Geoderma 402 (2021) 115193

Contents lists available at ScienceDirect

Geoderma

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

A comparison between three legacy soil maps of Zambia at national scale: The spatial patterns of legend units and their relation to soil properties

I. Mukumbuta [a], L.M. Chabala [a], S. Sichinga [b], C. Miti [a], R.M. Lark [c,*]

[a] University of Zambia, School of Agricultural Sciences, Great East Road Campus, Lusaka, Zambia
[b] Soil Survey Unit, Zambia Agriculture Research Institute, Mt Makulu Central Research Station, Chilanga, Lusaka, Zambia
[c] University of Nottingham, School of Biosciences and Future Food Research Beacon, Sutton Bonington, Nottinghamshire LE12 5RD, UK

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:
Soil surveys
Legacy soil information
Zambia

ABSTRACT

We examined three soil maps of Zambia, two published at scales of 1:1 million – the Exploratory Soil Map of Zambia (ESMZ) and the Vegetation–Soil Map produced by Trapnell and colleagues in 1947 – and one at 1:3 million, the Soil Atlas of Africa (SAA). We estimated components of variance for measurements of clay, sand and organic carbon content and bulk density of the soil across the country using models which included different mean values for soil map units as random effects. For all but organic carbon content there was significant variation accounted for by differences between legend units for two of the maps, ESMZ with legend units based on the FAO-Unesco and SAA with legend units based on the World Reference Base respectively. This was despite their small cartographic scale. For the Vegetation–Soil Map, we examined differences between broad soil physiographic units. These did not account for significant variation in the soil properties. There were clear similarities between the soil physiographic units of the Soil–Vegetation Map and broader physiographic units into which the legend units of the ESMZ are grouped. The spatial pattern of soil units of the SAA was the most spatially heterogeneous, as measured by the sum of indicator variograms, despite being at the smallest published scale. It was apparent that some of the soil variation within the largest physiographic unit of the Soil–Vegetation Map, the Plateau Soils, as expressed by the map units of the SAA was significantly associated with the different vegetation units mapped in 1947. These studies show how quantitative assessment of legacy soil information may help us understand its potential and limitations.

1. Introduction

Soil information is essential for planning agricultural research, development and extension, and to support policy makers, farmers and environmental managers who are concerned with food security and environmental protection, in particular the protection of soil resources. While many countries have limited capacity to undertake soil surveys, they often have a legacy of soil maps. Soil survey was conducted during the Twentieth Century in much of the Global South by Colonial Authorities; and after independence, by local soil surveyors often assisted by technical staff from the previous Colonial power (the United Kingdom in the case of Zambia), or by international agencies (Young, 2017). The potential value of this legacy of soil surveys has been recognized by initiatives to preserve soil maps and memoirs and to make them available online (e.g. through the International Soil Reference and Information Centre (ISRIC) library and map collectionhttps://www.isric.org/explore/library, and the World Soil Survey Archive and Catalogue (WOSSAC),https://www.wossac.com/).

The *Vegetation–Soil map of Northern Rhodesia* produced by C.G. Trapnell and colleagues (Trapnell et al., 1947) followed more than a decade of innovative field investigations across the country. Note that 'Northern Rhodesia' was the name of Zambia from 1911 – 1964 when the country was a British protectorate. We use the name 'Zambia' in this paper to refer to the country unless we are quoting the title of a publication in which 'Northern Rhodesia' is used. The significance of Trapnell's work has been widely recognized in the development of soil survey (Young, 2017), understanding of farming systems (Allan, 1965; Moore and Vaughan, 1994) and in the broader story of the development of science and agricultural expertise in Africa (Worthington, 1938; Hodge, 2007; Tilley, 2011). Soil survey in Zambia continued after 1947, and after independence. Soil maps were produced at a range of scales, including small scale national maps. The most recent national map of

---

* Corresponding author.
E-mail address: murray.lark@nottingham.ac.uk (R.M. Lark).

https://doi.org/10.1016/j.geoderma.2021.115193
Received 7 January 2021; Received in revised form 23 April 2021; Accepted 30 April 2021
Available online 21 May 2021

Zambia (Soil Survey Section, 1991), the Exploratory Soil Map of Zambia, was produced at the same scale, 1:1 million, as the vegetation–soil map of Trapnell et al. (1947). The soils of Zambia are also mapped at 1:3 million as part of the Soil Atlas of Africa produced under the auspices of the Joint Research Centre of the European Union (Jones et al., 2013).

It is recognized that legacy soil maps are not just of historical interest, but may provide a useful basis to address contemporary problems and to design new surveys. Systematic approaches to the evaluation of legacy soil surveys have been developed (Forbes et al., 1982) and applied (e.g. Cambule et al., 2015;Rasaei et al., 2020). The objective of this study was to examine how small-scale legacy surveys might be examined and compared. For example, there are differences between the three maps of Zambia which were mentioned in the previous paragraph in terms of legend units and cartographic scale. The survey methodologies differ, not least because of the availability of airphotography, remote sensor data and modern methods of cartography to the producers of the more recent maps. Trapnell and colleagues had access to airphotography for only a small part of the country. To evaluate and compare the maps we might ask how effectively they each partition the landscape of Zambia in terms of accounting for the variation of soil properties. If a map of soils represents the variation of soil properties in a region effectively then we may expect that a substantial proportion of the variation of key soil properties can be modelled statistically as between-map unit variation rather than variation within the map units.

In this study we use a database of measurements of soil properties across Zambia to compare the vegetation–soil map units of Trapnell et al. (1947), the map units of the Exploratory Soil Map of Zambia (Soil Survey Section, 1991) and those parts of the Soil Atlas of Africa which fall within Zambia. We focussed on properties which influence, *inter alia*, the water-holding capacity of the soil: particle size distribution, soil organic carbon content and bulk density. We then compare the three maps by examining spatial statistics of the map units as observed on a half-degree national grid and on a finer grid in a part of the country which encompasses substantial soil variation.

2. Methods

2.1. The surveys

2.1.1. Vegetation–Soil Map of Northern Rhodesia, Trapnell et al. (1947)

The Vegetation–Soil Map published by Trapnell et al. (1947), was based primarily on the notes of the field reconnaissance traverses conducted by them from 1932 to 1944. Some of their observations were published (Trapnell and Clothier, 1937; Trapnell, 1943), the former including a provisional map of legend units based on soil and vegetation types. The traverse records have since been published (Smith and Trapnell, 2001). Whilst the Vegetation–Soil map was based primarily on these traverses, it also reflects information from forestry surveys in western Zambia and additional surveys conducted at larger scales in parts of the country by staff of the Agriculture Department and Forestry Branch. Full details of these additional sources are presented in the survey memoir (Trapnell et al., 1947; Smith and Trapnell, 2001).

As an ecologist, Trapnell structured his traverse records primarily around the major vegetation types observed in the field, supplemented with information obtained by interviewing local people. Early in his field work Trapnell recognized that vegetation cover could provide a basis for predicting the conditions of the underlying soil, and a set of soil classes for use in vegetation–soil survey of Zambia was proposed by him for discussion at the Zanzibar meeting of East African Soil Chemists in 1934, Trapnell (1935). This classification developed over time as Trapnell covered more of the country, and is present in its essentials in the legend units of the 1947 map. That said, the soil units are not described in the manner of a conventional soil survey. Due to staffing retrenchments in Zambia, planned soil analyses were not conducted so Trapnell et al. (1947) do not present any soil chemistry data, neither are soil texture descriptions based on mechanical analysis. Furthermore, Trapnell's description of soils is not, in general, based on profile descriptions and horizonization, and so cannot be correlated directly with classes in other soil classifications. Trapnell's soil units are listed in Table 3. These were described primarily in terms of physiography, and for some (notably the Lower Valley soils), it was recognized that the classes were broad and encompass substantial soil variation.

