



# Mineral–nutrient relationships in African soils assessed using cluster analysis of X-ray powder diffraction patterns and compositional methods

Benjamin M. Butler<sup>a,\*</sup>, Javier Palarea-Albaladejo<sup>b</sup>, Keith D. Shepherd<sup>c</sup>, Kamau M. Nyambura<sup>d</sup>, Erick K. Towett<sup>c</sup>, Andrew M. Sila<sup>c</sup>, Stephen Hillier<sup>a,e</sup>

<sup>a</sup> The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK

<sup>b</sup> Biomathematics & Statistics Scotland, JCMB, The King's Buildings, Edinburgh EH9 3FD, UK

<sup>c</sup> World Agroforestry Centre (ICRAF), P.O. Box 30677-00100 GPO, Nairobi, Kenya

<sup>d</sup> International Fertiliser Development Centre, c/o icipe Campus, P.O. Box 30772-00100, Nairobi, Kenya

<sup>e</sup> Department of Soil and Environment, Swedish University of Agricultural Sciences (SLU), SE-75007 Uppsala, Sweden

## ARTICLE INFO

Handling Editor: Yvan Capowicz

### Keywords:

Macro-nutrients  
Micro-nutrients  
Soil mineralogy  
X-ray powder diffraction  
Cluster analysis  
Compositional data analysis

## ABSTRACT

Soil mineral compositions are often complex and spatially diverse, with each mineral exhibiting characteristic chemical properties that determine the intrinsic total concentration of soil nutrients and their phyto-availability. Defining soil mineral–nutrient relationships is therefore important for understanding the inherent fertility of soils for sustainable nutrient management, and data-driven approaches such as cluster analysis allow for these relations to be assessed in new detail.

Here the fuzzy-c-means clustering algorithm was applied to an X-ray powder diffraction (XRPD) dataset of 935 soils from sub-Saharan Africa, with each diffractogram representing a digital signature of a soil's mineralogy. Nine mineralogically distinct clusters were objectively selected from the soil mineralogy continuum by retaining samples exceeding the 75% quantile of the membership coefficients in each cluster, yielding a dataset of 239 soils. As such, samples within each cluster represented mineralogically similar soils from different agro-ecological environments of sub-Saharan Africa. Mineral quantification based on the mean diffractogram of each cluster illustrated substantial mineralogical diversity between the nine groups with respect to quartz, K-feldspar, plagioclase, Fe/Al/Ti-(hydr)oxides, phyllosilicates (1:1 and 2:1), ferromagnesians, and calcite.

Mineral–nutrient relationships were defined using the clustered XRPD patterns and corresponding measurements of total and/or extractable (Mehlich-3) nutrient concentrations (B, Mg, K, Ca, Mn, Fe, Ni, Cu and Zn) in combination with log-ratio compositional data analysis. Fe/Al/Ti/Mn-(hydr)oxides and feldspars were found to be the primary control of total nutrient concentrations, whereas 2:1 phyllosilicates were the main source of all extractable nutrients except for Fe and Zn. Kaolin minerals were the most abundant phyllosilicate group within the dataset but did not represent a nutrient source, which reflects the lack of nutrients within their chemical composition and their low cation exchange capacity. Results highlight how the mineral composition controls the total nutrient reserves and their phyto-availability in soils of sub-Saharan Africa. The typical characterisation of soils and their parent material based on the clay particle size fraction (i.e. texture) and/or the overall silica component (i.e. acid and basic rock types) alone may therefore mask the intricacies of mineral contributions to soil nutrient concentrations.

## 1. Introduction

Minerals are the major component of most soils. Through direct inheritance from the parent material and subsequent alteration by chemical weathering, the soil mineral composition can be spatially diverse - reflecting the many soil forming factors (Jenny, 1994). Minerals present within the soil environment exhibit characteristic crystal

structures, chemical compositions and chemical properties (Dixon and Schulze, 2002) that determine the total reserves of essential plant nutrients (White and Brown, 2010). Aside from these total nutrient reserves, the fate, phyto-availability and toxicity of these nutrients depends upon the form in which they occur (Cornu et al., 2009) along with associated dissolution–precipitation, adsorption–desorption and reduction–oxidation reactions (Singh and Schulze, 2015).

\* Corresponding author.

E-mail address: [benjamin.butler@hutton.ac.uk](mailto:benjamin.butler@hutton.ac.uk) (B.M. Butler).

<https://doi.org/10.1016/j.geoderma.2020.114474>

Received 14 December 2019; Received in revised form 19 May 2020; Accepted 22 May 2020

Available online 09 June 2020

0016-7061/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Soil macro- and micro-nutrient concentrations and their phyto-availability are inherently related to soil fertility (Ajiyoye et al., 2019). Understanding mineral contributions to soil nutrient reserves is a key priority at present given the threat of negative soil nutrient balances to food security (Sutton et al., 2013; Omuto and Vargas, 2018). Further, estimates suggest that of cultivated soils suffer from growth-limiting problems relating to nutrient deficiencies and/or toxicities (Cakmak, 2002; Schjoerring et al., 2019; Keskinen et al., 2019). In particular, many African soils are considered to have a typically low fertility caused by a lack of volcanic/tectonic rejuvenation, resulting in long and repeated cycles of weathering, erosion and leaching predominating over millions of years. This leaves the soils poor in nutrients, especially soils derived from basement complex rocks and aeolian sands (Voortman et al., 2003). Furthermore, since the use of fertiliser in sub-Saharan African crop production is very limited, there is a dependence on inherent reserves of soil nutrients (Vitousek et al., 2009; Vanlauwe et al., 2014; Keskinen et al., 2019), of which mineralogy is a key component.

Indeed, soil fertility degradation has been described as the single most important constraint on food security in sub-Saharan Africa (Smaling, 1994). By 2050, the population of sub-Saharan Africa is expected to increase 2.5-fold and demand for cereals approximately triple (Van Ittersum et al., 2016). To maintain the current level of cereal self-sufficiency of approximately 80% by 2050, nearly complete closure of the gap between current farm yields and water-limited yield potential is needed, which is in the range of 20% to 50% (Van Ittersum et al., 2016). Meeting these demands for food production will require an increase in nutrient inputs combined with a better understanding of the intrinsic soil mineral reserves of plant nutrients.

X-ray powder diffraction (XRPD) measurements are commonly used to identify and quantify soil mineralogy because diffraction data are fundamentally related to the crystal structure and crystal chemistry of the soil minerals present (Schulze, 1989). A typical soil XRPD pattern (diffractogram) is comprised of discrete 'Bragg' diffraction peaks varying in intensity ( $y$ ) that are distributed along an experimental axis ( $x$ ) usually expressed in degrees  $2\theta$ . The 'Bragg' peaks rise above a background that often includes diffuse scattering from X-ray amorphous components such as soil organic matter or volcanic glass. The mineralogical detail associated with soil XRPD analysis therefore makes it arguably the most powerful approach to accurately identify and quantify the complex suite of minerals in the soil environment. Classically, identifying and quantifying soil minerals from XRPD data is a time and labour intensive process, requiring each diffractogram to be manually inspected and analysed in combination with mineral databases [e.g. the Powder Diffraction File; ICDD, 2019] and specialised computer software. Such workflows become inconvenient when processing large numbers of samples, and the recent acquisition of high-throughput XRPD datasets containing thousands of soil diffractograms has promoted application of alternative, data-driven, approaches to soil XRPD data for the first time (Butler et al., 2018; Butler et al., 2019; Hillier and Butler, 2018). These approaches remove the need for classical expert interpretation in the initial stages of analysis, which can be particularly challenging and time consuming in diverse datasets of this size.

One such high-throughput soil XRPD dataset is that of the Africa Soil Information Service (AFSIS) Sentinel sites, which contains approximately 2000 georeferenced soil samples from sub-Saharan Africa that have each been analysed by XRPD (Vågen et al., 2010; Towett et al., 2013). Associated with each XRPD measurement is a range of site attributes and soil property measurements, including total and extractable nutrient concentrations. By treating each diffractogram within this dataset as a reproducible mineralogical signature of a soil, the data can be combined with data-driven analysis and related to the associated nutrient concentrations in order to identify and interpret mineral contributions to soil nutrient concentrations. The application of data-driven analysis to soil XRPD data defines a new concept for soil

mineralogy research which has been labelled 'Digital Soil Mineralogy' (Hillier and Butler, 2018).

Cluster analysis is a branch of multivariate statistical analysis for unsupervised machine learning and has notable potential to facilitate expert interpretation of soil mineralogy–nutrient relationships in large XRPD datasets by grouping the data into a manageable number of classes. Each cluster defined from soil XRPD data should represent a group of mineralogically similar soils that are mineralogically distinct from those in other clusters. Recently, a suitable protocol for clustering soil XRPD data has been determined (Butler et al., 2019), and the present study uses this protocol to investigate mineral–nutrient relationships in soils sampled from a diverse range of agro-ecological environments in sub-Saharan Africa (Vågen et al., 2010). Analysis includes a suite of macro-nutrients (Mg, K and Ca) and micro-nutrients (B, Mn, Fe, Ni, Cu and Zn), both in terms of total and extractable concentrations. Novel statistical approaches are applied that account for the compositional nature of nutrient concentration data. That is, multivariate data conveying relative information, with values of the constituting components representing fractions of a total (e.g. 100 or  $10^6$  when expressed respectively in percentage or ppm units). Such data are only meaningful in relation to each other, regardless of the measurement scale used, and compositional data analysis methods seek to ensure that meaningful conclusions are obtained on this basis (Aitchison, 1982). In combining soil XRPD patterns and nutrient concentrations with cluster analysis and compositional methods, we generate new understanding that can contribute to a move towards more sustainable and mineralogically tailored land management practices.

## 2. Sample collection and laboratory analyses

### 2.1. Soil sampling

Georeferenced samples associated with the AFSIS project (Vågen et al., 2010) were taken from the set of sixty  $10 \times 10$  km 'Sentinel' sites distributed across sub-Saharan Africa (Fig. 1). The Sentinel sites were designed to be statistically representative of the variability in climate, topography and vegetation of sub-Saharan Africa (Vågen et al., 2010; Hengl et al., 2017). Field sampling was conducted based on the field methods collectively referred to as the Land Degradation Surveillance Framework (LDSF) protocol (Vågen et al., 2010; Vågen et al., 2015). Each Sentinel site was divided into 16 grid cells of equal size, and a soil sampled at a random location within each grid cell (Fig. 1).

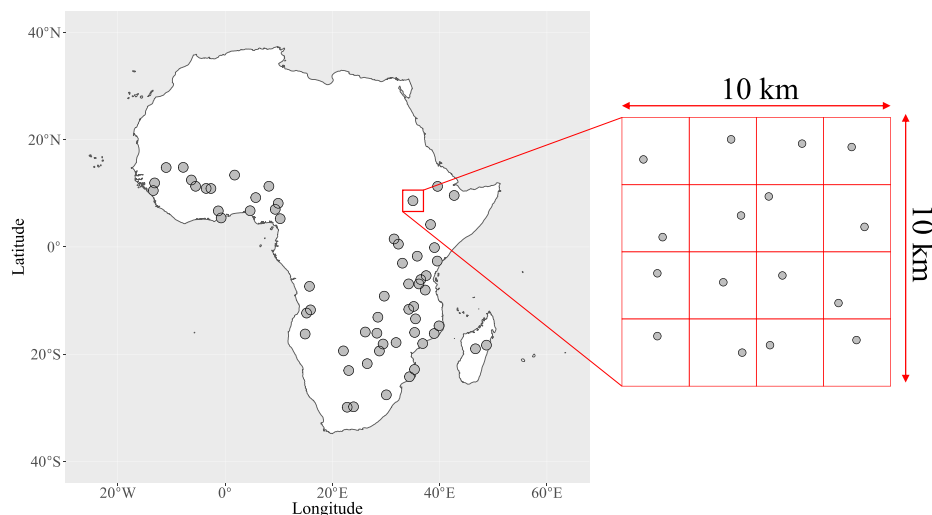
Only subsoil samples were investigated here with the aim of minimising the potential effects of land management practices (see Table 1 in Towett et al., 2015) and soil organic matter on nutrient concentrations, both of which would be more prominent in topsoils. In doing so the naturally occurring mineral contributions to soil element concentrations could be focused upon. Subsoil samples were collected at depths of 20 – 50 cm from each grid cell at each Sentinel site, giving a total of 960 soil samples (16 grid cells per site  $\times$  60 Sentinel sites) of which 935 were measured by XRPD.

### 2.2. Laboratory analyses

A total of 10 elements are examined herein to understand how nutrient and micro-nutrient concentrations relate to the mineralogy of African soils. These elements include:

1. The total organic carbon (TOC) concentration, used as an estimate of soil organic matter content.
2. Mg, K and Ca concentrations examined as essential plant macro-nutrients.
3. B, Mn, Fe, Ni, Cu and Zn concentrations examined as essential plant micro-nutrients.

