Specifically, to give an example, light-textured soils can achieve high tuber yields but are at risk of loosing a large proportion of applied N and K (e.g., O'Sullivan, 2010) — which are both demanded in relatively large quantities by yam — to the environment (e.g., Diby et al., 2011). Hence, more frequent and local mineral applications of these nutrients after crop emergence, eventually combined with organic mulch, could improve the fertilizer efficiency and mitigate negative environmental impacts under these soil conditions. To estimate the availability of specific (micro)nutrients, however,

295 more efforts need to be made to measure them in fine temporal and spatial resolution.

The mid-IR model accurately estimated C (RMSE = 1.6g kg^{-1} soil; Table 1; Figure 3). Mostly, only field-scale spectroscopic models achieve such accuracy (Nocita et al., 2015; Guerrero et al., 2016), whereas the predictive accuracy reported for larger-scale application of spectroscopic models is lower than for our model (Rossel and Webster, 2012; Stevens et al., 2013; Sila et al., 2016). Models covering a wide geographical range of soils often result in high prediction errors (Stenberg and Rossel,

300 2010). Despite different soil types and climate regimes across a wide geographic spacing between the calibration fields, we achieved an accurate spectroscopic estimation of total C. The model was also able to reliably estimate a range of other important soil properties than total C. Specifically, other soil variables eligible for a mid-IR quantification include total N, total S, total Ca, total K, total Al, exchangeable Ca, Fe DTPA, CEC$_{\text{eff.}}$, and clay content ($R^2 > 0.75$). The high correlations of total C to N, S, exchangeable Ca, exchangeable Mg, CEC$_{\text{eff.}}$, total Ca, Al, and clay content (Figure 2) are consistent with Johnson et al.

305 (2019), who reported very similar associations of clay content and exchangeable cations (Ca, Mg, K) as well as CEC$_{\text{eff.}}$ in soils from rice fields ($0.54 \leq r \leq 0.65$) — nevertheless they spectrally modeled a considerable soil variability (20 countries in sub-

Saharan Africa; 42 study sites) and a larger sample size ($n = 285$) using PLS regression. At the same time, the measured range and the error in spectral estimates of CEC were larger compared to ours (RMSE = 6.7 cmol(+) kg^{-1} vs. 1.4 cmol(+) kg^{-1}; range = 1.9–66.5 cmol(+) kg^{-1} vs. 0.9–14.6 cmol(+) kg). Even though, total K and Fe(DTPA) were poorly correlated to total

310 C, their spectroscopic estimates were relatively accurate. This suggests that the mid-IR prediction of other soil properties is largely based on their correlation with total C as well as other absorption features of many organic and mineral soil components having a specific IR adsorption.

We also found reasonable prediction accuracy for Cu(DTPA) ($R^2 = 0.74$) and Mn(DTPA) ($R^2 = 0.55$), despite that soil nutrients that are extraction-based or dependent on surface chemistry usually have variable predictive performance (Janik

315 et al., 1998). Since relationships between soil composition and soil matrix exchange processes are typically complex, some properties may not be represented in the models in a straight-forward manner (Janik et al., 1998; Nocita et al., 2015). Although the spectral estimates for resin-extractable P were not eligible for accurate soil diagnostics ($R^2 = 0.40$), they might be sufficient for the screening of site conditions. Singh et al. (2019) also reported varying applicability of available P within four regions of cocoa smallholder farms in Papua New Guinea ($0.12 \leq R^2 \leq 0.58$). A further aspect of interest is that resin-extractable

320 P was mainly correlated to total P ($r = 0.78$), CEC$_{eff.}$ ($r = 0.60$) and total C ($r = 0.52$), while it poorly did to silt, sand and clay ($-0.19 \leq r \leq 0.16$). Despite that iron and aluminum oxides as well as clay minerals commonly cause P sorption (Gérard, 2016), we found no correlation between resin-extractable P and total Fe or Fe (DTPA) ($r = 0.05$ and -0.03). This most likely is because the form rather than the abundance of iron oxides (poorly crystalline Fe) and kaolins controls the P adsorbed and available to crops in soils (Hartwig and Loeppert, 1993).