The vegetation–soil map units are organized primarily in terms of vegetation classes. The highest-level classes are broadly structural (Evergreen and semi-decidous types, *Brachystegia-Isoberlinia* woodlands, Other deciduous woodlands and forests, Deciduous thickets, High grass-woodlands or Chipya, Tree-Grassland and Scrub-Grassland, and lastly Grasslands). Within these classes, units were identified primarily in terms of species composition or regional occurrence. For example, within the *Brachystegia-Isoberlinia* woodlands, a distinction is made between 'Northern *Brachystegia* woodlands', 'Northern *Brachystegia*–*Isoberlinia globiflora* woodland' and 'Eastern *Brachystegia–Isoberlinia globiflora* woodland'. The map units for the survey then combine these vegetation classes with the soil units so, for example, unit **E1** comprises 'Eastern *Brachystegia–Isoberlinia globiflora* woodland on Escarpment Hill soils', whereas unit **P6** comprises 'Eastern *Brachystegia–Isoberlinia globiflora* woodland on Plateau soils'. Note that vegetation unit names are given as in the original map legend. According to the table of synonymy for botanical names published by Smith and Trapnell (2001), *Isoberlinia globiflora* and *I. paniculata* are now known as *Julbernardia globiflora* and *J. paniculata* respectively. The genus *Isoberlinia* is extant, however, so what Trapnell et al. (1947) denoted as *Isoberlinia* woodland would have contained species which now belong to both genera.

The Vegetation–Soil Map of Trapnell et al. (1947) was published at a scale of 1:1 million. It was based on extensive field work, but only limited use of air photography. For this reason the map units are shown in solid colour only where Trapnell and colleagues regarded them as supported by field observations, and with hachuring where their delineation was regarded as 'inferential'. Over parts of the country (approximately 17% of the half-degree grid nodes on the map) there is no unit delineated at all.

2.1.2. Exploratory Soil Map of Zambia Soil Survey Section (1991)

The Exploratory soil map of Zambia (ESMZ) was completed in 1991 (Soil Survey Section, 1991; Soil Survey Unit, 2011). The purpose of ESMZ was to make available a broad overview of soil types and their distribution across Zambia. The map was based on the output of previous work by the Soil Survey Unit of Zambia which included extensive field work, aerial photo interpretation and soil analysis. The soil survey team made use of available soil survey reports either published or in draft form as part of the process towards map production, in particular they used maps from a programme of surveys at Province level (e.g. Ting-Tiang, 1987). Soils were classified according to the 3rd draft of the revised Legend of the Soil Map of the World (FAO, 1985) generally up to the sub-unit level.

The map legend of ESMZ comprises units based on physiography and dominant soil type (Chileshe, 1988). The physiographic units were based on the Geomorphic Legend of Dalal-Clayton et al. (1985) along with the dominant slope. For example, the code **Pd** denoted dissected plateau with slopes in the range 5–16%, **Pu** denotes undulating plateau with slopes 0–5%. **Vd** denotes footslopes and dissected upper valleys of the rift valley, slopes 5–16%.

The second component of the legend was the soil code. This is a number, which, in combination with the geomorphic unit, denotes a dominant soil subunit, in most cases a major soil unit that occupied at least 30% of the map unit. Whilst inclusions were noted in such a unit, none of these occupied 30% of the map unit. So, for example, unit **Pu1** comprises ortho-rhodic Ferralsols. Soil Complexes were reported when there were two soil types each occupying at least 30% of the soil unit and there was no characteristic spatial relationship between the two soils while Soil Associations were reported when a spatial relationship existed between the soil types in the map unit.

2.1.3. *Soil Atlas of Africa Jones et al. (2013)*

The Soil Atlas of Africa (SAA) was produced to give an overview of the soil resources of the continent. It presents map units named according to the classes of the World Reference Base (2006), Reference Soil Groups (RSG) with a prefix qualifier at a published scale of 1:3 million. Jones et al. (2013) and Dewitte et al. (2013) describe the process by which the SAA was produced. The underlying source of information is the Harmonized World Soil Database (FAO et al., 2012), which, for Zambia, Malawi and much of western Africa, drew on the FAO-74 Digital Soil Map of the World (FAO-Unesco, 1974). This raster database was converted to a vector database of soil polygons, which were processed to correct artefacts and errors identified in the identification of soil classes. Dewitte et al. (2013) show that, in Zambia, extensive polygon corrections were made across the country, and note that this was done with reference to a draft national soil map published by SADCC (1991) at a scale of 1:2 million. The SADCC map of Zambia was generalized to the scale of 1:2 million from the ESMZ sheets at 1:1 million, (Mashuta Kalebe, *pers. comm.* 2020), and so, while the SAA soil atlas of Africa is at a smaller cartographic scale than ESMZ, and makes use of some other information (FAO-74), it is not independent of it.

The polygons of the SAA are available in digital form from the European Soil Data Centre, ESDAC (Panagos Panagos et al., 2012), and this format was used to extract the delineated Reference Soil Group at locations of interest in Zambia.

2.2. *Data on soil properties*

There are relatively few point data available in Zambia on the soil properties of interest. As part of a project on the impacts of conservation agriculture on soil and groundwater, a collection was made of such data from legacy sources in the country. These sources include soil surveys where analytical data were provided (Cheatle, 1980; Dalal-Clayton, 1974, 1980; Lee, 1968), a previous study synthesizing such data from survey reports (Maclean, 1970), and a set of detailed profile descriptions associated with a meeting on soil classification which took place in Zambia in the early 1980s (Woode, 1985). In addition we extracted data for Zambia from the World Soil Information System (WoSIS), described by Batjes et al. (2017) as it stood in 2016 (Batjes et al., 2016). This last set comprises data collected from 1977 to 2000, most entries have dates in the early 1990s. In this study we examined the data in this collection on soil bulk density, clay content, sand content (defined according to the USDA classification of particle size) and organic carbon content (SOC). Reported data on organic matter content (SOM) were converted to organic carbon on the assumption that $\mathrm{SOC} = \mathrm{SOM}/1.72$, following Landon (1991). There were 443 entries in the database, collected from 81 profiles. Some of these profiles were isolated observations with known locations (e.g. observations provided by Woode, 1985), but most of them belonged to clusters of observations from known locations (e.g. sets of profile descriptions from a single large-scale survey, or from a particular farm or experimental station). The distribution of 52 unique locations, each corresponding to a single profile or cluster of profiles, is shown in Fig. 1.

2.3. *Sampling the maps*

We sampled the three soil maps to identify the delineated legend units at locations of interest.

First, the map sheets of Trapnell et al. (1947) were scanned in raster format and georeferenced. The geocoded locations with soil data (see Section 2.2) were then displayed on the map and the delineated map units were extracted for each by visual inspection. We did not distinguish between sites where the map indicates a particular vegetation–soil unit as ‘inferred’ with hachuring, or where the map unit colour was solid (near a traverse line or in an area where the original interpretation was supported by airphotography). The map units from the ESMZ and the legend units for the SAA were extracted directly for these points of interest from the digital shape files of these respective maps.

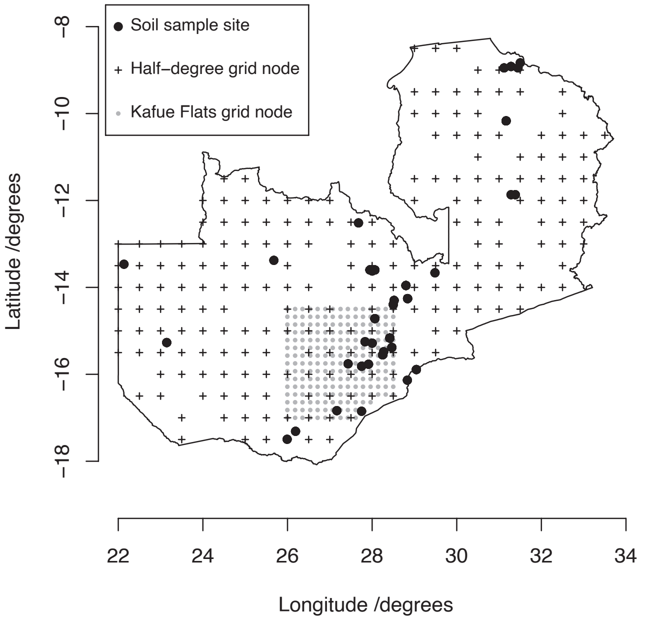

Fig. 1. Map of Zambia showing the distribution of locations with soil data (solid black symbols), locations on the half-degree national grid which are not on blank spaces in the map of Trapnell et al. (1947) (crosses) and locations on the Kafue Flats grid (solid grey symbols).