Further to these 10 elements, soil pH and clay particle size fraction



**Fig. 1.** The survey design used for the sampling of Sentinel sites. Each of the sixty  $10 \times 10$  km sites was separated into a  $4 \times 4$  grid, and a soil at a random location within each grid cell sampled.

were also examined as common soil properties linked to plant nutrient provision. All laboratory analyses were conducted on the dry sieved  $<2$  mm size fraction.

### 2.2.1. Determination of total organic carbon, pH and clay particle size fraction

Samples were prepared for TOC analysis by acidification with hydrochloric acid and drying at  $60^\circ\text{C}$  using a modification of the protocol described in Harris et al., 2001. The carbon concentrations were then determined by thermal oxidation on a Thermoquest FlashEA 1112 equipped with an autoanalyser (Vågen et al., 2010).

Soil pH was determined with a combination electrode on the supernatant of a 1:2 soil:solution ratio. Clay particle size fraction was determined by laser diffraction as the percentage volume of  $<2\ \mu\text{m}$  sized particles (assuming spherical morphology) on a Horiba LA-950V2 Particle Size Analyser with a detectable size range of  $0.01 - 3000\ \mu\text{m}$  (Vågen et al., 2010). It is important to distinguish that the clay particle size fraction includes any mineral particle  $<2\ \mu\text{m}$  in diameter. Whilst this size delineation is most often associated with the clay minerals (i.e. phyllosilicates), this does not rule out the fact that other rock forming minerals can sometimes be present in the clay particle size fraction, and conversely phyllosilicates can have particle sizes  $>2\ \mu\text{m}$ . As such the clay particle size fraction may differ to the percentage of phyllosilicates/clay minerals determined from quantitative mineralogical analysis (Section 2.2.3).

### 2.2.2. Determination of nutrient concentrations

Total X-ray fluorescence spectroscopy (TXRF) was used to determine the total concentrations of K, Ca, Mn, Fe, Ni, Cu, and Zn (hereafter denoted by subscript 'T', e.g.  $\text{K}_T$ ). Detailed accounts of the TXRF protocol and limits of detection are provided in Towett et al., 2013; Towett et al., 2015. Briefly, 50 mg of air-dried ( $40^\circ\text{C}$ ) and ground ( $20-50\ \mu\text{m}$ ) soil sample was mixed with 2.5 ml of Triton X100 (Fisher) solution (0.1 vol.%) to form a soil suspension, and spiked with  $40\ \mu\text{l}$  of  $1000\ \text{mg l}^{-1}$  Selenium (Fluka) as the internal standard. The resulting suspension was placed into an ultrasonic water bath at room temperature and sonicated for 15 min, and then mixed well using a digital shaker.  $10\ \mu\text{l}$  of the turbid soil solution was then dispensed onto a clean siliconised quartz glass sample carrier and dried for 10 – 15 min at  $52^\circ\text{C}$ . Samples were analysed in triplicate on an S2 PICOFOX TXRF (Bruker) with a data acquisition time of 1000 s per sample. Spectral evaluation and element quantification were performed using the software SPECTRA 6.3 (Bruker).

Phyto-available nutrient concentrations were estimated using the

Mehlich-3 extraction protocol (Mehlich, 1984; Ziadi and Tran, 2007), hereafter referred to as 'M3'. The M3 extracting solution is composed of 0.2 N ethanoic acid ( $\text{CH}_3\text{COOH}$ ), 0.25 N ammonium nitrate ( $\text{NH}_4\text{NO}_3$ ), 0.015 N ammonium fluoride ( $\text{NH}_4\text{F}$ ), 0.013 N nitric acid ( $\text{HNO}_3$ ), and 0.001 M ethylenediaminetetraacetic acid ( $\text{C}_{10}\text{H}_{16}\text{N}_2\text{O}_8$ ). The method allows for multiple elements to be analysed from a single extraction, namely B, Mg, K, Ca, Mn, Fe, Cu and Zn in this case - hereafter denoted by subscript 'M', e.g.  $\text{K}_M$ . All extracted elements were quantified by inductively-coupled plasma optical emission spectroscopy using a Perkin Elmer Optima 8300 instrument.

### 2.2.3. X-ray powder diffraction and mineral quantification

Sub-sampling for XRPD analysis was conducted by coning and quartering. Sub-samples were then prepared for XRPD by McCrone milling 3 g of sieved ( $<2$  mm) and air-dried soil for 12 min in ethanol. Excess ethanol was removed by centrifugation and each sample re-suspended in 1.5 ml of hexane. The samples were then oven dried at  $80^\circ\text{C}$  before being disaggregated and ground by hand in a mortar and pestle before passing through a  $250\ \mu\text{m}$  sieve. Loading into the instrument was carried out by loosely filling the sample holders with the finely ground powders, before flattening the surface with the sharp edge of a razor blade, with personnel instructed to take care to apply minimum pressure and to avoid shearing motion. The combination of methodical milling and loading was designed to produce samples with appropriate particle statistics and to minimise preferred orientation (Zhang et al., 2003).

After loading, XRPD data were collected on a Bruker desktop D2 PHASER diffractometer, with Ni-filter,  $\text{Cu-K}\alpha$  radiation with the X-ray tube operated at 30 kV and 10 mA. The beam was collimated using a 0.6 mm divergence slit, a 1 mm anti-scatter slit and a 2.5 mm Soller slit. Samples were rotated continuously at 15 rpm during data collection over the angular range of  $3$  to  $75^\circ 2\theta$ , counting for 96 s per  $0.02^\circ$  step with a Lynxeye position sensitive detector.

To quantitatively summarise the mineralogy of each cluster, untreated diffractograms were aligned relative to a standard quartz pattern (ICDD, 2019; Butler et al., 2019), and the mean diffractogram of each cluster computed. Phases in each mean diffractogram were identified using the Powder Diffraction File database (ICDD, 2019), taking all peaks of each mineral component into account, and subsequently quantified by the full pattern summation method (Omotoso et al., 2006) as implemented in the 'powdR' package (Butler and Hillier, 2020) of the R language and environment for statistical computing (R Core Team, 2018). This approach treats an observed diffractogram as the sum of contributions from individual crystalline, para-crystalline and

**Table 1**  
Soil property summary statistics. Each property is summarised by its 10th, 25th, 50th, 75th and 90th percentiles for the whole dataset ( $n = 935$ ). Clay represents the clay particle size fraction ( $<2 \mu\text{m}$ ). All data are summarised by their geometric mean and geometric standard deviation except for pH, where the arithmetic mean and arithmetic standard deviation are used. The geometric standard deviation is related to the geometric mean by multiplication and division, represented by  $\times$ . The arithmetic standard deviation is related to the arithmetic mean by addition and subtraction, represented by  $\pm$ . All concentrations are in units of  $\text{mg kg}^{-1}$  except for pH and Clay ( $<2 \mu\text{m}$  particle size fraction; % by volume).

Property	Percentile, all soils					Mean and standard deviation								
	10th	25th	50th	75th	90th	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
pH	4.94	5.35	6.01	6.69	8.13	6.10 $\pm$ 0.99	5.56 $\pm$ 1.07	5.39 $\pm$ 0.56	5.87 $\pm$ 0.80	5.82 $\pm$ 0.76	6.60 $\pm$ 1.29	6.02 $\pm$ 0.91	8.02 $\pm$ 0.90	5.86 $\pm$ 0.52
Clay (%)	12.42	23.88	43.21	67.26	78.77	10.89 $\times$ 2.15	29.21 $\times$ 1.38	55.78 $\times$ 1.27	25.09 $\times$ 1.45	61.19 $\times$ 1.25	47.43 $\times$ 1.45	70.7 $\times$ 1.2	73.4 $\times$ 1.18	77.75 $\times$ 1.14
TOC	2178	3113	5489	11139	20176	2089 $\times$ 1.71	3593 $\times$ 1.77	7137 $\times$ 1.79	4134 $\times$ 1.73	8588 $\times$ 2.11	6473 $\times$ 2.13	12813 $\times$ 1.96	10503 $\times$ 1.70	23054 $\times$ 2.41
K <sub>T</sub>	1117	2482	7046	16948	26556	2505 $\times$ 2.27	2530 $\times$ 2.75	2890 $\times$ 1.98	28819 $\times$ 1.36	5861 $\times$ 2.23	13877 $\times$ 1.78	6088 $\times$ 3.35	21573 $\times$ 1.32	4516 $\times$ 1.89
Ca <sub>T</sub>	183	400	1254	5028	19873	342 $\times$ 2.14	277 $\times$ 2.75	479 $\times$ 1.97	1844 $\times$ 2.23	1165 $\times$ 2.14	3796 $\times$ 2.37	1842 $\times$ 4.33	19558 $\times$ 2.33	3071 $\times$ 2.71
Mn <sub>T</sub>	36	96	216	530	1029	38 $\times$ 2.23	60 $\times$ 3.10	173 $\times$ 2.07	170 $\times$ 1.69	339 $\times$ 2.13	457 $\times$ 1.59	721 $\times$ 2.20	558 $\times$ 1.91	1105 $\times$ 2.01
Fe <sub>T</sub>	3518	8714	20675	36643	61561	3020 $\times$ 1.78	3525 $\times$ 5.65	19759 $\times$ 1.29	11161 $\times$ 1.49	36032 $\times$ 1.29	26140 $\times$ 1.34	54828 $\times$ 1.41	36489 $\times$ 1.17	90529 $\times$ 1.33
Ni <sub>T</sub>	2	5	13	25	47	3 $\times$ 2.27	3 $\times$ 2.56	12 $\times$ 1.81	5 $\times$ 2.21	25 $\times$ 1.51	18 $\times$ 1.50	26 $\times$ 2.14	72 $\times$ 2.34	37 $\times$ 1.88
Cr <sub>T</sub>	3	5	12	20	31	3 $\times$ 1.77	4 $\times$ 2.08	12 $\times$ 1.87	6 $\times$ 1.51	21 $\times$ 1.46	17 $\times$ 1.45	25 $\times$ 1.57	21 $\times$ 1.22	29 $\times$ 1.88
Zn <sub>T</sub>	4	8	19	35	59	4 $\times$ 1.97	6 $\times$ 1.84	19 $\times$ 1.63	12 $\times$ 1.45	22 $\times$ 1.57	30 $\times$ 1.69	40 $\times$ 1.67	55 $\times$ 1.32	52 $\times$ 1.65
B <sub>M</sub>	0.001	0.001	0.081	0.310	0.829	0.003 $\times$ 11.61	0.002 $\times$ 10.66	0.027 $\times$ 11.93	0.013 $\times$ 16.09	0.064 $\times$ 6.85	0.106 $\times$ 14.64	0.133 $\times$ 7.34	2.265 $\times$ 1.87	0.044 $\times$ 9.65
Mg <sub>M</sub>	39	71	150	386	815	79 $\times$ 1.99	53 $\times$ 2.40	82 $\times$ 1.87	129 $\times$ 1.63	255 $\times$ 2.60	331 $\times$ 1.73	273 $\times$ 3.11	1414 $\times$ 1.64	335 $\times$ 2.35
K <sub>M</sub>	29	45	77	158	334	51 $\times$ 2.32	52 $\times$ 2.15	58 $\times$ 1.85	84 $\times$ 2.38	89 $\times$ 1.76	157 $\times$ 2.60	91 $\times$ 2.12	677 $\times$ 1.96	115 $\times$ 2.27
Ca <sub>M</sub>	122	251	570	1540	5104	279 $\times$ 2.08	176 $\times$ 2.97	315 $\times$ 2.28	521 $\times$ 1.93	782 $\times$ 2.25	1352 $\times$ 2.22	881 $\times$ 3.44	7050 $\times$ 2.48	1433 $\times$ 3.38
Mn <sub>M</sub>	5	18	62	130	334	17 $\times$ 3.05	18 $\times$ 5.71	37 $\times$ 3.40	37 $\times$ 2.65	62 $\times$ 3.95	137 $\times$ 1.94	111 $\times$ 3.66	83 $\times$ 4.30	55 $\times$ 3.70
Fe <sub>M</sub>	40	58	86	126	175	72 $\times$ 1.92	101 $\times$ 1.54	95 $\times$ 1.61	95 $\times$ 1.60	101 $\times$ 1.45	91 $\times$ 1.71	102 $\times$ 1.47	43 $\times$ 2.02	102 $\times$ 1.57
Cr <sub>M</sub>	0.2	0.5	1.1	2.4	3.8	0.1 $\times$ 4.62	0.4 $\times$ 4.15	1.0 $\times$ 2.69	0.8 $\times$ 1.68	2.1 $\times$ 1.87	2.6 $\times$ 1.54	2.2 $\times$ 2.13	3.0 $\times$ 1.33	1.9 $\times$ 2.57
Zn <sub>M</sub>	0.4	0.6	1.0	1.5	2.4	0.8 $\times$ 1.65	0.9 $\times$ 1.62	1.0 $\times$ 1.80	0.6 $\times$ 8.3	1.0 $\times$ 2.05	1.3 $\times$ 1.89	1.1 $\times$ 2.03	1.0 $\times$ 1.88	0.8 $\times$ 2.40

amorphous components within it. Using a library of prior measured diffractograms of pure phases (Eberl, 2003) and their Reference Intensity Ratios (Hillier, 2000), an observed pattern can be modelled as the sum of these pure components and their concentrations can be accurately determined (Omotoso et al., 2006). Given the overlapping features of some clay minerals in bulk XRPD measurements (i.e. randomly oriented milled subsamples of the <2 mm fraction), clay minerals characterised as dioctahedral smectites along with mixed-layered (interstratified) clay minerals with dioctahedral character were grouped into a general category labelled as 'dioctahedral expandable' phyllosilicates.