325 ## 3.3 Interpretation of spectral features

All mid-IR spectra that we measured for soils in the four landscapes exhibited a similar pattern of absorbance (Figure 4). The O-Si-O absorptions in quartz at 1080 cm^{-1}, 800–780 cm^{-1} and 700 cm^{-1} were a prominent feature in the spectra due to relatively high sand contents across the landscapes (range 30% to 92%, median 76%). Our spectra further had hydroxyl (OH) absorptions that are typical for kaolin minerals, at 3695 cm^{-1} (surface OH), 3620 cm^{-1} (inner OH), 914 cm^{-1} (inner

330 OH), and 936 cm^{-1} (surface OH) (Madejová et al., 2002). The spectral pattern between the hydroxyl bands at 3695 cm^{-1} and 3620 cm^{-1} was relatively consistent and the intensity ratio of these flanking peaks was close to 1. This is typical for halloysite (0.8–0.9) while the ratio for kaolinite is often higher (1.2–1.5) and dickite lower (0.6–0.8) (Lyon and Tuddenham, 1960). The two weak intermediary stretching absorptions at around 3657 cm^{-1} and 3670 cm^{-1} indicate surface hydroxyls. Together with the absorption at 936 cm^{-1}, the spectra would suggest the presence of rather well-ordered prismatic halloysite (Hillier et al.,

335 2016). This aligns well with the spectral patterns of soils that were assigned to the Halloysite archetype through similarity mapping (by comparison to the pure mineral spectra) by Sila et al. (2016). Our spectra confirm the presence of kaolin minerals, which reflects the advanced state of mineral weathering in these tropical soil types.

Our accurate predictions, which are comparable to field-scale calibrations, are most likely because of the relatively uniform mid-IR spectra we obtained for our samples. This suggests a relatively homogeneous soil chemical composition, particularly

340 with regard to the mineralogy in the sampled soils. Still, the data set presented here is relatively small and no randomized

spatial sampling strategy was used for selecting field locations. Therefore, we propose to implement a spectroscopy-driven approach to diagnose soils from other yam fields as an effort to broaden the library to achieve better spatial coverage of soil variability.

4   Conclusions

345   We developed models with mid-IR spectra to estimate soil chemical and physical properties relevant to production of yam and other staple crops in four landscapes in the yam belt of West Africa. We tested the models for the important soil properties that are applied widely for agronomic performance evaluation. We showed that mid-IR spectroscopy models have the potential to cost-effectively and rapidly determine the distribution and variability of important soil properties across highly variable yam production landscapes in West Africa. Specifically, total C, total N, total S, total Fe, total Al, total K, total Ca, exchangeable Ca,

350   CEC_{eff}, Fe(DTPA), and clay content can be quantified with RPD > 2 and $R^2 > 0.75$ when aiming to predict in the range of soil property values found in the environmental conditions covered by this study. We achieved spectral estimates with quite small uncertainties, that are typically reported for libraries at the scale of a field or farm. The correlation analysis of measured values together with spectral inference helps improve our understanding of how soil properties are interrelated with soil functional composition. This study delivered parsimonious, unbiased and accurate mid-IR spectroscopy-based models to monitor and

355   predict soil quality and to manage crop nutrition. Hence, we envision this pilot study as being a starting point to continuously update and adapt the mid-IR model library for more efficient site-specific and agronomically relevant soil estimates in the West African yam belt. This can bring better capacity to diagnose and and long-term monitor soils compared with traditional wet chemistry, and will hopefully ameliorate the soil conditions for sustainably meeting the demand of yam and other important staple crops in the regions.

360   *Code and data availability.*   All data and code to reproduce the results of this publication are publicly available under GNU General Public License v3.0, and can be accessed via the Zenodo archive and the corresponding github public repository (Baumann, 2018)

Author contributions.   Philipp Baumann carried out the research and analysis (soil sampling, sample preparation, soil chemical analysis, infrared spectroscopy, statistical modeling) under continuous support of the YAMSYS project team, and took the lead in writing the manuscript. All co-authors helped to improve the manuscript. Johan Six, Juhwan Lee and Emmanuel Frossard framed the idea of delivering validated

365   models for soil properties relevant for yam growth in the four pilot regions in Burkina Faso and Ivory Coast. Valérie Hgaza and Delwende Kiba contributed to the selection representative yam fields that were sampled for our work.

Competing interests.   We declare that we have no conflict of interest.