We wished to compare the spatial pattern of the legend units of the three surveys, and their mutual correspondences. An approach based on the sum of indicator variograms (SIV), which is estimated from point observations on the maps (the legend units delineated at point locations) is described below (Section 2.1.2). This, in effect, examines the cumulative distribution function of map unit inter-boundary spacings, as resolved by the sampling, and so gives insight into the scale-dependent pattern of the map units. A SIV represents the spatial pattern of sampled classes at a range of scales limited at the fine end (high spatial frequency) by the sampling interval. When sampling a map this lower limit could be made arbitrarily small. However, we preferred to sample on a relatively coarse grid to examine variation at national scale, without the assumption of stationarity in the short-range variation over widely contrasting environments, and to examine short-range variation in a limited subregion of particular interest.

We therefore sampled the mapsheets at national scale by identifying the map unit delineated at the intersection of lines of latitude and longitude at half-degree intervals. There were 213 such grid nodes within Zambia, and the legend units were identified there for all three maps. Note that some of these grid nodes were on blank areas in the map of Trapnell et al. (1947). The grid nodes where a legend unit from the survey of Trapnell et al. (1947) could be identified are shown in Fig. 1. Inspection of the maps suggested that sampling at half-degree intervals was adequate to capture the broad national pattern of soil variation attributable to the north–south climatic trend, and environmental differences between broad geological and geomorphological domains — the valleys, plateau, wetlands and Kalahari coversands.

To examine and compare the mapped spatial variation at shorter-scaled we focussed on a smaller region. We extracted map units for all three maps from the nodes of a grid of 15 rows and columns at latitude -17 to -14.5 degrees south and longitude 26–28.5 degrees north (a 2.5×2.5 degree square of interval $0.16\dot{6}$, three times the resolution of the half-degree national grid). The sampled area is centred on the Kafue Flats in Southern, Central and Lusaka provinces. The region is of particular pedological interest because it includes Upper Valley soils, formed where rejuvenation of the plateau land surface by adjustments of the drainage to successive changes in baselevel, which, along with the presence of the edge of the Kalahari sands in the west and the Kafue alluvial deposits, generates variation in the topographic and parent material factors of soil formation. Of the 225 grid nodes a total of 189 were useable, with others lying either in Lake Kariba or at sites which are blank on Trapnell's map. Again, the soil map units were extracted manually after projection onto the scanned and georeferenced version of Trapnell's map, and from the shapefiles for SAA and for ESMZ.

2.4. Data analysis

2.4.1. Linear mixed models for soil properties: nested samples and their analysis

We used a linear mixed model to analyse the variation in soil properties, with a component attributable to differences between soil map units. For this reason a different model was fitted for each soil property and each of the sets of soil map units under consideration.

Because of the relatively limited number of locations with measurements of soil properties (52) we did not attempt to analyse the soil data with respect to the basic legend units — the Vegetation–Soil units of Trapnell et al. (1947), the physiographic-soil units of the ESMZ of Zambia or the WRB sub-units with prefix qualifier from SAA. Rather, we considered the soil units in the case of Trapnell et al. (1947), that is to say the soil categories which are combined with vegetation units to define the map units. There were eight such soil units represented in the sample of soil physical data, although data on bulk density were not available for one of them (see Table 3). In the case of ESMZ we aggregated the legend units from the map in two ways, first, by the FAO-Unesco soil units (without qualifiers), ten of these units were represented in the data set (8 with measurements on bulk density). Note that four of the legend units were complexes, but the dominant soil class in each of these units was indicated and this one was specified. We also considered the physiographic units which contribute to the ESMZ legend units, of which 5 were represented in the sample of soil physical data (Table 3). In the case of the SAA we considered the Representative Soil Groups of WRB (IUSS Working Group WRB, 2006), of which 8 were represented in the soil physical data set (7 for data on bulk density).

The data that we use in this study have a complex structure. Any one measurement is taken from a particular depth interval in the soil profile, and in almost all cases there is more than one measurement from a particular profile. Furthermore, soil profiles are typically clustered in groups from a particular location. For example, soil profiles collected in a particular large-scale survey of a farm or from a single experimental station. These locations can be allocated to soil map units. These data were not selected according to a probability design, although the locations are distributed across the country (Fig. 1), and represent 7 or 8 (depending on the property) of Trapnell's 10 soil units (Table 3). The analysis of the data, however, cannot require the condition that the observations are independent, or conditionally independent, which could be ensured only by an appropriate sampling design. We therefore used a linear mixed model with an appropriate set of random effects to characterize the nested levels of dependence induced by the structure of the data.

We model the data, as described below, in terms of random contributions from nested sources of variation: differences among the map units delineated at the sample location, differences among locations within the soil map units, differences among the profiles within the locations and finally a residual term, the variation between measurements within a profile. These sources of variation were treated as random effects within a linear mixed model. This accounts for correlations (e.g. among observations within the same profile, or at the same location), and the fact that the observations are not, therefore, independent of each other. These are the random effects in the model.

In addition we considered, for each soil property, the possibility that there are systematic differences between shallow and deeper soil observations. Because the depth intervals in use vary between profiles and studies, we divided all observations into the topsoil or subsoil category according to the following rule. A soil sample is defined as 'topsoil' if its upper depth limit is at zero or if the upper depth limit is below the surface but the lower depth limit is shallower than 20 cm. Otherwise a soil sample is classified as subsoil. We therefore considered models with two alternative fixed effects structures. The expected value of the soil property is either a constant mean, or it is a topsoil or subsoil mean value.

Prior to fitting a linear mixed model, as described below, we undertook exploratory analysis of the data, and of residuals from a simple model with depth effect included. We examined histograms of the residuals, and summary statistics, including the octile skewness of Brys et al. (2003). On the basis of the summary statistics we elected to analyse bulk density on the original units of measurement (g cm^{-3}), sand content and clay content, which are proportions, on a logit scale $\log(S/(1-S))$ and $\log(C/(1-C))$ respectively where S and C are the sand and clay proportions respectively) and soil organic carbon content on a log scale.

Let the fixed effects be represented by a $n \times m$ design matrix $\mathbf{X}$. If the fixed effect is just a constant mean then $m = 1$ and the elements of $\mathbf{X}$ are all 1. If the fixed effects are topsoil vs subsoil then $m = 2$ and all elements in the first column of $\mathbf{X}$ are equal to 1 and the elements in the second column are equal to zero (for rows corresponding to a topsoil observation) or 1 (for rows corresponding to a subsoil observation). The linear mixed model for an observation of variable z can then be written as

$$z_{ijkl} = \mathbf{x}_{ijkl}^{\mathrm{T}}\boldsymbol{\beta} + M_l + S_{kl} + P_{jkl} + \varepsilon_{ijkl} \tag{1}$$

where $\mathbf{x}_{ijkl}$ is a vector equal to the row of the overall design matrix $\mathbf{X}$ which corresponds to the i^{th} observation in the j^{th} profile at the k^{th} location within the l^{th} map unit. The superscript T denotes the transpose of a matrix. The vector $\boldsymbol{\beta}$ contains the fixed effects coefficients. In the event of a model with a constant mean the only fixed effect $\boldsymbol{\beta}$ is a singleton, equal to the mean. In the event of a model with depth category as the fixed effect, as described above, the first element in $\boldsymbol{\beta}$ is the mean value for the topsoil and the second element is the difference between the mean for the subsoil and the mean for the topsoil. The remaining terms in the equation are random effects, for example, the variation of map unit means about the mean value provided by the fixed effects model is described by the random variable M where

$$M \sim \mathscr{N}(0, \sigma_M) \tag{2}$$

and $\mathscr{N}(\mu, \sigma)$ denotes a normal random variable of mean μ and standard deviation σ. The random effects parameters of the linear mixed model are therefore the within-profile, between-profile within-location, between-location within-map unit and between-map unit variances, $\sigma_\varepsilon^2, \sigma_P^2$, σ_S^2 and σ_M^2 respectively. Under an alternative model, in which there is no variation attributable to differences between map units, the term M_l is dropped from the model and σ_S^2 becomes a between-location variance component (rather than between-location within map unit).