### 3. Statistical analyses

All statistical analyses outlined in subsequent sections were conducted using the R language and environment for statistical computing (R Core Team, 2018). Statistical test results were assessed at the usual 0.05 significance level.

#### 3.1. Cluster analysis of X-ray powder diffraction data

Prior to cluster analysis, the XRPD data were subset to the  $6 - 75^\circ 2\theta$  range before being pre-treated by alignment, binning (bin width = 5), square-root transformation and mean centering as recommended by Butler et al., 2019. This procedure was designed to stabilise sample independent variability of the signals to enhance comparison across samples, whilst reducing the overwhelming signal from strong diffractors such as quartz.

For the cluster analysis, principal component analysis [PCA, Jolliffe, 1986] was firstly applied to reduce the pre-processed XRPD dataset by projection onto a low-dimensional space (Rosset et al., 2016; Butler et al., 2019). Principal components that explained 99% of pre-processed data variability (in this case 21 principal components) were then used as the input for the fuzzy-c-means clustering algorithm (Bezdek et al., 1984; Meyer et al., 2017). Fuzzy clustering allows for classification uncertainty by considering soft boundaries between clusters so that for each sample it determines membership coefficients, defined in the [0,1] interval, to each cluster (using a cluster fuzziness hyper-parameter set to 2). The optimum number of clusters was objectively derived by applying the fuzzy-c-means algorithm to 19 iterations with cluster nodes ranging from 2 to 20, and selecting the instance with the highest partition coefficient from the clustering statistics (Bezdek et al., 1984; Rosset et al., 2016). To facilitate neater characterisation of clusters from the soil mineralogy continuum, only samples exceeding the 75% quantile of the membership coefficients in each cluster were retained for further inspection of soil properties across the clusters. This yielded a subset of the data containing 239 samples upon which all subsequent analysis was applied, allowing the investigation to focus on more distinct mineralogical groupings within the data.

#### 3.2. Compositional analysis of nutrient concentrations

Nutrient concentrations describing the relative make-up of a soil sample are inherently multivariate data. Regardless of the units of measurement, the data are constrained in that each chemical species measurement is a fraction or part of a total (e.g. total sample weight or volume) and, hence, changes in the relative abundance of one species necessarily implies changes in at least one of the others. It has been long recognised that, for example, ordinary correlation measures computed on data carrying relative information can result in spurious associations, with pairwise correlations between parts not being consistent when measured from a full composition or a subset of it (Pearson, 1897; Aitchison, 1986). Since the correlation structure of a dataset is a key piece of information in multivariate analysis, this is a fundamental drawback. Following on the seminal work by Aitchison, 1986, the mainstream approach to dealing with compositional data focuses on the

analysis of log-ratios between parts which contain the relative information. This eliminates technical issues like spurious correlations, singularity of the covariance matrix or data curvature, producing results which do not depend on the scale of the data nor the sample total. Moreover, through log-ratios the data are mapped onto the ordinary real space, which then facilitates analysis, modelling and visualisation using ordinary statistical methods on real-valued log-ratio coordinates.

A form of log-ratios, so called balances (Egozcue and Pawlowsky-Glahn, 2005), were used here to represent the total and M3 nutrient compositions ( $\text{TOC-K}_T\text{-Ca}_T\text{-Mn}_T\text{-Fe}_T\text{-Ni}_T\text{-Cu}_T\text{-Zn}_T$  and  $\text{B}_M\text{-Mg}_M\text{-K}_M\text{-Ca}_M\text{-Mn}_M\text{-Fe}_M\text{-Cu}_M\text{-Zn}_M$ , respectively, expressed in units of  $\text{mg kg}^{-1}$ ; see Section 2.2.2). This facilitated interpretability since these balances correspond to trade-offs between subsets of parts of the composition determined according to their co-dependence structure (measured in terms of proportionality by the variation matrix; see details on the construction of these balances in accompanying Supplementary Material).

##### 3.2.1. Values below the detection limit and average nutrient concentrations

The subset ( $n = 239$ ) M3 dataset included overall 5.23% cells with values below the detection limit, which were mostly concentrated in  $\text{B}_M$  (34.31%) and  $\text{Cu}_M$  (6.69%). These values were dealt with using statistical imputation to enable subsequent multivariate analysis. The log-ratio expectation-maximisation (EM) algorithm (Palarea-Albaladejo et al., 2007) implemented in the 'zCompositions' R package (Palarea-Albaladejo and Martin-Fernandez, 2015) was used for this which preserves the log-ratios between parts while accounting for the corresponding detection limit thresholds and the co-dependence structure of the observed data.

Following imputation of values below the detection limit, geometric means were used to summarise the various nutrient concentrations of each cluster (Section 3.1) in a way that is consistent with their relative scale (Aitchison, 1986; Montero-Serrano et al., 2010). It is worth noting that geometric means and associated geometric standard deviations are related to one-another by division and multiplication, as opposed to the traditional subtractions and additions of their arithmetic equivalents. Thus the 'divide on times' notation ( $\otimes$ ) is provided wherever geometric means and geometric standard deviations are reported.

##### 3.2.2. Multivariate analysis of variance

The entire nutrient compositions from the total and M3 datasets were statistically compared between the groups determined by cluster analysis of the XRPD data (Section 3.1) using multivariate analysis of variance (MANOVA) on log-ratio balances. More specifically, the non-parametric permutational MANOVA (PERMANOVA) introduced by Anderson, 2001 was applied since the ordinary assumption of multivariate normality for the standard MANOVA was not met. Results from PERMANOVA are invariant to the log-ratio balance representation chosen, as determined by a sequential binary partition into subsets of parts of the composition (Egozcue and Pawlowsky-Glahn, 2005), thus balances built according to co-dependences between parts were used as described in the Supplementary Material.

##### 3.2.3. Principal component analysis biplots

To visualise relationships between log-ratio nutrient balances and soil mineralogy, PCA biplots were produced by PCA of the pre-treated and subset XRPD data ( $n = 239$ ). Log-ratio balances of the total and M3 datasets were provided as supplementary variables to the PCA (Lê et al., 2008) so that they could subsequently be added as vectors to biplots based on the first three principal components. Mineral contributions to the principal components derived from the XRPD data were interpreted using the loadings of each component in combination with the known diffraction features of soil mineral components (ICDD, 2019).



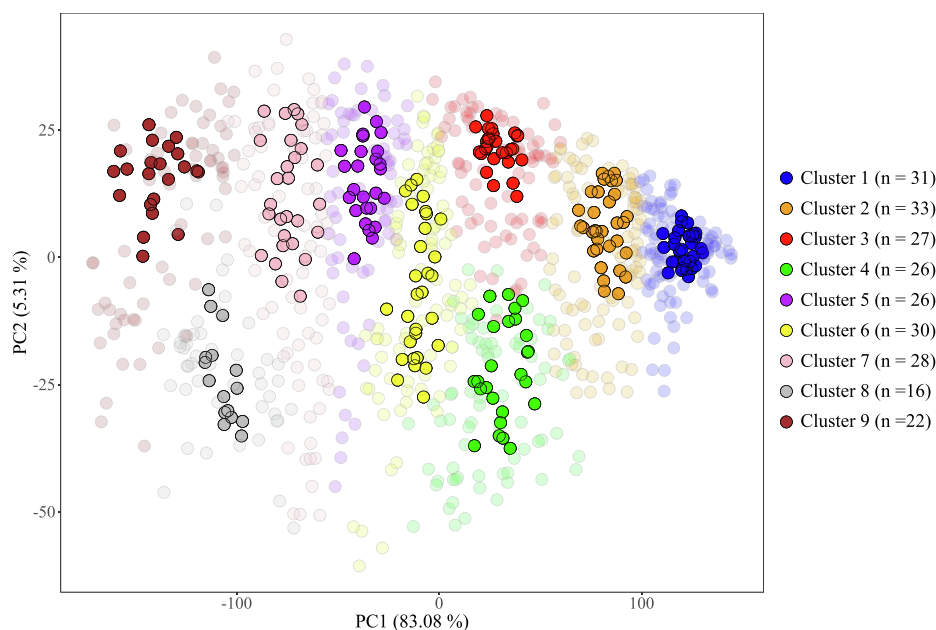


Fig. 2. Principal component scores of pre-treated X-ray power diffraction data and the resulting clusters defined by the fuzzy-c-means algorithm. Total number of samples = 935. Samples excluded from subsequent analysis are plotted as translucent symbols, whilst those retained are plotted as opaque symbols.

### 3.3. Analysis of variance of soil pH

Soil nutrient concentrations are directly and indirectly related to soil pH via the effects of changing  $H^+$  activity on adsorption–desorption and dissolution–precipitation reactions (Dixon and Schulze, 2002; Singh and Schulze, 2015). Thus measurements of pH are included within this investigation to help isolate mineralogical drivers of soil nutrient concentrations. The variation in pH between the mineralogically defined clusters (Section 3.1) was assessed using analysis of variance (ANOVA).

## 4. Results and discussion

### 4.1. Cluster analysis

The optimum number of clusters for the XRPD dataset, defined by maximisation of the partition coefficient of the fuzzy-c-means algorithm (Section 3.1), was found to be nine. A broad spread of the XRPD data when plotted in principal component space was observed, reflecting the soil mineralogy continuum (Fig. 2). Certain samples from this continuum were excluded from subsequent analysis (based on membership coefficients < 75% quantile threshold; Section 3.1), creating mineralogically discrete clusters more suitable for identifying mineral–nutrient relationships. Estimates of the quartz concentration for each cluster, derived from the quantitative mineral analysis (Sections 2.2.3 and 4.2), were used to order the clusters from 1 through to 9 by decreasing quartz concentration, and thus aid interpretation of the data. The resulting dataset comprised a total of 239 soil samples across the nine clusters, with cluster sizes ranging from 17 to 33 (Fig. 2).

The 239 samples in the nine clusters represented soils from 57 of the 60 Sentinel sites in total. The samples in each cluster display considerable spatial variation across sub-Saharan Africa (Fig. 3). The mean number of sites represented by each cluster is twelve, with mean distances between samples within each cluster being approximately 2000 – 3000 km. The only obvious exception to this is Cluster 8, which is represented by only three sites, with a mean distance between samples of 813 km. Given that the Sentinel sites cover the major agro-ecological zones of sub-Saharan Africa (Vågen et al., 2010; Hengl et al., 2017), the spatial distribution of the clustered data suggests that the mineralogically similar soils within each cluster are representative of a

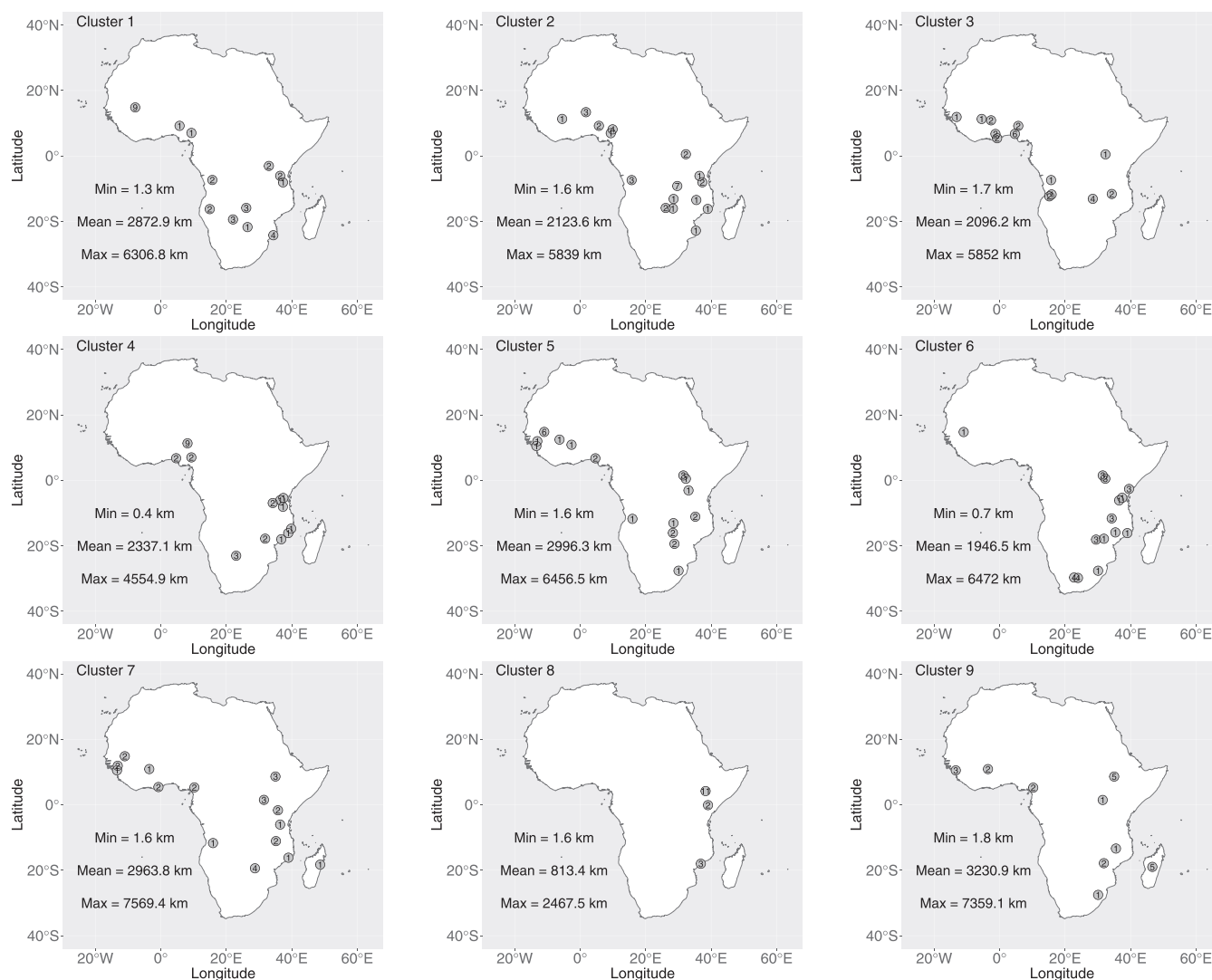
wide variety of land uses, climatic properties and ecosystems (see Table 1 in Towett et al., 2015).