Acknowledgements. This study has been done within the YAMSYS project (www.yamsys.org) funded by the food security module of the Swiss Programme for Research on Global Issues for Development (r4d programme; www.r4d.ch) (SNF project number: 400540_152017 / 1). We would like to express gratitude to the site managers — Marie Leance Kouassi, Marcel Soma, and Augustin Kangah N'da — and the 80 farmers that strongly supported our sampling endeavor and the idea of developing diagnostic and management innovations for improved yam growth. We would like to thank Nestor Pouya and Carole Werdenberg who assisted with soil sampling. Our thanks also goes to Dr. Bahar Aciksöz, Michele Wyler, and Patricia Schwitter, who helped with sample milling, weighting, and acid digestion. We would also like to thank Dr. Federica Tamburini for support in the CNS analysis, Éva Mészaros for her advice on the resin P protocol, and Björn Studer for the opportunity to perform XRF analyses on soils. We thank Raphael Viscarra Rossel and Marijn Van de Broek for their valuable feedback, which helped us to improve the manuscript.

References

Abbott, L. K. and Murphy, D. V., eds.: Soil Biological Fertility: A Key to Sustainable Land Use in Agriculture, Springer Netherlands, //www.springer.com/de/book/9781402017568, 2007.

380 Baumann, P.: philipp-baumann/yamsys-soilspec-publication: Pre- release of spectral models for the publication of the soil spectral library of the West African yam belt, https://doi.org/10.5281/zenodo.1174869, https://doi.org/10.5281/zenodo.1174869, 2018.

Baumann, P.: philipp-baumann/simplerspec: Beta release simplerspec 0.1.0 for zenodo, https://doi.org/10.5281/zenodo.3303637, https://doi.org/10.5281/zenodo.3303637, 2019.

Bouyoucos, G. J.: A recalibration of the hydrometer method for making mechanical analysis of soils, Agronomy journal, 43, 434–438, 1951.

385 Breiman, L., Friedman, J., Stone, C., and Olshen, R.: Classification and Regression Trees, The Wadsworth and Brooks-Cole statistics-probability series, Taylor & Francis, https://books.google.ch/books?id=JwQx-WOmSyQC, 1984.

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C., and Barthès, B. G.: Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field, Geoderma, 261, 151–159, https://doi.org/10.1016/j.geoderma.2015.07.007, http://linkinghub.elsevier.com/retrieve/pii/S0016706115300185, 2016.

390 Carsky, R. J., Asiedu, R., and Cornet, D.: Review of soil fertility management for yam-based systems in west africa, African Journal of Root and Tuber Crops, 8, 1, http://www.researchgate.net/profile/Denis_Cornet/publication/260988837_Review_of_soil_fertility_management_for_yam-based_systems_in_West_Africa/links/0c960532ef7c8ca5b2000000.pdf, 2010.

Chong, I.-G. and Jun, C.-H.: Performance of some variable selection methods when multicollinearity is present, Chemometrics and Intelligent Laboratory Systems, 78, 103–112, https://doi.org/10.1016/j.chemolab.2004.12.011, http://linkinghub.elsevier.com/retrieve/pii/S0169743905000031, 2005.

395 Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N. P., Bernoux, M., and Barthès, B. G.: National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy, Geoderma, 276, 41–52, https://doi.org/10.1016/j.geoderma.2016.04.021, http://linkinghub.elsevier.com/retrieve/pii/S001670611630180X, 2016.

Cornet, D., Sierra, J., Tournebize, R., Gabrielle, B., and Lewis, F. I.: Bayesian Network Modeling of Early Growth Stages Explains

400 Yam Interplant Yield Variability and Allows for Agronomic Improvements in West Africa, European Journal of Agronomy, 75, 80–88, https://doi.org/10.1016/j.eja.2016.01.009, http://www.sciencedirect.com/science/article/pii/S1161030116300090, 2016.

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A., and Brun, J. J.: Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS), European Journal of Soil Science, 60, 770–784, https://doi.org/10.1111/j.1365-2389.2009.01178.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2389.2009.01178.x/abstract, 2009.

405 Diby, L. N., Hgaza, V. K., Tie, T. B., ASSA, A., Carsky, R., Girardin, O., and Frossard, E.: Productivity of Yams (Dioscorea Spp.) as Affected by Soil Fertility, Journal of Animal & Plant Sciences, 5, 494–506, http://www.m.elewa.org/JAPS/2009/5.2/5.pdf, 2009.

Diby, L. N., Tie, B. T., Girardin, O., Sangakkara, R., and Frossard, E.: Growth and Nutrient Use Efficiencies of Yams (Dioscorea Spp.) Grown in Two Contrasting Soils of West Africa, https://doi.org/10.1155/2011/175958, https://www.hindawi.com/journals/ija/2011/175958/, 2011.