The parameters of the random effects in the model are estimated from data by residual maximum likelihood (REML) following Patterson and Thompson (1971). This is a likelihood calculated on generalized contrasts of the observations which filter out the unknown fixed effects and so reduce the bias which is well-known in ordinary maximum likelihood estimates of variance parameters.

Under the model in Eq. (1), a set of data in vector $\mathbf{z}$ have a multivariate normal distribution with mean vector given by $\mathbf{X}\boldsymbol{\beta}$ where $\mathbf{X}$ is the design matrix introduced above and β are the random effects coefficients. The covariance matrix for $\mathbf{z}$, $\mathbf{V}$, can be written as

$$\mathbf{V} = \sigma_M^2\mathbf{U}_M\mathbf{U}_M^{\mathrm{T}} + \sigma_S^2\mathbf{U}_S\mathbf{U}_S^{\mathrm{T}} + \sigma_P^2\mathbf{U}_P\mathbf{U}_P^{\mathrm{T}} + \sigma_\varepsilon^2\mathbf{I}_n, \tag{3}$$

where $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. The $n \times n_M$ matrix $\mathbf{U}_M$ is a design matrix for the between-map units random effect, where n_M indicates the number of map units. If the i^{th} observation corresponds to the j^{th} map unit, then $\mathbf{U}_M[i,j] = 1$, and all other entries in the ith row of $\mathbf{U}_M$ are zero. The matrix $\mathbf{U}_M$ therefore shows how the observations are distributed among the map units. The matrices $\mathbf{U}_S$ and $\mathbf{U}_P$ similarly show how the observations are distributed between the locations and profiles respectively.

For a set of data and design matrices for the fixed and random effects it is then possible, for any proposed set of valid values for the unknown variance components to compute the log of the residual likelihood, which is maximized with respect to a set of variance components, $\boldsymbol{\sigma} = \{\sigma_M^2 \sigma_S^2, \sigma_P^2, \sigma_\varepsilon^2\}$, to find REML estimates of the variance components from our observations, $\mathbf{z}$, is given by

$$\ell_{\mathrm{R}}(\boldsymbol{\psi}|\mathbf{z}) = -\frac{1}{2}\left(\ln\mathbf{V} + \ln\left|\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right| + \mathbf{z}^{\mathrm{T}}\mathbf{P}\mathbf{z}\right), \tag{4}$$

where $\mathbf{P}$ is

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}. \tag{5}$$

In this study we used the optim function in the R platform (R Core Team, 2020) to find values of the variance parameters which maximize $\ell_{\mathrm{R}}(\boldsymbol{\psi}|\mathbf{z})$. This uses the simplex algorithm of Nelder and Mead (1965).

Once REML estimates of the variance parameters have been obtained, then generalized least squares estimates of the fixed effects coefficients can be computed as

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{V}}^{-1}\mathbf{z}, \tag{6}$$

where $\widehat{\mathbf{V}}$ denotes the estimated covariance matrix obtained by inserting the REML estimates of the variance parameters into Eq. (3) above. The covariance matrix for the errors in $\widehat{\boldsymbol{\beta}}$, denoted by $\mathbf{C}$, can be obtained as

$$\mathbf{C} = \left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}. \tag{7}$$

For each soil variable under consideration, and for each random effects model (different soil map units) we evaluated the evidence for a depth effect by examining the 95% confidence interval for the fixed effect coefficient which represents the difference between subsoil and topsoil values. If this interval excluded zero, then the fixed effects model with different mean values for subsoil and topsoil was considered for all further inference. Otherwise the model with a single constant mean was used.

In this study we are interested in evidence that differences between soil map units account for variation of soil properties. This evidence is provided by the estimated between-map unit variance, $\widehat{\sigma_M^2}$. We may test the null hypothesis that the between-map unit variance is zero by the log-likelihood ratio statistic L, obtained by

$$L = 2\{\ell_{\mathrm{R}}(\boldsymbol{\psi}|\mathbf{z}) - \ell_{\mathrm{R}}(\boldsymbol{\psi}_{-M}|\mathbf{z})\}, \tag{8}$$

where $\ell_{\mathrm{R}}(\boldsymbol{\psi}_{-M}|\mathbf{z})$denotes the maximized residual likelihood for a model which differs from the one in Eq. (1) only in that it lacks the between-map unit random effect. Under the null hypothesis represented by this second model L is distributed as a mixture of χ^2 random variables with 0 and 1 degrees of freedom as described by Verbeke and Molenberghs (2003). A p-value to evaluate the evidence against the null hypothesis was computed on this basis.

In cases where there was evidence to reject the null model, with no between-map unit random effect we then quantified the between map-unit difference by the intra-unit correlation, r_{MU}, which was calculated as

$$r_{\mathrm{MU}} = \frac{\widehat{\sigma^2}_M}{\widehat{\sigma^2}_M + \widehat{\sigma^2}_S}, \tag{9}$$

where $\widehat{\sigma^2}_M$ and $\widehat{\sigma^2}_S$, are, respectively, the REML estimates of the between map-unit and between-location within-map unit variance components. Note that the remaining random effects in the model are not included because these are all nested within locations and so map unit differences could not account for them.

2.4.2. Comparing the spatial distribution of legend units at sample locations

The data on legend units at the half-degree grid and on the Kafue Flats grid were examined. We cross-tabulated the observed Trapnell et al. (1947) soil units with the ESMZ physiographic units and the Representative Soil Groups from SAA on each grid. For any table a cell shows the number of observations which belonged to the legend units represented by the corresponding row and column of the table. We did not attempt to examine the FAO-Unesco soil units for the ESMZ because these were numerous, and many of the legend units were identified as complexes or associations. We note that, for both grids, the number of ESMZ physiographic units was similar to the number of Trapnell soil units or Representative Soil Groups from the SAA (see Tables 4a,4b,5,6a,6b,6c,7).

To compare the spatial pattern of each set of legend units we estimated, for each grid, the corresponding sum of indicator variograms. The sum of indicator variograms for a particular set of map legend units, $I, \zeta_I(\boldsymbol{h})$, is the probability that two locations, separated by the 'lag' vector $\mathbf{h}$, would be found to belong to different units of that legend. If we denote the unit in legend I to be found at location $\mathbf{x}$ by $K_I(\mathbf{x})$ then

$$\zeta_I(\mathbf{h}) = P[K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})]. \tag{10}$$

The function $\zeta_I(\mathbf{h})$ is so-called because it is equal to the sum of the indicator variograms (Journel, 1983) for each of the units in legend I. The sum of indicator variograms has been used to characterise the spatial distribution of soil classes (Goovaerts and Webster, 1994; Lark and Beckett, 1998). If the estimated values of $\zeta_I(\mathbf{h})$ are plotted against the scalar value of $\mathbf{h}$, the value is expected to increase with distance up to a maximum which depends on the relative frequency of the units in the legend. The sum of indicator variograms will increase with distance more rapidly, approaching a maximum value at a shorter spatial range for a fine-scale pattern of units, with short spacing between the boundaries than it would for a legend with a coarser-scale pattern in which the units are more generalized and so the distance between boundaries is characteristically longer. We are not aware of previous studies to compare different classifications or legends with respect to the sum of indicator variograms. For this purpose we propose the joint function $\zeta_{I,J}(\mathbf{h})$ which is the probability that two locations separated by lag vector $\mathbf{h}$ will be found to correspond to different units on each of two legends, I and J:

$$\zeta_{I,J}(\mathbf{h}) = P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\}, \{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\}]. \tag{11}$$

From the rules for conditional probability we may rewrite this as:

$$\begin{aligned}\zeta_{I,J}(\mathbf{h}) = P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\}|\{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\}]P[K_J(\mathbf{x}) \\ \neq K_J(\mathbf{x}+\mathbf{h})] = P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\}|\{K_J(\mathbf{x}) \\ \neq K_J(\mathbf{x}+\mathbf{h})\}]\zeta_J(\boldsymbol{h}).\end{aligned} \tag{12}$$

Now consider two possible scenarios which we denote A and B.