Soil TOC concentration (i.e. organic matter), pH and clay particle size fraction can affect soil nutrient concentrations, and are therefore taken into account here when interpreting the results from the cluster analysis (Table 1). The average pH values for each cluster range from 5.39 (Cluster 3) to 8.02 (Cluster 8). Notably, the mean pH of Cluster 8 is close to the 90th percentile of the entire dataset (Table 1), thus the soils within this group are relatively unusual in this context. Indeed, ANOVA of pH (Section 3.3) between the clusters yields a statistically significant difference in means only when Cluster 8 is included, suggesting that pH should only be taken into account when interpreting results associated with this cluster. The average clay particle size fraction and soil TOC concentration of the nine clusters both display a general increase from Cluster 1 through to Cluster 9 (Table 1), with notable features that will be discussed in relation to soil mineralogy and nutrient concentrations below.

#### 4.1.1. Soil mineralogy and nutrient concentrations

Quantitative estimates for the mineral compositions of each cluster, derived from their mean diffractograms (Fig. 4), are expressed in terms of weight percent of soil mineral components (i.e. organic matter is not implicit in the composition; Table 2; Fig. 5). This quantitative information allows for improved interpretation of mineral–nutrient relationships from the cluster analysis.

A total of 17 minerals were identified across the nine clusters (Table 2). Quartz, phyllosilicates (divided into kaolin, expandable dioctahedral and illitic/micaceous mineral groups), feldspars (K-feldspar and plagioclase) and Fe/Al/Ti-(hydr)oxides (goethite, maghemite, ilmenite, hematite, gibbsite, magnetite and anatase) dominated the mineral compositions (Fig. 5), whilst amphibole (Clusters 6 and 8), pyroxene (Cluster 9), calcite (Cluster 8) and gypsum (Cluster 8) had more limited occurrence. Several of the nine clusters stand out based on the mineralogy of the mean diffractogram (Fig. 5; Table 2). These include the mineralogy of Cluster 1 being almost entirely quartz; Cluster 4 being by far the richest in K-feldspar minerals; Cluster 6 being particularly enriched in illitic/micaceous minerals; Cluster 8 displaying high concentrations of expandable dioctahedral phyllosilicates along with the presence of calcite and gypsum; and Cluster 9 containing high concentrations of Fe/Al/Ti-(hydr)oxides. Together these contrasting



**Fig. 3.** The spatial distribution of soil samples within each cluster. Symbols are labelled according to the number of samples at a given Sentinel site. Min, Mean and Max represent the minimum, mean and maximum distance between soil samples within the cluster, respectively.

soil mineralogies illustrate that the cluster analysis of soil XRPD data yielded nine groups of mineralogically distinct soils, with each group formed of soils from different agro-ecological environments of sub-Saharan Africa (Fig. 3).

Nutrient concentrations (total and/or M3 nutrients) also displayed substantial variation between the nine clusters (Table 1). This variation is visualised in Fig. 6 according to the deviation (in log-ratio scale) of each cluster's geometric mean from that of the entire dataset (Montero-Serrano et al., 2010). Results illustrate general nutrient deficiencies in Clusters 1–4, and enrichments in Clusters 5–9 (Fig. 6). Aside from this broad trend, the mineralogical and geochemical data can be combined to interpret far more specific mineral–nutrient relationships.

#### 4.1.2. Mineral–nutrient relationships

The mean diffractograms of Clusters 1 and 2 represent soils containing 98% and 85% quartz, respectively. The soils within these clusters are consistently deficient in almost all nutrients and micronutrients investigated (Fig. 6; Table 1) with average concentrations consistently near the 10th or 25th percentile of the entire dataset. This deficiency can be explained by the dominance of quartz and kaolin minerals in these soils (Fig. 5; Table 2). Quartz is a notoriously inert mineral that displays negligible contributions to total and extractable soil nutrients (Hardy and Cornu, 2006), whilst kaolinite is not generally considered a

significant source of plant nutrients (Dixon and Schulze, 2002). Indeed the consistent nutrient deficiencies of Clusters 1 and 2 suggest that such soils may require special management practices that account for their low nutrient status and potential susceptibility to nutrient leaching in response to fertiliser applications.

Clusters 3 and 4 have almost identical concentrations of quartz (Fig. 5; Table 2), but exhibit distinct  $K_T$  concentrations (Fig. 6). More specifically, the average  $K_T$  concentration of Cluster 3 is near the 25th percentile of the dataset, whilst that of Cluster 4 exceeds the 90th percentile (Table 1). These relatively high  $K_T$  concentrations in Cluster 4 are found to be driven by the enrichment of these soils in K-feldspar minerals (17%; Table 2), which represent a large soil  $K_T$  reservoir where present (Towett et al., 2015). The contrasting  $K_T$  concentrations between Clusters 3 and 4 despite their near identical quartz concentrations illustrates how important information can be lost when characterising soil parent material by its silica (i.e. quartz) content alone (Gray et al., 2016).

Clusters 8 and 9 display similar concentrations of both quartz and phyllosilicate mineral components (Fig. 5; Table 2), but differ greatly in M3 nutrient concentrations, with Cluster 8 showing far greater enrichment in  $B_M$ ,  $Mg_M$ ,  $K_M$  and  $Ca_M$  (Fig. 6; Table 1). This difference is driven in part by the contrasting phyllosilicate mineralogies of these clusters, with Cluster 8 dominated by expandable dioctahedral

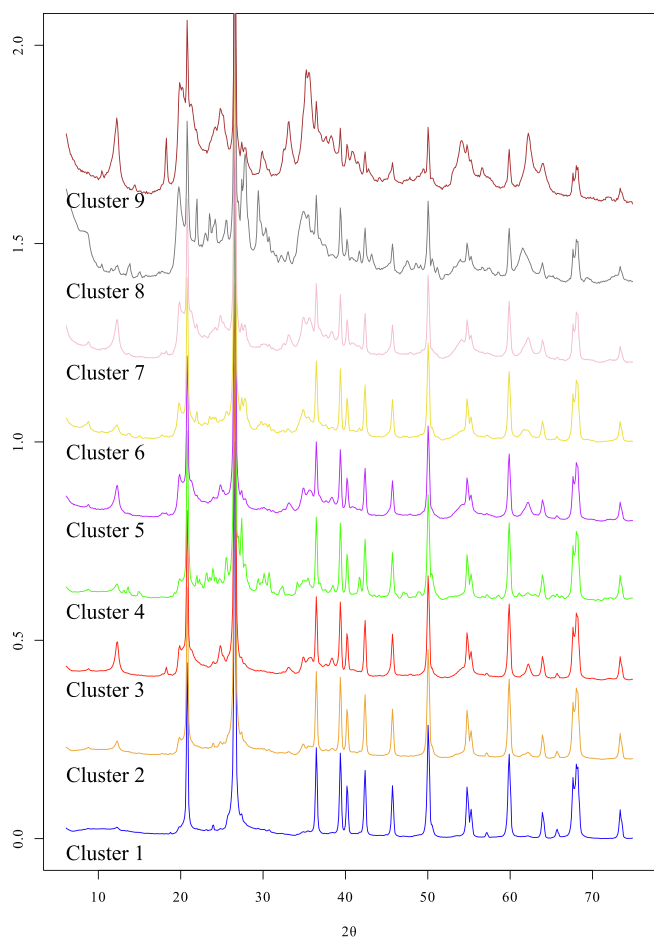


Fig. 4. The mean diffractograms for each cluster, which were subsequently used to derive quantitative estimates of each cluster's mineralogy. The y-axis has been square-root transformed to aid with comparison.

phyllosilicates (45%; Table 2) that can represent substantial sources of adsorbed cations (Velde and Barré, 2009), and Cluster 9 dominated by kaolin minerals (34%; Table 2) that offer limited cation exchange capacities in comparison (Dixon and Schulze, 2002). Further influence of mineralogy on the contrasts in nutrient concentrations between Clusters 8 and 9 relate to the presence of calcite in the soils of Cluster 8 (Table 2), resulting in an alkaline pH (Table 1) that would favour the adsorption of base cations to expandable dioctahedral phyllosilicates (Dixon and Schulze, 2002). Further to the similar concentrations of total phyllosilicate minerals in Clusters 8 and 9, their clay particle size fractions are almost identical (Cluster 8 =  $73.4 \times 1.18$ ; Cluster 9 =  $77.75 \times 1.14$  Table 1). Together the contrasting nutrient concentrations of Clusters 8 and 9 despite their similar phyllosilicate concentrations (Table 2) and particle size fractions (Table 1) illustrate how important information can be lost when summarising soil clay content as a size fraction (commonplace in soil science; Churchman, 2010) or total phyllosilicate concentration.

A useful attribute of the present dataset is the availability of both total and extractable concentrations of some elements (K, Ca, Mn, Fe, Cu and Zn), which offers a rare insight into how these properties can contrast, and how some of these contrasts are mineralogically driven. In this context, Cluster 4 stands out as having the highest mean  $K_T$  concentration of all Clusters, but this enrichment is not expressed in the  $K_M$  concentrations of these soils (Fig. 6; Table 1). This contrast between  $K_T$  and  $K_M$  in Cluster 4 illustrates how despite K-feldspar minerals being a major K reserve in these soils, it is not accessible to the M3 extraction, which is in agreement with the notorious resistance of K-feldspar to chemical extractions (Andrist-Rangel et al., 2010; Butler et al., 2018).

Table 2

Quantitative mineralogy estimated from the mean diffractogram of each cluster (Fig. 4). Kaol. = kaolin minerals; Exp. dioct. = expandable dioctahedral minerals; Ill. = illitic/micaceous minerals; K-fel. = K-feldspar minerals; Plag. = plagioclase minerals; Goet. = goethite; Magh. = maghemite; Ilim. = ilmenite; Hem. = hematite; Gibb. = gibbsite; Magn. = magnetite; Anat. = anatase; Amph. = amphibole minerals; Pyrox. = pyroxene minerals; Cal. = calcite; Gypsum. = gypsum;  $\sum$ Phyllo. = summed concentration of phyllosilicate minerals (i.e. kaolin and 2:1 phyllosilicates);  $\sum$ Fe/Al/Ti-(hydr)ox. = summed concentration (excluding values <1%) of Fe/Al/Ti-(hydr)oxide minerals.