410 Enyi, B. a. C.: Effect of Staking, Nitrogen and Potassium on Growth and Development in Lesser Yams: Dioscorea Esculenta, Annals of Applied Biology, 72, 211–219, https://doi.org/10.1111/j.1744-7348.1972.tb01287.x, 1972.

FAO: World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps., FAO, oCLC: 1004475995, 2014.

Food and Agriculture Organization of the United Nations: FAOSTAT statistics database, www.fao.org/faostat/, 2019.

415 Frossard, E., Aighewi, B. A., Aké, S., Barjolle, D., Baumann, P., Bernet, T., Dao, D., Diby, L. N., Floquet, A., Hgaza, V. K., Ilboudo, L. J., Kiba, D. I., Mongbo, R. L., Nacro, H. B., Nicolay, G. L., Oka, E., Ouattara, Y. F., Pouya, N., Senanayake, R. L., Six, J., and Traoré, O. I.: The Challenge of Improving Soil Fertility in Yam Cropping Systems of West Africa, Frontiers in Plant Science, 8, https://doi.org/10.3389/fpls.2017.01953, http://journal.frontiersin.org/article/10.3389/fpls.2017.01953/full, 2017.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A. M., Gabarrón-Galeote, M. A., Ruiz-Sinoga, J. D., Zornoza, R., and Viscarra Rossel,
420 R. A.: Do We Really Need Large Spectral Libraries for Local Scale SOC Assessment with NIR Spectroscopy?, Soil and Tillage Research, 155, 501–509, https://doi.org/10.1016/j.still.2015.07.008, http://linkinghub.elsevier.com/retrieve/pii/S0167198715001567, 2016.

Gérard, F.: Clay minerals, iron/aluminum oxides, and their contribution to phosphate sorption in soils — A myth revisited, Geoderma, 262, 213–226, https://doi.org/10.1016/j.geoderma.2015.08.036, http://www.sciencedirect.com/science/article/pii/S0016706115300653, 2016.

Hartwig, R. C. and Loeppert, R. H.: EVALUATION OF SOIL IRON, in: Iron Chelation in Plants and Soil Microorganisms, pp. 465–482, El-
425 sevier, https://doi.org/10.1016/B978-0-12-079870-4.50027-2, https://linkinghub.elsevier.com/retrieve/pii/B9780120798704500272, 1993.

Hendershot, W. H. and Duquette, M.: A simple barium chloride method for determining cation exchange capacity and exchangeable cations, Soil Science Society of America Journal, 50, 605–608, 1986.

Hgaza, V. K., Diby, L. N., Tié, T. B., Tschannen, A., Aké, S., Assa, A., and Frossard, E.: Growth and Distribution of Roots of Dioscorea Alata L. Do Not Respond to Mineral Fertilizer Application, Open Plant Science Journal, 5, 14–22, 2011.

430 Hillier, S., Brydson, R., Delbos, E., Fraser, T., Gray, N., Pendlowski, H., Phillips, I., Robertson, J., and Wilson, I.: Correlations among the mineralogical and physical properties of halloysite nanotubes (HNTs), Clay Minerals, 51, 325–350, https://doi.org/10.1180/claymin.2016.051.3.11, https://www.cambridge.org/core/product/identifier/S0009855800006774/type/journal_article, 2016.

Janik, L. J., Skjemstad, J. O., and Merry, R. H.: Can mid infrared diffuse reflectance analysis replace soil extractions?, Australian Journal of
435 Experimental Agriculture, 38, 681, https://doi.org/10.1071/EA97144, http://www.publish.csiro.au/?paper=EA97144, 1998.

Johnson, J.-M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K. D., and Saito, K.: Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa, Geoderma, 354, 113 840, https://doi.org/10.1016/j.geoderma.2019.06.043, https://linkinghub.elsevier.com/retrieve/pii/S001670611930357X, 2019.

Kang, B. T. and Wilson, J. E.: Effect of mound size and fertilizer on white Guinea yam (Dioscorea rotundata) in Southern Nigeria, Plant and
440 Soil, 61, 319–327, https://doi.org/10.1007/BF02182013, http://link.springer.com/10.1007/BF02182013, 1981.

Kassi, S.-P. A., Koné, A. W., Tondoh, J. E., and Koffi, B. Y.: Chromoleana Odorata Fallow-Cropping Cycles Maintain Soil Carbon Stocks and Yam Yields 40 Years after Conversion of Native- to Farmland, Implications for Forest Conservation, Agriculture, Ecosystems & Environment, 247, 298–307, https://doi.org/10.1016/j.agee.2017.06.044, https://linkinghub.elsevier.com/retrieve/pii/S0167880917302876, 2017.