A. The two sets of legend units are entirely independent of each other, that is to say knowing that for some particular $\mathbf{x}$, $\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\}$ tells us nothing about whether $\{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\}$. In that case

$$\begin{aligned}P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\}|\{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\}] = P[\{K_I(\mathbf{x}) \\ \neq K_I(\mathbf{x}+\mathbf{h})\}] = \zeta_I(\boldsymbol{h}),\end{aligned} \tag{13}$$

and so, from Eq. (12), we may write

$$\zeta_{I,J}(\mathbf{h}) = \zeta_I(\mathbf{h})\zeta_J(\mathbf{h}). \tag{14}$$

B. The units of legend I are nested perfectly within the units of legend

J. By this we mean that knowing the unit in legend I at some location $\mathbf{x}$ allows us to state the unit in legend J (but not necessarily *vice versa*). In this case it follows that

$$\{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\} \Rightarrow \{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\},$$

so

$$P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\} | \{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\}] = 1,$$

and so, from Eq. (12), we may write

$$\zeta_{I,J}(\mathbf{h}) = \zeta_J(\mathbf{h}). \tag{15}$$

Furthermore, it can be shown that $\zeta_I(\mathbf{h}) > \zeta_J(\mathbf{h}) = \zeta_{I,J}(\mathbf{h})$

As a special case of our second condition, if two legends are actually equivalent, by which we mean that each unit in legend I corresponds exactly one and only one unit in legend J, and *vice versa*, then

$$\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\} \Leftrightarrow \{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\},$$

and so

$$\zeta_{I,J}(\mathbf{h}) = \zeta_J(\mathbf{h}) = \zeta_I(\mathbf{h}). \tag{16}$$

It is clear from Eq. (12) that $\zeta_J(\mathbf{h})$ is an upper bound for $\zeta_{I,J}(\mathbf{h})$. The value $\zeta_{I,J}(\mathbf{h}) < \zeta_I(\mathbf{h})\zeta_J(\mathbf{h})$ is not, in general, a lower bound because one can see from the same equation that $\zeta_{I,J}(\mathbf{h}) < \zeta_I(\mathbf{h})\zeta_J(\mathbf{h})$ if and only if

$$P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\} | \{K_J(\mathbf{x}) \neq K_J(\mathbf{x}+\mathbf{h})\}] \langle P[\{K_I(\mathbf{x}) \neq K_I(\mathbf{x}+\mathbf{h})\}]. \tag{17}$$

This condition might be met, for example, if one legend is for soil units and the other for vegetation, and vegetation is more spatially heterogeneous where the soil is more spatially uniform, and *vice versa*. In cases, such as those that interest us, where one legend is expected to be a more or less generalized version of the other we would expect that $\zeta_I(\mathbf{h}) > \zeta_J(\mathbf{h}) \geqslant \zeta_{I,J}(\mathbf{h}) > \zeta_I(\mathbf{h})\zeta_J(\mathbf{h})$. These bounds on $\zeta_{I,J}(\mathbf{h})$ help its interpretation. If two legends represent quite independent partitions of space into map units then $\zeta_{I,J}(\mathbf{h})$ is at its lower bound. If two legends are equivalent, or one is nested in the other, then $\zeta_{I,J}(\mathbf{h})$ is at its upper bound. In cases where the legends are not mutually independent, but one approximates to a generalized form of the other, $\zeta_{I,J}(\mathbf{h})$ will lie inbetween these two bounds.

We first examined the behaviour of these functions with simulated patterns of legend units. We generated a spatial pattern of legend units on nodes of the half-degree grid across Zambia by simulating, at the nodes, a standard gaussian random variable, Z_1 with a Matérn spatial correlation function (Stein, 1999) with a smoothness parameter $\nu = 2$. We then allocated points on the grid nodes to classes defined by intervals on the variable Z_1. One such set of five legend units, for example, were generated by dividing the simulated values, z_1 at their quintiles. Estimates of $\zeta_I(\mathbf{h})$, $\zeta_J(\mathbf{h})$ and $\zeta_{I,J}(\mathbf{h})$ for different simulated patterns are shown in Fig. 2, which also shows the values of $\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})$. In Fig. 2a two patterns of legend units were simulated by dividing two independent realizations of the spatial random function at their quintiles. The patterns are therefore independent, so note that $\zeta_{I,J}(\mathbf{h})$ is at the lower bound $\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})$. In Fig. 2b a pattern of classes was generated at quantiles 0.2 and 0.6 of the simulated values (legend J) and another nested set of classes (legend I at the quintiles of the same simulated values. The classes of legend I are therefore perfectly nested within the classes of legend J and so the values of $\zeta_{I,J}(\mathbf{h})$ are at the upper bound ($\zeta_J(\mathbf{h})$). In the cases shown in Figs. 2c and 2d the two sets of legend units, I and J, are approximately nested, in that they are defined by more or less different ranges of values of the same realization of the gaussian random variable. Note that the values of $\zeta_{I,J}(\mathbf{h})$ sit between the bounds in these cases. In both cases the classes in legend I were generated by dividing a realization of Z_1 at its quintiles. The classes in legend J were generated by dividing the same realization into five intervals at quantiles $(0.22, 0.42, 0.62, 0.82$ (Fig. 2c) or $(0.3, 0.5, 0.7, 0.9)$ (Fig. 2d). In the first of these two cases the overlap between the two sets of legend classes is greatest, and $\zeta_{I,J}(\mathbf{h})$ is closest to the upper boundary.

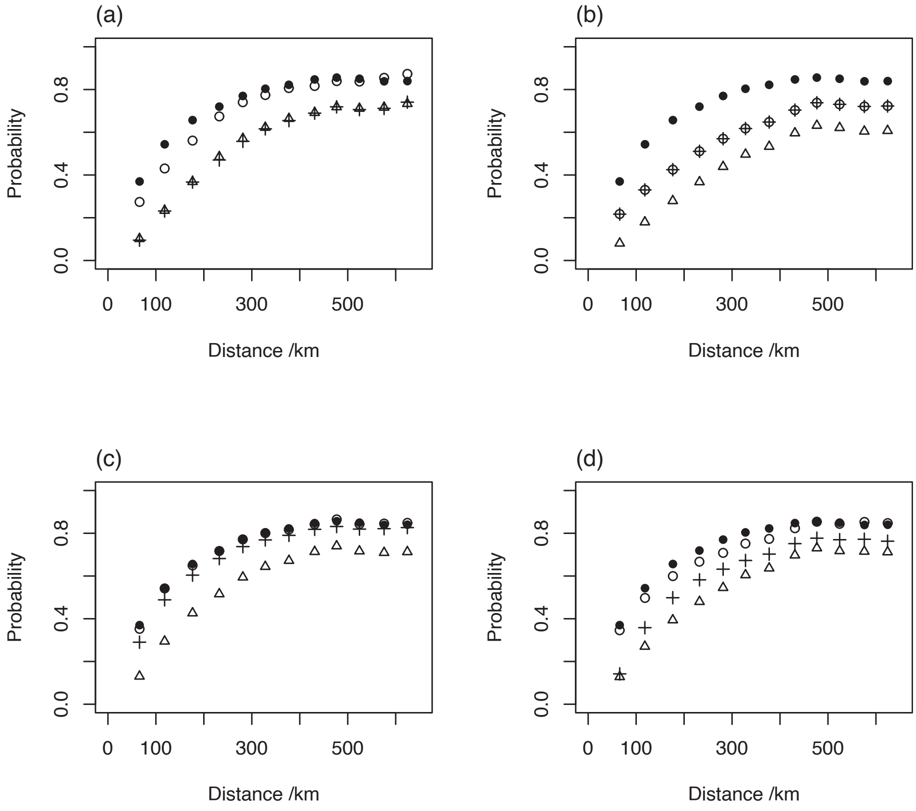

Fig. 2. Functions from simulated patterns of spatial classes. In each plot $\zeta_I(\mathbf{h})$ is shown by a solid disc, $\zeta_J(\mathbf{h})$ by an open circle and $\zeta_{I,J}(\mathbf{h})$ by a cross. The open triangle denotes $\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})$. In all cases the points in legend I are simulated as the quintiles of a simulated standard gaussian random variable, Z_1, with a Matérn variogram function. The points in legend J are simulated as follows. (a) quintiles of a second independent random variable, Z_2, with the same parameters as Z_1. (b) quantiles (0.4,0.6,1) of Z_1, so legend I is nested perfectly within J (c) quantiles (0.22,0.42,0.62,0.82) of Z_1 (d) quantiles (0.3,0.5,0.7,0.9,) of Z_1.