Cluster	2:1 phyllosilicates										Feldspars				Fe/Al/Ti-(hydr) oxides				Ferromagnesian					
	Quartz	Kaol.	Exp. dioct.		Ill.	K-fel.		Plag.	Goet.	Magh.	Ilim.	Hem.	Gibb.	Magn.	Anat.	Amph.		Pyrox.	Cal.	Gyp.	$\sum$ Phyllo.	$\sum$ Fe/Al/Ti-(hydr)ox.		
			Exp. dioct.	Ill.		K-fel.	Plag.									Amph.	Pyrox.							
1	98	2	0	0	0	<1	0	0	0	0	0	0	0	0	<1	0	0	0	0	0	0	2	0	0
2	85	7	1	4	0	1	0	0	0	0	0	0	0	0	<1	0	0	0	0	0	12	1	1	9
3	66	22	2	0	3	17	4	7	0	<1	1	1	1	0	<1	0	0	0	0	0	24	9	2	2
4	64	8	2	3	9	2	0	1	1	0	0	0	0	0	<1	0	0	0	0	0	13	12	12	8
5	48	24	9	4	9	4	4	10	0	0	2	0	0	0	<1	3	0	0	0	0	30	8	20	6
6	47	9	12	9	3	7	4	7	1	0	0	0	0	0	<1	0	0	0	0	0	43	6	46	47
7	33	27	13	3	1	9	2	15	0	2	2	1	0	0	<1	3	0	0	5	1	46	6	36	36
8	17	1	45	1	0	3	2	18	4	4	5	3	4	2	<1	0	4	0	0	0	47	0	0	0
9	10	34	13	0	0	3	2	18	4	4	5	3	4	2	<1	0	4	0	0	0	47	0	0	0



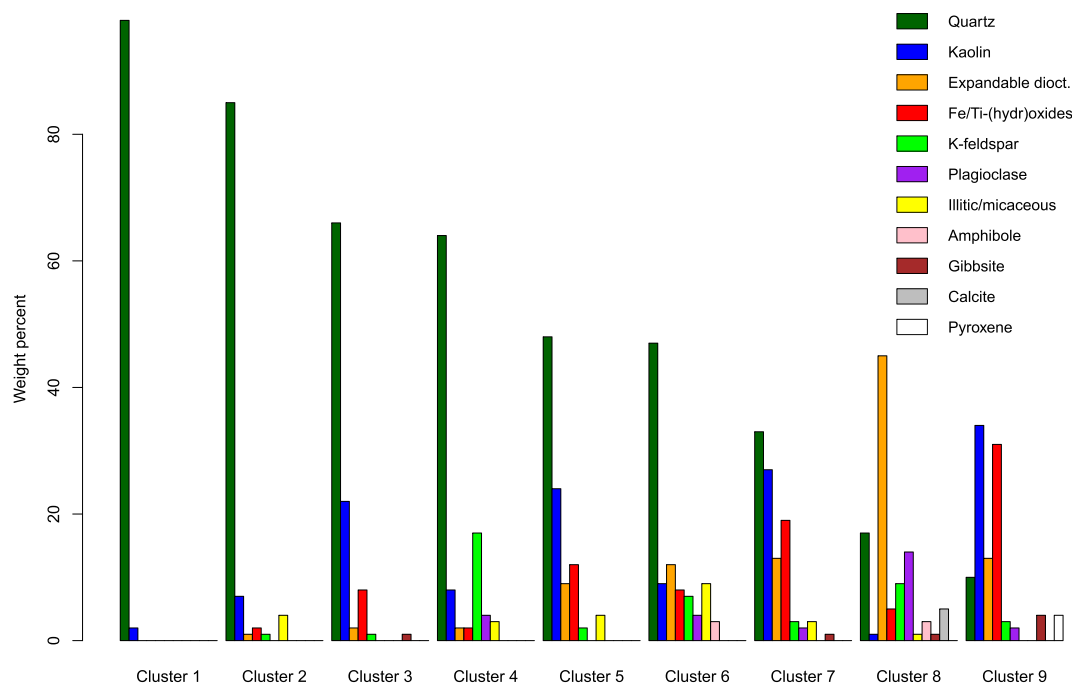


Fig. 5. Simplified mineralogy of each cluster, with Fe/Ti-(hydr)oxides grouped together. The Fe/Ti-(hydr)oxide bar therefore represents the sum of goethite, maghemite, ilmenite, hematite and magnetite. A more detailed summary is provided in Table 2.

Further distinctions between total and extractable nutrients are provided in the soils of Cluster 9, which show enrichments in  $Ca_T$ ,  $Mn_T$ ,  $Fe_T$ ,  $Cu_T$  and  $Zn_T$  that are not reflected in their M3 equivalents. Thus the total reserves of these nutrients, likely present within the high concentration of Fe/Al/Ti/Mn-(hydr)oxides (Table 2), are not generally accessible to the M3 extraction.

$Fe_T$  and  $Zn_T$  concentrations vary considerably between the nine clusters, but this is not replicated for either  $Fe_M$  or  $Zn_M$ . The only exception to this lack of variability is the  $Fe_M$  deficiency of Cluster 8, with these soils having an average  $Fe_M$  concentration approximately half that of all other clusters (Fig. 6; Table 1). This relative  $Fe_M$  deficiency of Cluster 8 (with an average near the 10th percentile of the dataset; Table 1) is explained by the alkaline pH of these soils (caused by the presence of calcite; Fig. 5) since the solubility of Fe-(hydr)oxides is particularly dependent upon pH (Lindsay and Schwab, 1982). The consistent lack of variability of  $Zn_M$  concentrations between the nine mineralogically diverse clusters suggests that neither major soil mineral constituents nor pH are key drivers of  $Zn_M$  in this case, and further research could seek to identify potentially important mineral sources of phyto-available Zn given recent concerns about Zn deficiencies in crops and humans (Alloway, 2009).

#### 4.2. Compositional data analysis

Further mineral–nutrient relationships within the dataset were explored by applying compositional data analysis to the nine clusters (Section 3.2). On a general basis, the PERMANOVA applied to nutrient balances (Sections 3.2.2 and 3.2) supports a statistically significant difference in (geometric) mean nutrient compositions between clusters ( $p < 0.0001$ ) for both total and M3 nutrients, reflecting how the contrasting soil mineralogies of the nine clusters yield differences in nutrient compositions.

##### 4.2.1. Balances

Dendrograms based on the variation matrices of the total and M3 datasets (Fig. 7; Supplementary Material) show the overall grouping structures of the parts of the respective compositions according to their co-dependence. The dendrogram of the total nutrient dataset presents

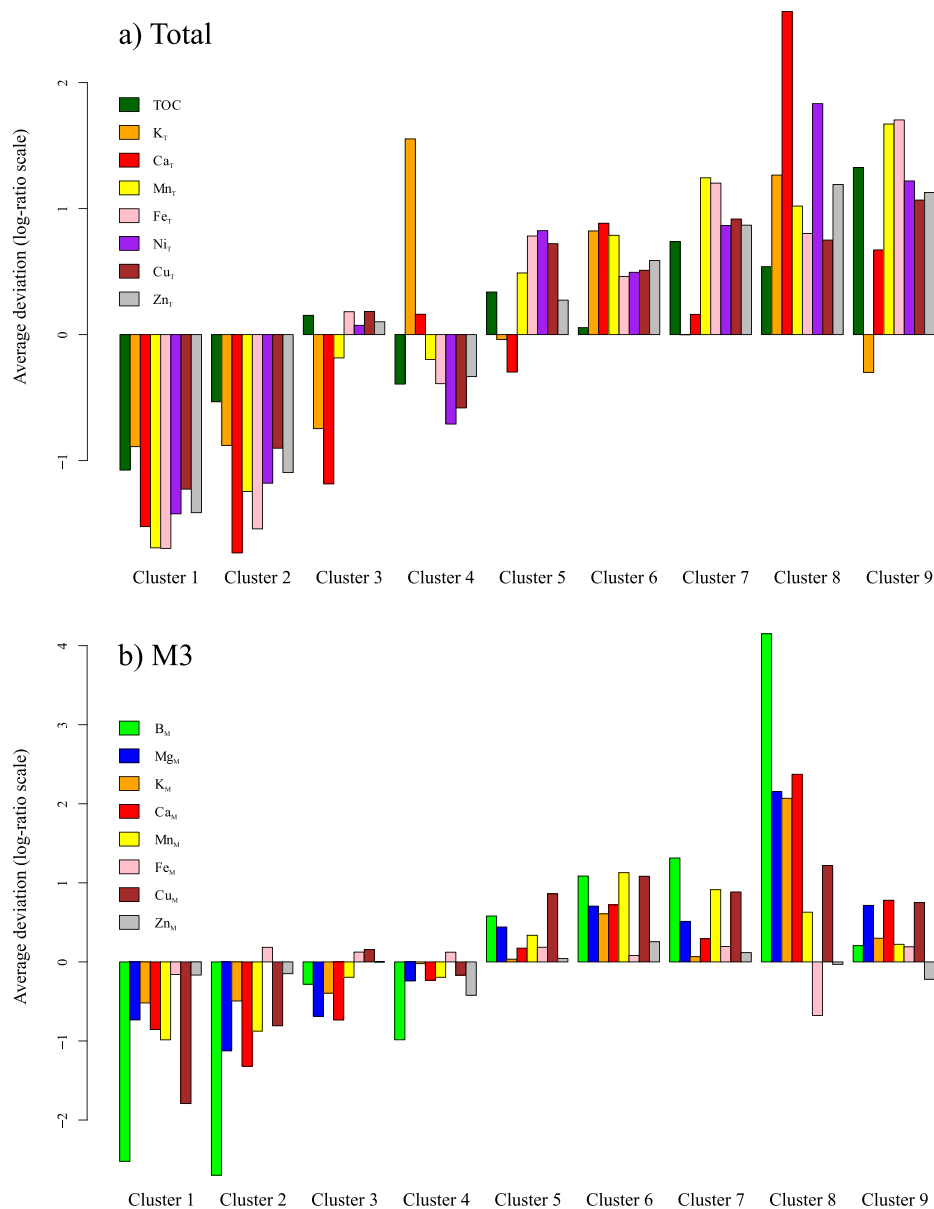
two clear groupings of co-dependent components (Fig. 7a), the first formed of  $TOC-Mn_T-Ni_T-Zn_T-Fe_T-Cu_T$ , and the second formed of  $K_T-Ca_T$ . The groupings were used to define seven balances between the components of the total nutrient dataset (Table 3). The dendrogram of the M3 nutrient dataset presents about four groups of co-dependent components (Fig. 7b). The first of these groups comprises  $B_M$  alone, which is clearly the most independent element in the M3 dataset. The remaining three groups are comprised of  $Fe_M-Zn_M$ ,  $Mg_M-K_M-Ca_M$ , and  $Mn_M-Cu_M$ . Again these groupings were used to define seven balances between the components of the M3 nutrient dataset (Table 3).

##### 4.2.2. Principal component biplots

Principal component analysis applied to the subset ( $n = 239$ ; Section 3.2.3) pre-treated XRPD data from the cluster analysis was used to represent the multivariate data in three dimensions (i.e. the first three principal components). These principal components (PCs) together explained 95.54% of total variation of XRPD data ( $PC1 = 90.13\%$ ,  $PC2 = 3.94\%$  and  $PC3 = 1.47\%$ ), and are plotted in combination with the balances in biplots provided in Fig. 8. More specifically, PC scores in Fig. 8 are represented by point data (coloured by the Cluster factor), whilst the balances are represented as vectors. It is worth noting that since the balances were supplementary variables in the PCA, the PC scores remain solely defined by the XRPD data. In general the biplots illustrate how a range of total and M3 nutrient balances correspond to the principal component scores, and therefore to soil mineralogy. Balances  $b_1^1$ ,  $b_2^1$ ,  $b_3^1$ ,  $b_4^1$  and  $b_5^1$  display the strongest correlations to the PCs, as reflected by their length and proximity to the corresponding PC reference axes presented in Figs. 8a–c.

The loadings of the three principal component dimensions (Fig. 9) represent the positive or negative contributions of the XRPD variables to the PCs. The Powder Diffraction File database (ICDD, 2019) and the mean diffractograms of the nine clusters were used to assign likely mineral components to regions of notably high or low loading values (Fig. 9) so that mineral–nutrient relationships could be interpreted from Fig. 8.

The loadings of PC1 are dominated by negative values in regions of the XRPD data that are associated with quartz peaks (Fig. 9a). Thus increasing quartz peak intensity would promote decreased PC1 values



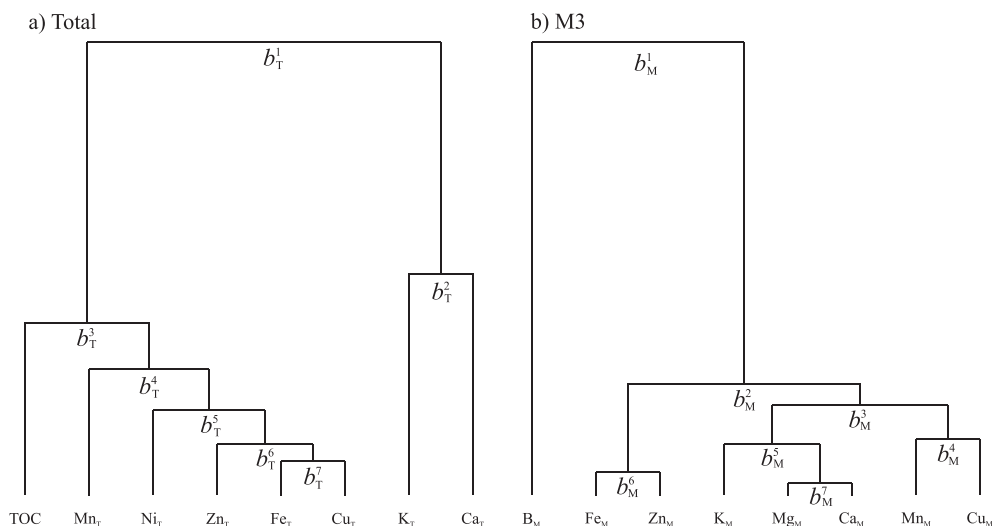
**Fig. 6.** Total (a) and M3 (b) average nutrient concentrations of each cluster expressed as deviation in log-ratio scale from the overall geometric mean of the dataset (i.e. all 935 subsoils; at zero reference line). Values below zero represent average concentrations lower than that for the entire dataset, whilst values above zero represent the opposite.

in Fig. 8. The high proportion of variability accounted for by PC1 (90.13%) is explained by the near ubiquitous presence of quartz in soil (Dixon and Schulze, 2002) along with it being a relatively strong diffractor (Butler et al., 2019). In contrast to PC1, the loadings of PC2 and PC3 represent more subtle features within the diffraction data, accounting for 3.94% and 1.47% of XRPD data variability, respectively. Despite this, the loadings of PC2 and PC3 contain various regions characterised by distinctly positive or negative values that relate to the diffraction features of several common mineral components within the dataset. Namely, positive PC2 loadings are associated with kaolin minerals and Fe/Al-(hydr)oxides, whilst negative PC2 loadings are associated with 2:1 phyllosilicates (i.e. both expandable 2:1 and micaceous 2:1 phyllosilicates), K-feldspar and plagioclase (Fig. 9b). Positive PC3 loadings are associated with K-feldspar and kaolin minerals, whilst negative PC3 loadings are associated with 2:1 phyllosilicates, plagioclase and calcite (Fig. 9c). Together these interpretations of the PC loadings provide a mineralogical meaning for the PCs presented in Fig. 8, allowing further mineral–nutrient relationships to be explored.