Kim, J.-H.: Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap, Computational Statistics & Data
445 Analysis, 53, 3735–3745, https://doi.org/10.1016/j.csda.2009.04.009, http://linkinghub.elsevier.com/retrieve/pii/S0167947309001601, 2009.

Kouno, K., Tuchiya, Y., and Ando, T.: Measurement of soil microbial biomass phosphorus by an anion exchange membrane method, Soil Biology and Biochemistry, 27, 1353 – 1357, https://doi.org/http://dx.doi.org/10.1016/0038-0717(95)00057-L, http://www.sciencedirect.com/science/article/pii/003807179500057L, 1995.

450   Kuhn, M., Wing, J., Weston, S., A., W., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T.: caret: Classification and Regression Training, https://CRAN.R-project.org/package=caret, r package version 6.0-82, 2019.

Lindsay, W. L. and Norvell, W. A.: Development of a DTPA soil test for zinc, iron, manganese, and copper, Soil science society of America journal, 42, 421–428, 1978.

455   Lyon, R. J. P. and Tuddenham, W. M.: Infra-Red Determination of the Kaolin Group Minerals, Nature, 185, 835–836, https://doi.org/10.1038/185835a0, https://www.nature.com/articles/185835a0, number: 4716 Publisher: Nature Publishing Group, 1960.

Madejová, J., Kečkés, J., Pálková, H., and Komadel, P.: Identification of components in smectite/kaolinite mixtures, Clay Minerals, 37, 377–388, https://doi.org/10.1180/0009855023720042, http://www.catchword.com/cgi-bin/cgi?ini=xref&body=linker&reqdoi=10.1180/0009855023720042, 2002.

460   Martens, H. and Naes, T.: Multivariate Calibration, Wiley Chichester, 1989.

Mevik, B.-H., Wehrens, R., and Liland, K. H.: pls: Partial Least Squares and Principal Component Regression, https://CRAN.R-project.org/package=pls, r package version 2.7-1, 2019.

Molinaro, A. M., Simon, R., and Pfeiffer, R. M.: Prediction error estimation: a comparison of resampling methods, Bioinformatics, 21, 3301–3307, https://doi.org/10.1093/bioinformatics/bti499, https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/

465   bti499, 2005.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J. M., Shepherd, K. D., Stenberg, B., Towett, E. K., Vargas, R., and Wetterlind, J.: Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring, in: Advances in Agronomy, vol. 132, pp. 139–159, Elsevier, http://linkinghub.elsevier.com/retrieve/

470   pii/S0065211315000425, dOI: 10.1016/bs.agron.2015.02.002, 2015.

Nuernberg, N. J., Leal, J. E., and Sumner, M. E.: Evaluation of an anion-exchange membrane for extracting plant available phosphorus in soils, Communications in Soil Science and Plant Analysis, 29, 467–479, https://doi.org/10.1080/00103629809369959, http://www.tandfonline.com/doi/abs/10.1080/00103629809369959, 1998.

Ohno, T. and Zibilske, L. M.: Determination of Low Concentrations of Phosphorus in Soil Extracts Using Malachite Green, Soil Science So-

475   ciety of America Journal, 55, 892, https://doi.org/10.2136/sssaj1991.03615995005500030046x, https://www.soils.org/publications/sssaj/abstracts/55/3/SS0550030892, 1991.

O'Sullivan, J. N.: Yam nutrition nutrient disorders and soil fertility management, ACIAR, Canberra, oCLC: 1074816829, 2010.

O'Sullivan, J. N. and Jenner, R.: Nutrient Deficiencies in Greater Yam and Their Effects on Leaf Nutrient Concentrations, Journal of Plant Nutrition, 29, 1663–1674, https://doi.org/10.1080/01904160600851569, http://www.tandfonline.com/doi/abs/10.1080/01904160600851569,

480   2006.

Padwick, G. W.: Fifty Years of *Experimental Agriculture* II. The Maintenance of Soil Fertility in Tropical Africa: A Review, Experimental Agriculture, 19, 293–310, https://doi.org/10.1017/S001447970001276X, https://www.cambridge.org/core/product/identifier/S001447970001276X/type/journal_article, 1983.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https:

485   //www.R-project.org/, 2017.

Rossel, R. A. V. and Webster, R.: Predicting soil properties from the Australian soil visible–near infrared spectroscopic database, European Journal of Soil Science, 63, 848–860, https://doi.org/10.1111/j.1365-2389.2012.01495.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2389.2012.01495.x/abstract, 2012.