We estimated the functions $\zeta_I(\mathbf{h}), \zeta_J(\mathbf{h})$ and $\zeta_{I,J}(\mathbf{h})$ for each pair-wise comparison of legends on the half-degree national grid and on the Kafue Flats grid separately (where I and J denote the two legends in no particular order). Each pairwise comparison between points was considered in turn. The distance between two locations, and the final bearing of the great circle line joining the points was computed using the `distVincentySphere` function from the `geosphere` package for the R platform (Hijmans, 2019; R Core Team, 2020). The comparisons were sorted into lag bins defined on distance (Webster and Oliver, 2007) of width 50 km (half-degree grid) or 20 km (Kafue Flats grid). Within each lag bin the proportion of point to point comparisons which fell within different legend units was computed for each legend, and, for each pair of legends, the proportion of comparisons for which points fell in different units on both legends were computed. These provide estimates of $\zeta_I(\mathbf{h})$, $\zeta_J(\mathbf{h})$ and $\zeta_{I,J}(\mathbf{h})$ which can be plotted for interpretation. The values of $\zeta_{I,J}(\mathbf{h})$ at each lag were then normalized over the range defined by $\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})$ and $\min\{\zeta_I(\mathbf{h}), \zeta_J(\mathbf{h})\}$, so that the normalized value,

$$\frac{\min\{\zeta_I(\mathbf{h}), \zeta_J(\mathbf{h})\} - \zeta_{I,J}(\mathbf{h})}{\{\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})\} - \min\{\zeta_I(\mathbf{h}), \zeta_J(\mathbf{h})\}}, \qquad (18)$$

falls in the interval [0,1].

3. Results

3.1. Linear mixed models

Table 1 shows summary statistics for the data on bulk density, and for the residuals from exploratory models for the main effect of depth in the case of sand and clay content and soil organic carbon content. The results are shown after a selected data transformation (logit for the proportions of sand and clay, and log for soil organic carbon).

Table 2 shows the results from the linear mixed modelling. The confidence intervals for the depth effect exclude zero in the case of sand, clay and organic carbon content, but not for bulk density. The estimated depth effect shows that the organic carbon content is generally larger in the topsoil than the subsoil, as is the sand content, with clay content larger in the subsoil.

Note that there was no evidence to reject the null hypothesis that the between-map unit variance is zero in the case of the soil units from the map of Trapnell et al. (1947), for any of the soil variables. In the case of soil organic carbon content, this was true for all the legends.

For the other variables there was evidence for a non-zero between-map unit variance for the SAA (Representative Soil Groups) and for the ESMZ soil units (FAO–Unesco classes). In the case of bulk density there was also evidence for a non-zero between-map unit variance in the case of the ESMZ physiographic units, but the intra-class correlation for ESMZ physiographic units was markedly smaller than that for the soil classes in ESMZ and SAA. The intra-class correlations are largest for clay content.

3.2. Cross-tabulations

For the Half-degree grid, the cross-tabulation of Trapnell's soil units and the RSG of the SAA (Table 4a), we note the following key points.

1. Trapnell's Plateau soils are divided primarily between three substantial Representative Soil Groups, the Ferralsols, Acrisols and Leptosols. Trapnell et al. (1947) describe the Plateau soils as those formed on the 'mature topography of older land forms.' These soils typically have a horizon of iron nodules as the profile grades to the regolith, which is ascribed to the intense rainfall received and poor drainage. Because the soils have been subject to prolonged and intensive rainfall they are typically base deficient and very acidic. The texture of the Plateau soils is described as variable, with three basic types: (i) light-coloured sandy loams, (ii) pallid sandy soils over a buff-coloured subsoil (found mainly on the central plateau) and (iii) clay soils with a pale yellow to orange colour (mainly on the northern plateau). In addition, Trapnell et al. describe two groups of Plateau soils with ironstone pellets or massive concretions near the surface. These may be found on the main plateau, resulting from surface denudation (truncation), or on valley flanks, again as a result of erosion. It is notable on the SAA that Ferralsols are found primarily in the northern plateau areas and Acrisols in the central and southern plateau. This is consistent with type (ii) above corresponding primarily to Acrisols, and type (iii) to Ferralsols. Similarly the truncated soils, whether on the plateau or valley flanks, are consistent with Leptosols, shallow soils over continuous rock, and stony or gravelly soils (World Reference Base, 2006).
2. Trapnell's Upper Valley soils are sparse on this grid, and correspond mainly to the Luvisols RSG. Trapnell and Clothier (1937) state that the Upper Valley soils were first distinguished because of their distinctive vegetation cover, but that they are fundamentally differentiated from the Plateau soils by environmental factors. While the Plateau soils largely reflect pedogenesis under previous climatic conditions, the Upper Valley soils are formed in lower-lying environments, where recent modification of the landsurface has occurred and soil formation has taken place under somewhat warmer and drier conditions than those which dominated the development of the Plateau soils (Trapnell, 1943). The soils are typically colluvial or residual (Trapnell and Clothier, 1937), with a larger base saturation than Plateau soils, and increased content of soluble bases in the subsoil, Trapnell et al., 1947). The correspondence with Luvisols on the SAA is consistent with these observations. Luvisols have a pronounced textural contrast, with an argic subsoil (World Reference Base, 2006), consistent with the observation by Trapnell et al. (1947) of an increase in exchangeable base content with depth, Trapnell's (1943) description of various Upper Valley soil profiles in which the subsoils have a more clayey texture than the surface horizons. Luvisols are also characteristically formed in unconsolidate parent materials including colluvium (World Reference Base, 2006), which is also consistent with the observations of Trapnell and Clothier (1937) about the Upper Valley soils.
3. Trapnell's Lower Valley Soils are sparse on this grid and, mainly correspond to the Luvisols. Some Leptosols correspond to Trapnell's Lower Valley Soils. Trapnell et al. (1947) note that the soils within the Lower Valley group are very heterogeneous, and Trapnell (1943) notes that they include immature skeletal soils (which is consistent with the correspondence to Leptosols, which could also occur because of the occurrence of truncated stony soils on valley flanks as discussed in respect of the Plateau soils above. Trapnell (1943) also notes that the Lower Valley Soils formed in Karoo sedimentary beds,

Table 1
Summary statistics of soil properties. If the selected fixed effects model is for a constant mean, then these summary statistics are for the raw data. If depth was selected as a fixed effect, then summary statistics are for residuals from an exploratory fit of this model.

Soil property	Units	Fixed effect	Mean	Median	Standard deviation	Quartile 1	Quartile 3	Skewness	Octile skewness	n
Bulk density	g cm^{-3}	Overall mean	1.45	1.46	0.21	1.32	1.61	−0.10	0.02	285
Sand content	$\log(S/(1-S))$	Depth	0.00	0.00	1.09	−0.64	0.52	1.36	−0.11	426
Clay content	$\log(C/(1-C))$	Depth	0.00	0.13	1.04	−0.45	0.67	−1.24	−0.11	426
Soil organic C	log % mass	Depth	0.00	−0.25	1.21	−0.74	0.55	0.61	0.35	314

Table 2
Results for linear mixed models with between-map unit random effect considered for inclusion.