#### 4.2.3. Total nutrient balances

Of the seven log-ratio balances defined from the total nutrient dataset (Table 3),  $b_T^1$  and  $b_T^2$  stand out as being related to soil mineralogy based on PCA biplot analysis (Fig. 8).

The balance  $b_T^1$  accounts for 40.94% of variability in the total nutrient dataset (Table 3). This balance partitions the two most distinct groups of co-dependent variables in the total nutrient dataset (Fig. 7), representing TOC–Mn<sub>T</sub>–Fe<sub>T</sub>–Ni<sub>T</sub>–Cu<sub>T</sub>–Zn<sub>T</sub>:K<sub>T</sub>–Ca<sub>T</sub>, and correlates positively with PC2 scores of the XRPD data (Fig. 8). Positive PC2 scores (and increased  $b_T^1$  values) are driven by enrichment of the soils in Fe/Al-(hydr)oxides and kaolin minerals relative to feldspars and 2:1 phyllosilicates. Based on this interpretation of PC2 and estimated chemical compositions of these minerals, it is Fe/Al/Ti-(hydr)oxides that are likely to be the main driver of increased  $b_T^1$  values given their association with Mn, Fe, Ni, Cu and Zn (Brummer et al., 1988; Neaman et al., 2008), whilst both 2:1 phyllosilicates and feldspars represent known sources of K<sub>T</sub> and Ca<sub>T</sub> (Towett et al., 2015) that are likely to promote decreased  $b_T^1$  values. Soils rich in Fe/Al-(hydr)oxides are often also



**Fig. 7.** Groupings of the total (a) and M3 (b) nutrient concentration datasets derived from their respective variation matrices (see Tables S1 and S2 and Supplementary Material), used to define meaningful nutrient balances (Table 3).

**Table 3**

Nutrient balances for the total and M3 datasets, defined from the overall grouping structure of the variation matrix for each dataset (Fig. 7 and Supplementary Material), along with their contributions to total variance.

Balance	Components	Total variance (%)
<i>Total</i>		
$b_T^1$	TOC-Mn <sub>T</sub> -Fe <sub>T</sub> -Ni <sub>T</sub> -Cu <sub>T</sub> -Zn <sub>T</sub> -K <sub>T</sub> -Ca <sub>T</sub>	40.94
$b_T^2$	K <sub>T</sub> :Ca <sub>T</sub>	16.64
$b_T^3$	TOC:Mn <sub>T</sub> -Fe <sub>T</sub> -Ni <sub>T</sub> -Cu <sub>T</sub> -Zn <sub>T</sub>	14.42
$b_T^4$	Mn <sub>T</sub> :Fe <sub>T</sub> -Ni <sub>T</sub> -Cu <sub>T</sub> -Zn <sub>T</sub>	9.01
$b_T^5$	Ni <sub>T</sub> :Fe <sub>T</sub> -Cu <sub>T</sub> -Zn <sub>T</sub>	7.54
$b_T^6$	Zn <sub>T</sub> :Fe <sub>T</sub> -Cu <sub>T</sub>	5.79
$b_T^7$	Fe <sub>T</sub> :Cu <sub>T</sub>	5.65
<i>M3</i>		
$b_M^1$	B <sub>M</sub> :Mg <sub>M</sub> -K <sub>M</sub> -Ca <sub>M</sub> -Mn <sub>M</sub> -Fe <sub>M</sub> -Cu <sub>M</sub> -Zn <sub>M</sub>	51.75
$b_M^2$	Fe <sub>M</sub> -Zn <sub>M</sub> :Mg <sub>M</sub> -K <sub>M</sub> -Ca <sub>M</sub> -Mn <sub>M</sub> -Cu <sub>M</sub>	17.47
$b_M^3$	Mg <sub>M</sub> -K <sub>M</sub> -Ca <sub>M</sub> :Mn <sub>M</sub> -Cu <sub>M</sub>	11.48
$b_M^4$	Mn <sub>M</sub> :Cu <sub>M</sub>	7.81
$b_M^5$	K <sub>M</sub> :Mg <sub>M</sub> -Ca <sub>M</sub>	5.61
$b_M^6$	Fe <sub>M</sub> :Zn <sub>M</sub>	4.21
$b_M^7$	Mg <sub>M</sub> :Ca <sub>M</sub>	1.68

enriched in Mn-oxides (Neaman et al., 2008), which are difficult to characterise or even identify from complex XRPD data unless especially abundant. Together, Fe/Al/Ti/Mn-(hydr)oxides in soil may also act as strong adsorbents of Ni (Bruemmer et al., 1988), Cu and Zn (Violante et al., 2003), which may partly explain the co-dependence of these elements in  $b_T^1$ . Further, the inclusion of TOC in the TOC-Mn<sub>T</sub>-Fe<sub>T</sub>-Ni<sub>T</sub>-Cu<sub>T</sub>-Zn<sub>T</sub> sub-composition also aligns with the consensus that Fe/Al/Ti/Mn-oxides play an important role in the stabilisation of organic matter in soils and sediments (Lützwow et al., 2006; Wagai and Mayer, 2007; Lalonde et al., 2012). In contrast to the geochemistry of Fe/Al-(hydr)oxides, K-feldspar and plagioclase minerals do not usually represent a source of structural or adsorbed transition metal micro-nutrients, and the dominance of feldspars in defining K<sub>T</sub> and Ca<sub>T</sub> concentrations likely contributes to the two independent sub-components of  $b_T^1$ .

The balance  $b_T^2$  accounts for 17% of variability in the total nutrient dataset, and contrasts to  $b_T^1$  because it partitions the two components of the K<sub>T</sub>-Ca<sub>T</sub> sub-composition, yielding K<sub>T</sub>:Ca<sub>T</sub> (Fig. 7; Table 3). Biplots illustrate how  $b_T^2$  positively correlates with PC3 of the XRPD data (Fig. 8). Positive PC3 scores (and increased  $b_T^2$  values) are largely

associated with soils rich in K-feldspar and kaolin minerals relative to plagioclase, 2:1 phyllosilicates and calcite, whilst the reverse applies to negative PC3 scores. Pure K-feldspar minerals contain approximately 14% K by weight (Andrist-Rangel et al., 2006) and represent the main K<sub>T</sub> reservoir in most soils (Towett et al., 2015). Ca-enriched 2:1 phyllosilicates such as Ca-montmorillonite contain up to 2% Ca, plagioclase up to 14% Ca (anorthite), and calcite up to 40% Ca. Based on this information, 2:1 phyllosilicates, plagioclase and calcite could all act as a considerable source of Ca<sub>T</sub>, which is reflected in the Ca<sub>T</sub> concentration of Cluster 8 being 5 times higher than that of any other cluster (Table 1). As found for  $b_T^1$ , this analysis illustrates mineralogical controls on total nutrient concentrations, and highlights how K-feldspar, plagioclase, 2:1 phyllosilicates and calcite control  $b_T^2$ .

In summary, total nutrient compositions in the present dataset primarily relate to the concentrations of Fe/Al/Ti/Mn-(hydr)oxides, feldspars, 2:1 phyllosilicates, and calcite. TOC, Mn<sub>T</sub>, Fe<sub>T</sub>, Ni<sub>T</sub>, Cu<sub>T</sub> and Zn<sub>T</sub> form a group of co-dependent variables that reflect how Fe/Al/Ti/Mn-(hydr)oxides are the main host of these transition metal micro-nutrients in soil, and that these minerals have a tendency to promote the stabilisation of soil organic matter (Wagai and Mayer, 2007; Zhao et al., 2016). K<sub>T</sub> and Ca<sub>T</sub> concentrations are largely independent of the TOC-Mn<sub>T</sub>-Fe<sub>T</sub>-Ni<sub>T</sub>-Cu<sub>T</sub>-Zn<sub>T</sub> sub-composition and primarily relate to K-feldspar and plagioclase minerals along with contributions from 2:1 phyllosilicates and calcite.

#### 4.2.4. M3 nutrient balances

Of the seven log-ratio balances defined from the M3 nutrient dataset (Table 3),  $b_M^1$ ,  $b_M^2$  and  $b_M^3$  are found to have specific relationships to soil mineralogy based on PCA biplot analysis (Fig. 8).

B<sub>M</sub> is the most independent nutrient of the M3 dataset (Fig. 7) and is partitioned from all other M3 nutrients in  $b_M^1$  (B<sub>M</sub>:Mg<sub>M</sub>-K<sub>M</sub>-Ca<sub>M</sub>-Mn<sub>M</sub>-Fe<sub>M</sub>-Cu<sub>M</sub>-Zn<sub>M</sub>), which accounts for 51.75% of variability in the M3 dataset (Table 3). This independence is primarily controlled by the enrichment of B<sub>M</sub> in the soils of Cluster 8, which have a mean B<sub>M</sub> concentration 17 times higher than that of any other cluster (Table 1). Thus  $b_M^1$  is consistently associated with Cluster 8 in the PCA biplots (Fig. 8), driven by low quartz concentrations (positive PC1 scores), enrichment in 2:1 phyllosilicates (negative PC2 and PC3 scores) and the presence of calcite (negative PC3 scores). The notable enrichment of Cluster 8 in 2:1 phyllosilicates, especially expandable dioctahedral minerals, combined with the presence of calcite (Table 2), therefore act to create soils with high B<sub>M</sub> concentrations. In this context the presence of calcite explains the alkaline pH of the soils of Cluster 8 (pH = 8.02 ± 0.90; Table 1), which would affect B adsorption to 2:1