Rossel, R. V. and Behrens, T.: 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, http://linkinghub.elsevier.com/retrieve/pii/S0016706109004315, 2010.

Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Analytical Chemistry, 36, 1627–1639, https://doi.org/10.1021/ac60214a047, 1964.

Sila, A. M., Shepherd, K. D., and Pokhariyal, G. P.: Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties, Chemometrics and Intelligent Laboratory Systems, 153, 92–105, https://doi.org/10.1016/j.chemolab.2016.02.013, http://linkinghub.elsevier.com/retrieve/pii/S0169743916300351, 2016.

Singh, K., Vasava, H. B., Snoeck, D., Das, B. S., Yinil, D., Field, D., Sanderson, T., Fidelis, C., Majeed, I., and Panigrahi, N.: Assessment of cocoa input needs using soil types and soil spectral analysis, Soil Use and Management, 35, 492–502, https://doi.org/10.1111/sum.12499, https://onlinelibrary.wiley.com/doi/abs/10.1111/sum.12499, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sum.12499, 2019.

Soares, M. R. and Alleoni, L. R. F.: Contribution of Soil Organic Carbon to the Ion Exchange Capacity of Tropical Soils, Journal of Sustainable Agriculture, 32, 439–462, https://doi.org/10.1080/10440040802257348, http://www.tandfonline.com/doi/abs/10.1080/10440040802257348, 2008.

Stenberg, B. and Rossel, R. A. V.: Diffuse Reflectance Spectroscopy for High-Resolution Soil Sensing, in: Proximal Soil Sensing, edited by Rossel, R. A. V., McBratney, A. B., and Minasny, B., Progress in Soil Science, pp. 29–47, Springer Netherlands, http://link.springer.com/chapter/10.1007/978-90-481-8859-8_3, dOI: 10.1007/978-90-481-8859-8_3, 2010.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., and van Wesemael, B.: Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy, PLoS ONE, 8, e66 409, https://doi.org/10.1371/journal.pone.0066409, http://dx.plos.org/10.1371/journal.pone.0066409, 2013.

Syers, J. K., Campbell, A. S., and Walker, T. W.: Contribution of organic carbon and clay to cation exchange capacity in a chronosequence of sandy soils, Plant and Soil, 33, 104–112, https://doi.org/10.1007/BF01378202, http://link.springer.com/article/10.1007/BF01378202, 1970.

UNEP: Land Health Surveillance: An Evidence-Based Approach to Land Ecosystem Management. Illustrated with a Case Study in the West Africa Sahel, 2012.

Vagen, T.-G., Shepherd, K. D., Walsh, M. G., Winowiecki, L., Desta, L. T., and Tondoh, J. E.: AfSIS technical specifications: Soil Health Surveillance., http://www.worldagroforestry.org/sites/default/files/afsisSoilHealthTechSpecs_v1_smaller.pdf, 2010.

Viscarra Rossel, R., Behrens, T., Ben-Dor, E., Brown, D., Demattê, J., Shepherd, K., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B., Bartholomeus, H., Bayer, A., Bernoux, M., Böttcher, K., Brodský, L., Du, C., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C., Knadel, M., Morrás, H., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. R., Sanborn, P., Sellitto, V., Sudduth, K., Rawlins, B., Walter, C., Winowiecki, L., Hong, S., and Ji, W.: A global spectral library to characterize the world's soil, Earth-Science Reviews, 155, 198–230, https://doi.org/10.1016/j.earscirev.2016.01.012, http://linkinghub.elsevier.com/retrieve/pii/S0012825216300113, 2016.

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., and Skjemstad, J. O.: Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties, Geoderma, 131, 59–75, https://doi.org/10.1016/j.geoderma.2005.03.007, http://www.sciencedirect.com/science/article/pii/S0016706105000728, 2006.

Wold, S., Martens, H., and Wold, H.: The Multivariate Calibration Problem in Chemistry Solved by the PLS Method, in: Matrix Pencils,

525   edited by Kågström, B. and Ruhe, A., vol. 973, pp. 286–293, Springer Berlin Heidelberg, https://doi.org/10.1007/BFb0062108, http://link.springer.com/10.1007/BFb0062108, 1983.

Wold, S., Johansson, E., and Cocchi, M.: PLS-partial least squares projections to latent structures, 3D QSAR in drug design, 1, 523–550, 1993.