Property	Legend unit	*P*	Variance components[1]					Map unit effect[2]		Depth effect[3]	
			Within-profile	Between-profile within-location	Between-location (within-map unit)[4]	Between-map unit	r_{MU}[5]	*L*	*p*	Depth	Confidence interval
Bulk density	Trapnell-soil	7	0.013	8.52E-05	0.038	—	—	0.00	1.000	−0.021	−0.056, 0.014
	ESM-FAO	8	0.013	0.001	0.020	0.019	0.49	6.99	0.004	−0.029	−0.063, 0.004
	ESM-phys	5	0.013	0.001	0.028	0.011	0.29	3.15	0.038	−0.029	−0.063, 0.004
	SAA-RSG	7	0.013	0.001	0.021	0.015	0.42	7.91	0.002	−0.030	−0.063, 0.004
Sand content	Trapnell-soil	8	0.18	0.22	0.86	—	—	0.00	1.000	0.68	0.57, 0.78
	ESM-FAO	10	0.20	0.35	0.47	0.48	0.51	4.99	0.013	0.62	0.51, 0.73
	ESM-phys	5	0.20	0.36	0.82	—	—	0.08	0.39	0.62	0.51, 0.73
	SAA-RSG	8	0.20	0.36	0.45	0.42	0.48	6.43	0.005	0.62	0.51, 0.73
Clay content	Trapnell-soil	8	0.32	0.31	0.64	—	—	0.00	1.000	−0.92	−1.06,−0.79
	ESM-FAO	10	0.32	0.31	0.19	0.55	0.74	11.79	<0.0001	−0.88	−1.01,−0.75
	ESM-phys	5	0.32	0.33	0.58	—	—	1.91	0.080	−0.88	−1.06,−0.79
	SAA-RSG	8	0.32	0.35	0.17	0.61	0.78	9.51	0.001	−0.88	−1.01,−0.75
Organic carbon	Trapnell-soil	8	0.33	0.18	1.24	—	—	0.00	1.000	1.11	0.95,1.27
	ESM-FAO	10	0.34	0.19	1.27	—	—	0.45	0.25	1.11	0.95,1.27
	ESM-phys	5	0.34	0.19	1.27	—	—	0.13	0.36	1.11	0.95,1.27
	SAA-RSG	8	0.34	0.19	1.27	—	—	0.00	1.000	1.11	0.95,1.27

[1] These are the variance components from the finally-selected model (so, if there was no evidence to reject a null hypothesis of a zero variance component for the between-map unit effect, then these variance components are from a model with just three random effects).

[2] Based on a comparison between the residual log-likelihoods for models with or without a between-map unit effect, and with fixed effect either a constant mean or a topsoil vs subsoil effect, this choice being based on the generalized least squares estimate of fixed effects coefficients for a model with random effects selected on the basis of a log-likelihood ratio test.

[3] Generalized least-squares estimates based on a random effects model selected on the basis of a log-likelihood ratio test. A soil sample is defined as topsoil if its upper depth limit is at zero or if the upper depth limit is below the surface but the lower depth limit is shallower than 20 cm. The effect recorded here is the topsoil mean minus the subsoil mean.

[4] Only interpreted as between-location within map unit in a model with a between-map unit random effect, otherwise just between-location.

[5] The correlation is based on the between-location-within-map unit and between-map unit variance components.

Table 3
Symbols for Legend Units used in the analysis. The asterisk, *, indicates those units for which soil physical data were available, and the obelus, †, indicates a unit which appeared in the half-degree grid or Kafue Flats grid.

Trapnell soil unit		Soil Atlas of Africa Representative Soil Group		Exploratory Soil Map of Zambia physiographic units	
B	Lake Basin soils†	AC	Acrisols†*	A	Floodplains†*
E	Escarpment Hill†	AR	Arenosols†	D	Dambos†
K	Kalahari Sand†*	FL	Fluvisols†	Es	Sedimentary plains†*
L	Lower Valley†*	FR	Ferralsols†*	H	Hills and minor scarps†
P	Plateau†*	GL	Gleysols†*	He	Escarpment hills†
S	Grey Clays and alluvium†*	HS	Histosols†	M	Mountains and major scarps†
Sk	Transitional sands†*	LP	Leptosols†*	Pd	Dissected plateau†*
SP	Units on Plateau and Ironstone†*	LV	Luvisols†*	Pu	Undulating plateau†*
U	Upper Valley†*	LX	Lixisols†	S	Swamp unit†
R	Red Earths/ Loams*	NT	Nitisols*	ST	Swamp unit†
		PZ	Podzols†*	Vd	Footslopes and dissected upper valleys of rift valley†
		VR	Vertisols†*	Vt	Older alluvial plains and higher river terraces in rift valley†*

where not eroded, have a pronounced texture contrast between light-brown surface loams and dark-brown clay loam subsoil, again consistent with the correspondence to Luvisols on the SAA.

4. Trapnell's Kalahari Sands correspond closely to Arenosols on the SAA. World Reference Base (2006) describes Arenosols as soils with little or no development, formed in unconsolidated and typically translocated sandy parent material which may be calcareous. Again, this is consistent with the description by Trapnell and Clothier (1937) of deep Kalahari Sand soils, with 'frosting' indicative of aeolian origin, and with very low fertility. The Kalahari Sand is an important regional superficial geological unit, comprising Quaternary aeolian deposits (Thomas, 1987). Trapnell and Clothier (1937) note that in places where the aeolian deposits are thinner the Kalahari Sand soils may have greater textural variation with clay derived from underlying material such as Karroo of older formations. There may be heavier soils on plains or dambos, so the occurrence of some Gleysols on the SAA at locations mapped as Kalahari Sand by Trapnell et al. (1947) is not necessarily inconsistent.
5. Trapnell's Escarpment Hill soils correspond closely to Leptosols, (although most Leptosols correspond to Plateau soils, and it has been shown above that this is consistent with Trapnell's account of Plateau soils. The Escarpment Hill Soils are described by Trapnell et al. (1947) as shallow and stony soils, immature due to their location on steep slopes. The Leptosols of the SAA therefore correspond to a range of environments (escarpment slopes or truncated profiles on the plateau or valley flanks), which Trapnell's map differentiates.

We examined the Plateau soil unit of Trapnell et al. (1947) more closely, cross-tabulating the six separate vegetation–soil units which it comprises in this sample against the SAA RSG (Table 4b). We tested the null hypothesis that the Trapnell vegetation–soil units and the RSG are independently assorted by means of a log-linear model implemented with the `loglm` function in the MASS package for the R platform (Venables and Ripley, 2002; R Core Team, 2020). There was substantial evidence against the null hypothesis (deviance $= 71.2$, 30 d.f., $p = 3.3 \times 10^{-5}$). Note that the Plateau soils of Trapnell under Central *Isoberlinia*

Table 4a

Cross-tabulation of units from Trapnell and Representative Soil Groups on the Soil Atlas of Africa. National half-degree grid.

		Soil Atlas of Africa Representative Soil Group										
		AC	AR	FL	FR	GL	HS	LP	LV	LX	PZ	VR
Trapnell Soil Unit	B	1	0	0	3	0	1	0	1	0	0	0
	E	0	0	0	0	0	0	13	4	0	0	0
	K	0	27	0	2	5	0	0	0	0	4	1
	L	0	1	0	0	1	0	4	8	0	0	0
	P	22	3	0	40	3	0	21	2	1	0	0
	S	3	1	1	4	4	3	0	0	0	0	5
	Sk	0	4	0	0	10	0	0	0	0	0	0
	SP	1	0	0	0	0	0	0	0	0	0	0
	U	1	0	0	1	0	0	0	7	0	0	0

Table 4b

Cross-tabulation of Trapnell's vegetation units within the Plateau soil unit and Representative Soil Groups on the Soil Atlas of Africa. National half-degree grid.

Trapnell units		SAA Representative Soil Group						
Symbol	Vegetation unit	AC	AR	FR	GL	LP	LV	LX
P2	Northern *Brachystegia* Woodland	0	1	9	0	2	0	0
P3	Northern *B. – Isoberlinia globiflora* Woodland	0	0	9	0	3	0	0
P4	Northern *B. – I. paniculata* Woodland	2	0	14	0	9	0	0
P5	Central *I. paniculata – B.* Woodland	14	2	5	2	3	0	0
P6	Eastern *B. – I.* Woodland	2	0	3	0	1	2	1
P7	Southern *I. globiflora – B.* Woodland	4	0	0	1	3	0	0

paniculata – Brachystegia woodland (unit P5) corresponded predominantly to Acrisols of the SAA, whereas Plateau soils under the Northern *Brachystegia*-dominated woodland (units P2, P3 and P4) corresponded primarily to Ferralsols.

For the Half-degree grid, the cross-tabulation of Trapnell's soil units and the physiographic units of the ESMZ (Table 4b), we note the following key points.