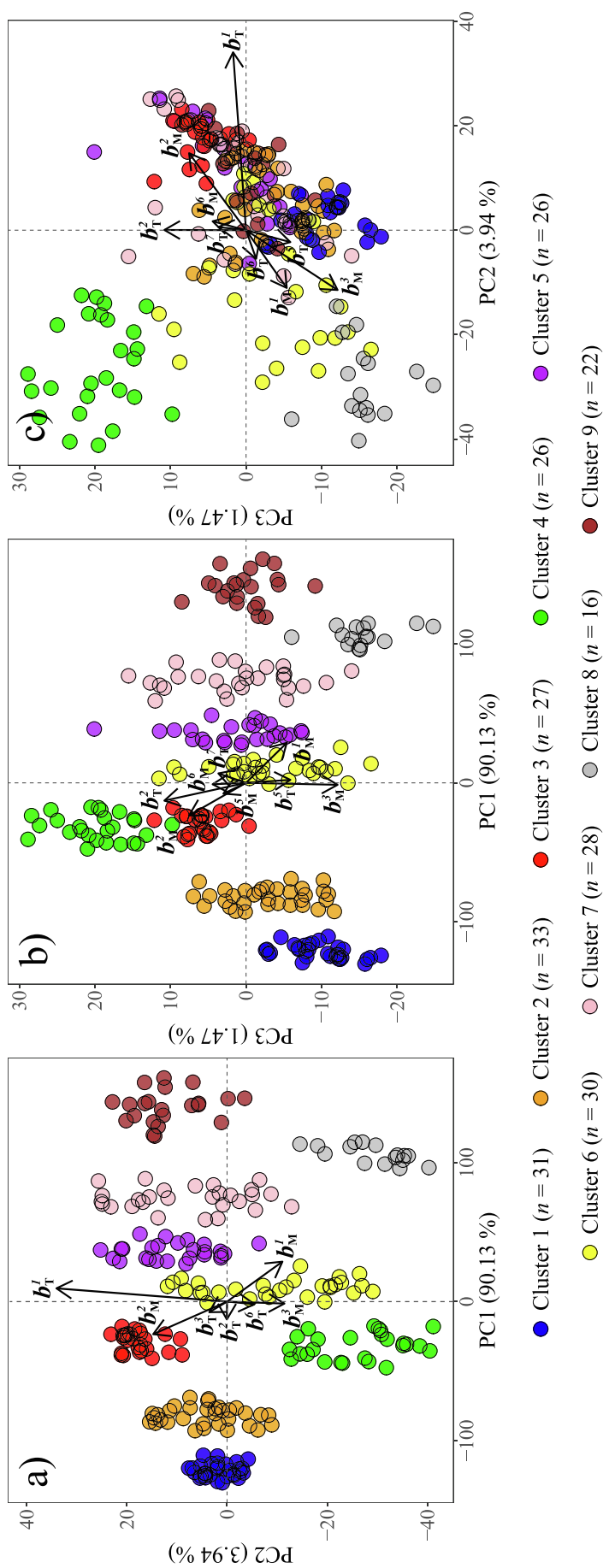
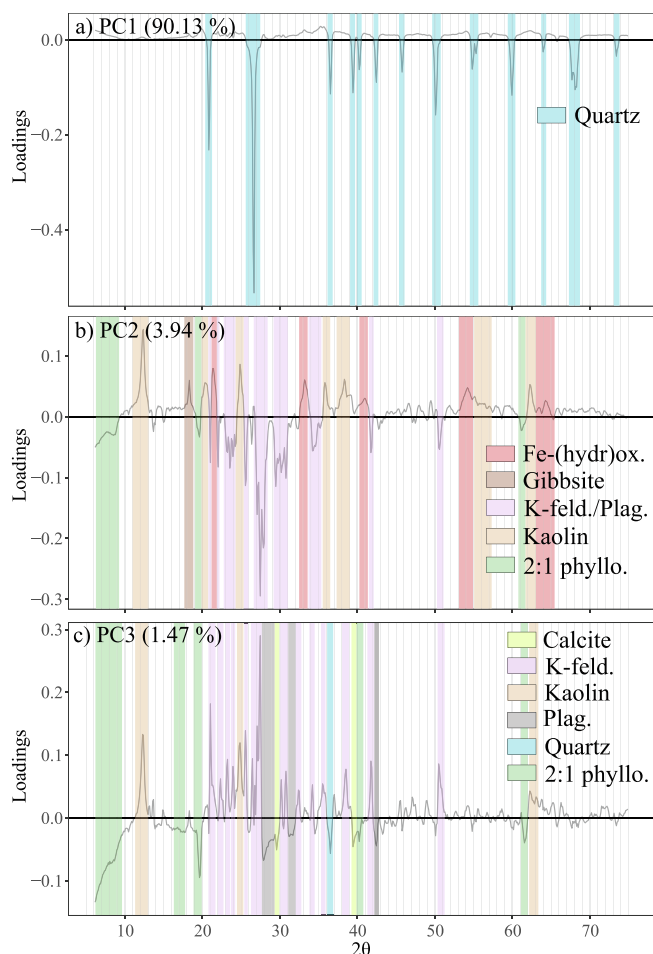
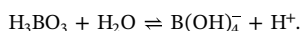


Fig. 8. Principal component analysis biplots of the pre-treated and subset ( $n = 239$ ) XRPD data (Section 3.2.3), with the total and M3 nutrient balances (Table 3) added as supplementary variables. Only balances that displayed a discernible vector length are included.



**Fig. 9.** The loadings of PC1 (a), PC2 (b) and PC3 (c) from PCA of the pre-treated and subset ( $n = 239$ ) XRPD data (Section 3.2.3). Highlighted regions represent the soil minerals associated with that specific region of XRPD variables.

phyllosilicates (Karahana et al., 2006). More specifically, aqueous  $B_M$  concentrations are driven by the equilibrium reaction of boric acid ( $H_3BO_3$ ):



Below pH 7,  $H_3BO_3$  predominates, whilst alkaline solutions favour  $B(OH)_4^-$  (Russell and Wild, 1988). In alkaline conditions (pH > 8; Karahana et al., 2006) 2:1 phyllosilicates have a high capacity for  $B(OH)_4^-$  adsorption, whilst at pH < 7 B is readily leached from the soil which can result in B deficiencies (Gupta et al., 1985). The alkaline pH of the soils of Cluster 8 would promote adsorption of B by the expandable dioctahedral minerals in the form of  $B(OH)_4^-$  (Karahana et al., 2006). All other clusters had similarly acidic pH's on average (Table 1), limiting the potential for  $B(OH)_4^-$  adsorption by phyllosilicates. Together this not only highlights the combined importance of calcite, phyllosilicate minerals and pH on soil B availability, but also suggests that the M3 extraction results in the desorption of  $B(OH)_4^-$  from expandable dioctahedral phyllosilicates.

The balance  $b_M^2$  represents the contrast of  $Fe_M-Zn_M:Mg_M-K_M-Ca_M-Mn_M-Cu_M$  that accounts for 17.47% of variability in the M3 nutrient dataset. It is worth noting that of all M3 nutrients,  $Fe_M$  and  $Zn_M$  display the least variation between the nine clusters (Fig. 6), therefore drivers of variation in  $b_M^2$  will primarily relate to changes within the  $Mg_M-K_M-Ca_M-Mn_M-Cu_M$  sub-composition.  $b_M^2$  displays positive correlations with PCs 2 and 3, and negative correlations with PC1. Increased  $b_M^2$  values are therefore promoted by soils enriched in quartz, Fe/Al-(hydr)oxides, K-feldspar, and kaolin relative to 2:1 phyllosilicates,

plagioclase and calcite. Of the minerals within this list, quartz would not represent a source of any M3 nutrients (Hardy and Cornu, 2006), whilst structural Ca and K in plagioclase and K-feldspars, respectively, would not contribute considerably to the M3 nutrient concentrations since these framework silicates are particularly resistant even to quite aggressive chemical extractants (Andrist-Rangel et al., 2010; Reimann et al., 2014; Butler et al., 2018). Of the remaining minerals driving changes in  $b_M^2$ , 2:1 phyllosilicates represent the most likely source of elements within the  $Mg_M-K_M-Ca_M-Mn_M-Cu_M$  sub-composition. It is therefore considered that  $b_M^2$  represents how 2:1 phyllosilicates are less likely to represent a source of  $Fe_M$  or  $Zn_M$  than they are of  $Mg_M$ ,  $K_M$ ,  $Ca_M$ ,  $Mn_M$  or  $Cu_M$ .

The balance  $b_M^3$  partitions the  $Mg_M-K_M-Ca_M-Mn_M-Cu_M$  sub-component of  $b_M^2$ , representing  $Mg_M-K_M-Ca_M:Mn_M-Cu_M$  and accounting for 11.48% of variability in the M3 nutrient dataset. It is found that  $b_M^3$  correlates negatively with PCs 2 and 3 (Fig. 8). Negative correlation of  $b_M^3$  with PC2 suggests soils enriched in 2:1 phyllosilicates and feldspar minerals relative to kaolin and Fe/Al-(hydr)oxides favour increased  $b_M^3$  values. Similarly the negative correlation of  $b_M^3$  with PC3 is also driven by enrichment of the soil in 2:1 phyllosilicates along with the presence of calcite. Given these mineral–nutrient relationships, soils in Clusters 6 (enriched in 2:1 phyllosilicates) and 8 (enriched in 2:1 phyllosilicates and calcite) display the highest  $b_M^3$  values (Fig. 8). Since PC2 scores are not affected by calcite and yet still relate to  $b_M^3$ , the data suggest that 2:1 phyllosilicates are more likely to represent sources of nutrients within the  $Mg_M-K_M-Ca_M$  sub-composition than the  $Mn_M-Cu_M$  sub-composition - in agreement with most literature highlighting 2:1 phyllosilicates as sources of base cations (Velde and Barré, 2009; Churchman, 2010; Singh and Schulze, 2015). Therefore, assuming that 2:1 phyllosilicates represent the main source of M3 nutrient concentrations,  $b_M^2$  and  $b_M^3$  together suggest an approximate sequence of co-dependent soil nutrients that are associated with these minerals of  $Mg_M-K_M-Ca_M > Mn_M-Cu_M > Fe_M-Zn_M$ .

In summary, whilst Fe/Al/Ti/Mn-(hydr)oxides and feldspar minerals are key drivers of total soil nutrient concentrations in the present dataset (Section 4.2.3), it is 2:1 phyllosilicates that represent the main source of all M3 extractable nutrients except for  $Fe_M$  and  $Zn_M$ . Since 2:1 phyllosilicates are therefore not likely to represent a source of  $Fe_M$  or  $Zn_M$  in soils investigated here, potential deficiencies in these elements must be accounted for by other organic or mineral soil components.

#### 4.3. Future prospects

The mineral–nutrient relationships defined here from the cluster analysis and compositional methods illustrate how mineralogy is the key driver of soil nutrient concentrations in African soils. With increasing availability of geo-referenced soil XRPD datasets (Hillier and Butler, 2018) and associated geochemical data, further data-driven research could progress towards classifying soil nutrient statuses from XRPD measurements via the use of cluster analysis or other data-driven methods. These ‘Digital Mineralogy’ approaches could aid in defining mineralogically tailored nutrient management schemes that can account for the full suite of essential plant nutrients. Doing so would help land users avoid the risks of excessive fertiliser use (Sebilio et al., 2013) and nutrient mining (Jones et al., 2013), and aid in tackling human micro-nutrient deficiencies (Alloway, 2009).

## 5. Conclusions

Cluster analysis of soil XRPD data was used to define nine mineralogically distinct clusters from the soil mineralogy continuum. Each of the nine clusters derived from XRPD data was comprised of mineralogically similar soils from different agro-ecological environments of sub-Saharan Africa. Despite this spatial variation, meaningful differences between clusters were found for both the total and M3 nutrient compositions. Quartz and kaolin minerals did not show a notable



contribution to any of the total or M3 nutrient concentrations analysed. Feldspar minerals were found to be major sources of  $K_T$  (K-feldspar) and  $Ca_T$  (plagioclase), but did not directly contribute to  $K_M$  or  $Ca_M$ . Fe/Al/Ti/Mn-(hydr)oxides were found to be associated with TOC concentrations as reflected in the co-dependence of TOC,  $Mn_T$ ,  $Fe_T$ ,  $Ni_T$ ,  $Cu_T$  and  $Zn_T$  sub-composition. Compared to the contribution of Fe/Al/Ti/Mn-(hydr)oxides to total nutrient concentrations, their importance as a source of nutrients accessible to the M3 extraction was found to be limited. All M3 nutrient concentrations except for  $Fe_M$  and  $Zn_M$  were primarily driven by 2:1 phyllosilicates, namely expandable dioctahedral and illitic/micaceous minerals. The importance of 2:1 phyllosilicates as a source of phyto-available nutrients suggests that the relative abundance of these minerals could provide a powerful indicator of the inherent fertility in African soils.

The interpreted soil mineralogy–nutrient relationships highlight how the soil mineral composition drives soil nutrient concentrations and their phyto-availability. As such, characterising soils based on their silica (e.g. basic vs acidic) and/or clay (e.g. soil texture) components alone acts to over-simplify the intricacies of mineral contributions to soil nutrients.

This is the first application of cluster analysis to soil XRPD data, from which it is evident that its use in combination with compositional data analysis methods can allow for detailed interpretation of soil mineral–nutrient relationships. Information encoded within soil XRPD data is thus inherently related to the total and/or extractable concentrations of B, Mg, K, Ca, Mn, Fe, Ni, Cu and Zn. Data-driven analysis of soil XRPD data can therefore be used to extract this information in new detail.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This work was supported by a Macaulay Development Trust Fellowship, United Kingdom, Grant No. MDT-50. J. P.-A. was partly supported by the Spanish Ministry of Science, Innovation and Universities under the project CODAMET (RTI2018-095518-B-C21, 2019–2021). Soil sampling and analysis were financially supported by the World Agroforestry Centre (ICRAF) through the Africa Soil Information Service (AfSIS) project, funded by the Bill and Melinda Gates Foundation, United States, Grant No. 51353 and Alliance for a Green Revolution in Africa (AGRA). The authors acknowledge the regional AfSIS field crews who provided the soil samples, Beatrice Oware and Bella Kauma for assistance with XRPD analysis, as well as other technical staff of the ICRAF Soil–Plant Spectral Diagnostics Laboratory for assistance with sample preparation. ICRAF's contribution was undertaken within the CGIAR Research Program on Water, Land and Ecosystems of the CGIAR Program on Water, Land and Ecosystems. We also thank Dr. Andy Kindness for reviewing the manuscript prior to submission, and the three anonymous reviewers whose constructive comments helped improve this paper. The support of the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS) is also gratefully acknowledged.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.geoderma.2020.114474>.

#### References

- Aitchison, J., 1982. The statistical analysis of compositional data. *J. Roy. Stat. Soc. Ser. B (Methodol.)* 44 (2), 139–160.
- Aitchison, J., 1986. The statistical analysis of compositional data. Monographs on statistics and applied probability. Chapman and Hall.
- Ajiboye, G., Oyetunji, C., Mesele, S., Talbot, J., 2019. The role of soil mineralogical characteristics in sustainable soil fertility management: a case study of some tropical Alfisols in Nigeria. *Commun. Soil Sci. Plant Anal.* 50 (3), 333–349.
- Alloway, B., 2009. Soil factors associated with zinc deficiency in crops and humans. *Environ. Geochem. Health* 31 (5), 537–548.
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26 (1), 32–46.
- Andrist-Rangel, Y., Hillier, S., Öborn, I., Lilly, A., Towers, W., Edwards, A.C., Paterson, E., 2010. Assessing potassium reserves in northern temperate grassland soils: a perspective based on quantitative mineralogical analysis and aqua-regia extractable potassium. *Geoderma* 158 (3–4), 303–314.
- Andrist-Rangel, Y., Simonsson, M., Öborn, I., Hillier, S., 2006. Mineralogical budgeting of potassium in soil: a basis for understanding standard measures of reserve potassium. *J. Plant Nutr. Soil Sci.* 169 (5), 605–615.
- Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* 10 (2–3), 191–203.
- Brummer, G., Gerth, J., Tiller, K., 1988. Reaction kinetics of the adsorption and desorption of nickel, zinc and cadmium by goethite. I. Adsorption and diffusion of metals. *European J. Soil Sci.* 39 (1), 37–52.
- Butler, B., Hillier, S., 2020. *powdR: Full Pattern Summation of X-Ray Powder Diffraction Data. R package version 1.2.1.* URL: <https://CRAN.R-project.org/package=powdR>.
- Butler, B.M., O'Rourke, S.M., Hillier, S., 2018. Using rule-based regression models to predict and interpret soil properties from X-ray powder diffraction data. *Geoderma* 329, 43–53.
- Butler, B.M., Sila, A.M., Shepherd, K.D., Nyambura, K.M., Gilmore, C.J., Kourkoumelis, N., Hillier, S., 2019. Pre-treatment of soil X-ray powder diffraction data for cluster analysis. *Geoderma* 337, 413–424.
- Cakmak, I., 2002. Plant nutrition research: Priorities to meet human needs for food in sustainable ways. *Plant and Soil* 247 (1), 3–24.
- Churchman, G.J., 2010. Is the geological concept of clay minerals appropriate for soil science? A literature-based and philosophical analysis. *Phys. Chem. Earth, Parts A/B/C* 35 (15–18), 927–940.
- Cornu, S., Chevalier, M., Hardy, M., Bourenne, H., Josiere, O., Pernes, M., Jolivet, C., Boulonne, L., Arrouays, D., 2009. X-ray diffraction determination of minerals carrying trace elements in soil: Application to the French Soil Quality Monitoring Network. *Commun. Soil Sci. Plant Anal.* 40, 1138–1147.
- Dixon, J.B., Schulze, D.G., 2002. Soil mineralogy with environmental applications. SSSA Book Series. 7. Soil Science Society of America, Madison, WI.
- Eberl, D., 2003. User guide to RockJock - A program for determining quantitative mineralogy from X-ray diffraction data. US Geological Survey Tech. rep.
- Egozcue, J.J., Pawłowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37 (7), 795–828.
- Gray, J.M., Bishop, T.F., Wilford, J.R., 2016. Lithology and soil relationships for soil modelling and mapping. *Catena* 147, 429–440.
- Gupta, U.C., Jame, Y., Campbell, C., Leyshon, A., Nicholaichuk, W., 1985. Boron toxicity and deficiency: A review. *Can. J. Soil Sci.* 65 (3), 381–409.
- Hardy, M., Cornu, S., 2006. Location of natural trace elements in silty soils using particle-size fractionation. *Geoderma* 133 (3–4), 295–308.
- Harris, D., Horwath, W.R., Van Kessel, C., 2001. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. *Soil Sci. Soc. Am. J.* 65 (6), 1853–1856.
- Hengl, T., Leenaars, J.G., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., Wheeler, I., Kwabena, N.A., 2017. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr. Cycl. Agroecosyst.* 109 (1), 77–102.
- Hillier, S., 2000. Accurate quantitative analysis of clay and other minerals in sandstones by XRD: comparison of a Rietveld and a reference intensity ratio (RIR) method and the importance of sample preparation. *Clay Minerals* 35 (1), 291–302.
- Hillier, S., Butler, B.M., 2018. New XRD data-based approaches to soil mineralogy. *Spectroscopy* 33 (7), 34–36.
- ICDD, 2019. PDF-4+ 2019 (Database). International Center for Diffraction Data, Newtown Square, PA, USA.
- Jenny, H., 1994. Factors of Soil Formation: A System of Quantitative Pedology. McGraw Hill.
- Jolliffe, I., 1986. Principal Component Analysis. Springer Series in Statistics. Springer, New York.
- Jones, D.L., Cross, P., Withers, P.J., DeLuca, T.H., Robinson, D.A., Quilliam, R.S., Harris, I.M., Chadwick, D.R., Edwards-Jones, G., 2013. Nutrient stripping: the global disparity between food security and soil nutrient stocks. *J. Appl. Ecol.* 50 (4), 851–862.
- Karahan, S., Yurdakoc, M., Seki, Y., Yurdakoc, K., 2006. Removal of boron from aqueous solution by clays and modified clays. *J. Colloid Interface Sci.* 293 (1), 36–42.
- Keskinen, R., Nyambura, M., Heikkinen, J., Sila, A., Euroala, M., Towett, E., Shepherd, K., Esala, M., 2019. Readily available concentrations of selected micronutrients and harmful metals in soils of Sub-Saharan Africa. *Geoderma* 347, 203–209.
- Lalonde, K., Mucci, A., Ouellet, A., Gélinas, Y., 2012. Preservation of organic matter in sediments promoted by iron. *Nature* 483 (7388), 198.
- Lê, S., Josse, J., Husson, F., 2008. FactoMineR: a package for multivariate analysis. *J. Stat. Softw.* 25 (1), 1–18.
- Lindsay, W., Schwab, A., 1982. The chemistry of iron in soils and its availability to plants.

- J. Plant Nutr. 5 (4–7), 821–840.
- Lützow, M.V., Kögel-Knabner, I., Ekschmitt, K., Matzner, E., Guggenberger, G., Marschner, B., Flessa, H., 2006. Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions—a review. *European J. Soil Sci.* 57 (4), 426–445.
- Mehlich, A., 1984. Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. *Commun. Soil Sci. Plant Anal.* 15 (12), 1409–1416.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2017. e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. URL: <https://cran.r-project.org/package=e1071>.
- Montero-Serrano, J.C., Palarea-Albaladejo, J., Martín-Fernández, J.A., Martínez-Santana, M., Gutiérrez-Martín, J.V., 2010. Sedimentary chemofacies characterization by means of multivariate analysis. *Sed. Geol.* 228 (3–4), 218–228.
- Neaman, A., Martínez, C.E., Trolard, F., Bourrié, G., 2008. Trace element associations with Fe- and Mn-oxides in soil nodules: Comparison of selective dissolution with electron probe microanalysis. *Appl. Geochem.* 23 (4), 778–782.
- Omoso, O., McCarty, D.K., Hillier, S., Kleeberg, R., 2006. Some successful approaches to quantitative mineral analysis as revealed by the 3rd Reynolds Cup contest. *Clays Clay Miner.* 54 (6), 748–760.
- Omuto, C.T., Vargas, R., 2018. Soil nutrient loss assessment in Malawi. *FAO Tech. rep.*
- Palarea-Albaladejo, J., Martín-Fernández, J., 2015. zCompositions – R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics Intelligent Laboratory Syst.* 143, 85–96.
- Palarea-Albaladejo, J., Martín-Fernández, J.A., Gómez-García, J., 2007. A parametric approach for dealing with compositional rounded zeros. *Math. Geol.* 39 (7), 625–645.
- Pearson, K., 1897. Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. R. Soc. of London* 60 (359–367), 489–498.
- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P., 2014. Chemistry of Europe's agricultural soils, part A. Schweizerbart'sche Verlagsbuchhandlung.
- Russell, E., Wild, A., 1988. Russell's soil conditions and plant growth. Longman Scientific & Technical.
- Schjoerring, J.K., Cakmak, I., White, P.J., 2019. Plant nutrition and soil fertility: synergies for acquiring global green growth and sustainable development. *Plant and Soil* 434, 1–6.
- Schulze, D.G., 1989. An introduction to soil mineralogy. In: Amonette, J.E., Bleam, W.F., Schulze, D.G., Dixon, J.B. (Eds.), *Soil mineralogy with environmental applications*. Soil Science Society of America, Madison, WI, pp. 1–35.
- Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G., Mariotti, A., 2013. Long-term fate of nitrate fertilizer in agricultural soils. *Proc. Nat. Acad. Sci.* 110 (45), 18185–18189.
- Singh, B., Schulze, D.E., 2015. Soil minerals and plant nutrition. *Nature Educ. Knowledge* 6 (1), 1.
- Smaling, E., 1994. The balance may look fine when there is nothing you can mine: nutrient stocks and flows in west African soils. In: *Use of phosphate rock for sustainable agriculture in West Africa. Proceedings of a seminar on the use of local mineral resources for sustainable agriculture in West Africa, held at IFDC-Africa*. pp. 21–23.
- Sutton, M.A., Bleeker, A., Howard, C., Erisman, J., Abrol, Y., Bekunda, M., Datta, A., Davidson, E., de Vries, W., Oenema, O., et al., 2013. Our nutrient world: the challenge to produce more food and energy with less pollution. *Tech. rep.* Centre for Ecology & Hydrology.
- Towett, E.K., Shepherd, K.D., Cadisch, G., 2013. Quantification of total element concentrations in soils using total X-ray fluorescence spectroscopy (TXRF). *Sci. Total Environ.* 463–464, 374–388.
- Towett, E.K., Shepherd, K.D., Tondoh, J.E., Winowiecki, L.A., Lulseged, T., Nyambura, M., Sila, A., Vågen, T.-G., Cadisch, G., 2015. Total elemental composition of soils in Sub-Saharan Africa and relationship with soil forming factors. *Geoderma Regional* 5, 157–168.
- Vågen, T., Winowiecki, L., Desta, L.T., Tondoh, J., 2015. The land degradation surveillance framework (LDSF) – Field Guide. World Agroforestry Centre, Nairobi, Kenya. *Tech. rep.*
- Van Ittersum, M.K., Van Bussel, L.G.J., Wolf, J., Grassini, P., Van Wart, J., Guilpart, N., Claessens, L., de Groot, H., Wiebe, K., Mason-D'Croz, D., Yang, H., Boogaard, H., van Oort, P.A.J., van Loon, M.P., Saito, K., Adimo, O., Adjei-Nsiah, S., Agali, A., Bala, A., Chikowo, R., Kaizzi, K., Kouressy, M., Makoi, J.H.J.R., Ouattara, K., Tesfaye, K., Cassman, K.G., 2016. Can sub-Saharan Africa feed itself? *Proc. Natl. Acad. Sci.* 113, 14964–14969.
- Vanlauwe, B., Wendt, J., Giller, K.E., Corbeels, M., Gerard, B., Nolte, C., 2014. A fourth principle is required to define conservation agriculture in sub-Saharan Africa: the appropriate use of fertilizer to enhance crop productivity. *Field Crops Res.* 155, 10–13.
- Velde, P., Barré, P., 2009. *Soils, Plants and Clay Minerals: Mineral and Biologic Interactions*. Springer Science & Business Media.
- Violante, A., Ricciardella, M., Pigna, M., 2003. Adsorption of heavy metals on mixed Fe-Al oxides in the absence or presence of organic ligands. *Water, Air, Soil Pollut.* 145 (1–4), 289–306.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Dematté, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B.G., Bartholomew, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodsky, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Selltito, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. *Earth Sci. Rev.* 155 (January), 198–230.
- Vitousek, P.M., Naylor, R., Crews, T., David, M.B., Drinkwater, L., Holland, E., Johnes, P., Katzenberger, J., Martinelli, L., Matson, P., et al., 2009. Nutrient imbalances in agricultural development. *Science* 324 (5934), 1519–1520.
- Voortman, R.L., Sonneveld, B.G., Keyzer, M.A., 2003. African land ecology: opportunities and constraints for agricultural development. *Ambio: J. Human Environ.* 32 (5), 367–374.
- Vågen, T.G., Shepherd, K.D., Walsh, M.G., Winowiecki, L.A., Tamene Desta, L., Tondoh, J. E., 2010. AfSIS Technical Specifications – Soil Health Surveillance. *Tech. rep.*, CIAT (the AfSIS project), Nairobi, Kenya.
- Wagai, R., Mayer, L.M., 2007. Sorptive stabilization of organic matter in soils by hydrous iron oxides. *Geochim. Cosmochim. Acta* 71 (1), 25–35.
- White, P., Brown, P., 2010. Plant nutrition for sustainable development and global health. *Annals Botany* 105 (7), 1073–1080.
- Zhang, G., Germaine, J.T., Martin, R.T., Whittle, A.J., 2003. A simple sample-mounting method for random powder X-ray diffraction. *Clays Clay Miner.* 51 (2), 218–225.
- Zhao, Q., Poulson, S.R., Obrist, D., Sumaila, S., Dynes, J.J., McBeth, J.M., Yang, Y., 2016. Iron-bound organic carbon in forest soils: Quantification and characterization. *Biogeosciences* 13 (16), 4777–4788.
- Ziadi, N., Tran, T.S., 2007. Mehlich 3-extractable elements. *Soil Sampling Methods Anal.* 81–88.