1. Trapnell's Plateau soils almost all correspond to the ESMZ's Undulating Plateau with a few in Dissected Plateau, this is where most of the Dissected Plateau soils are found. Similarly, Trapnell's Upper Valley soils are mostly in Undulating Plateau, the rest in Dissected Plateau, although proportionally the Dissected Plateau is more predominant over the Plateau soils than the Upper Valley soils. The account of the geomorphic units from the ESMZ, given by Dalal-Clayton et al. (1985), indicates that the Dissected Plateau unit, with denser drainage than the Undulating Plateau and steeper incised stream valleys, is typically found at the Plateau margin as a transitional unit to the Escarpment Hill soils. This is consistent with the Upper Valley soils corresponding, inter alia, to Dissected Plateau soils on ESMZ, because, as noted above, the Upper Valley soils are formed predominantly in colluvial material where the plateau is subject to rejuvenation.
2. Trapnell's Kalahari Sands and Transitional sands (K and SK) correspond largely to the ESMZ's Es (Sedimentary Plains). This ESMZ unit corresponds to the aggraded plateau (Dalal-Clayton et al., 1985), predominantly deep Kalahari Sand with wide-spaced drainage channels, but also including the floodplains and terraces of the upper Zambezi river. Most of the ESMZ soil units in this physiographic unit are Arenosols (FAO-Unesco, 1974), consistent with the close correspondence between Trapnell's unit and the Arenosols on SAA.
3. Trapnell's Escarpment Hill soils correspond largely to the Escarpment Unit of ESMZ.

For the Kafue Flats grid, the cross-tabulation of Trapnell's soil units

Table 6a

Cross-tabulation of units from Trapnell and Representative Soil Groups on the Soil Atlas of Africa. Kafue Flats Grid.

		Soil Atlas of Africa Representative Soil Group						
		AC	AR	GL	HS	LP	LV	VR
multirow96ptTrapnell Soil Unit	E	2	0	0	0	16	3	0
	K	0	3	0	0	0	0	0
	L	0	0	0	0	0	6	0
	P	59	0	1	0	2	11	1
	P/U	0	0	0	0	0	1	0
	S	10	1	3	1	0	4	20
	SP	0	0	0	0	0	1	0
	Sw	0	0	0	3	0	0	0
	U	7	0	0	0	1	29	4

Table 5

Cross-tabulation of units from Trapnell and physiographic units of the Exploratory Soil Map of Zambia. National half-degree grid.

		ESMZ physiographic unit										
		A	D	Es	H	He	M	Pd	Pu	S	ST	Vd	Vt
Trapnell Soil Unit	B	1	0	1	0	0	0	0	4	0	0	0	0
	E	0	0	0	0	9	0	2	3	0	0	1	2
	K	1	4	32	0	0	0	0	2	0	0	0	0
	L	2	0	1	0	1	0	1	1	0	0	3	5
	P	0	2	4	5	5	1	9	62	1	0	0	3
	S	9	0	3	0	0	0	0	5	3	1	0	0
	Sk	2	0	12	0	0	0	0	0	0	0	0	0
	SP	0	0	0	0	0	0	0	1	0	0	0	0
	U	0	0	0	0	0	0	3	6	0	0	0	0

Table 6b

Cross-tabulation of Trapnell's vegetation units within the Plateau soil unit and Representative Soil Groups on the Soil Atlas of Africa. Kafue Flats grid. Note that vegetation unit names are given as in the original map legend. According to the table of synonymy for botanical names published by Smith and Trapnell (2001), *Isoberlinia globiflora* and *I. paniculata* are now known as *Julbernardia globiflora* and *J. paniculata* respectively.

Trapnell units		SAA Representative Soil Group			
Symbol	Vegatation unit	AC	GL	LV	VR
P5	Central *I. paniculata* – *B.* Woodland	31	1	2	0
P7	Southern *I. globiflora* – *B.* Woodland	28	0	9	1

and the RSG of the SAA (Table 4a), we note the following key points.

1. Trapnell's Plateau soils correspond predominantly to Acrisols. There are no Ferralsols mapped on the SAA in this grid. This is consistent with our inference above that the Ferralsols are likely to correspond to the yellow–orange clay Plateau soils which Trapnell et al. (1947) identify as prevalent on the northern plateau.
2. Trapnell's Upper Valley soils correspond predominantly to Luvisols. Again, this is consistent with our interpretation of the relationship between the RSG and Trapnell's units on the half-degree national grid.
3. Similarly, all occurrences of Trapnell's Lower Valley soils on the Kafue Flats grid correspond to Luvisols on the SAA.
4. There are few grid points on Trapnell's Kalahari Sands soils, but all correspond to Arenosols on the SAA, which also corresponds to what was found on the national grid.
5. Trapnell's unit S (Mopani/Mopani-grassland over clays, and some grey dambo clays and Valley grasslands) correspond mainly to Vertisols, and some Gleysols and Acrisols on the SAA. Trapnell et al. (1947) note that clay soils around the Kafue Flats and in similar low-lying areas are typically cracking clays prone to swelling and shrinking and 'crabhole' or gilgai microtopography. This is consistent with the delineation of many of these soils in this grid as Vertisols on the SAA. Trapnell and Clothier (1937) note that the Grey Mopani clays, in contrast, are not swelling and shrinking. They describe the clays as alluvial in origin, and having been subject to substantial eluviation, and hence a loss of silica and bases. This is consistent with the significant number of grid nodes from unit S which are mapped as Acrisols on SAA.
6. Trapnell's Escarpment Hill soils correspond predominantly to Leptosols, as was found on the half-degree national grid.

As with the national half-degree grid we examined more closely the cross-tabulation of the SAA RSGs with Trapnell's vegetation–soil units over Plateau soils (Table 6b) and Upper Valley soils (Table 6c). In the former case there are two vegetation units – P5 and P7, in which different species of *Isoberlinia* were found along with *Brachystegia.* As seen in Table 6b, both these units correspond primarily to the Acrisols of SAA, and while there are somewhat more grid nodes over Luvisols of SAA in unit P7 than P5, the composition of Trapnell's units with respect to soil units of the SAA are not notably different (note that there is weak evidence against the null hypothesis of random association between the two sets of classes from these data, deviance $= 7.52$, 3 d.f., $p = 0.06$). In the case of the Upper Valley soil units of Trapnell, both correspond predominantly to the Luvisols of SAA, although Vertisols of SAA are found only in unit U3 (there is somewhat stronger evidence against random association of the two sets of map legend units in this case, deviance $= 8.77$, 3 d.f., $p = 0.03$).

Table 6c

Cross-tabulation of Trapnell's vegetation units within the Upper Valley soil unit and Representative Soil Groups on the Soil Atlas of Africa. Kafue Flats grid. Note that vegetation unit names are given as in the original map legend. According to the table of synonymy for botanical names published by Smith and Trapnell (2001), *Afrormosia angolensis* is now known as *Pericopsis angolensis*. Other names of genera in this table are extant, although note that at the time of Trapnell the species now known as *Faidherbia albida* (Winterthorn) was included in the genus *Acacia*.

Trapnell units		SAA Representative Soil Group			
Symbol	Vegetation unit	AC	LP	LV	VR
U2	*Combretum–Afrormosia* and*Pterocarpus–Combretum* vegetation	5	1	18	0
U3	*Acacia–Combretum* and allied vegetation	2	0	11	4

Table 7

Cross-tabulation of units from Trapnell and physiographic units of the Exploratory Soil Map of Zambia. Kafue Flats grid.

		ESMZ Physiographic Unit							
		A	Es	H	He	Pd	Pu	S	Vd
Trapnell Soil Unit	E	0	0	0	18	1	1	0	1
	K	2	1	0	0	0	0	0	0
	L	0	0	0	3	0	1	0	2
	P	4	1	8	1	8	52	0	0
	P/U	0	1	0	0	0	0	0	0
	S	21	5	1	0	0	12	0	0
	SP	0	0	0	0	0	1	0	0
	Sw	0	0	0	0	0	0	3	0
	U	15	0	0	1	7	18	0	0

For the Kafue Flats grid, the cross-tabulation of the Trapnell's soil units and the physiographic units of the ESMZ (Table 6a), we note the following key points.

1. Trapnell's Plateau soils correspond predominantly to the Undulating Plateau on ESMZ, with some on the Dissected Plateau and Hills and minor scarps. This is consistent with what was seen on the half-degree national grid.
2. Trapnell's Upper Valley soils are divided almost equally between the Undulating Plateau (with some on the Dissected Plateau) and Floodplain soils of ESMZ. Given that the Upper Valley soils correspond primarily to Luvisols, formed in colluvium arising from rejuvenation of the plateau, this distribution between ESMZ units is interpretable, as the Upper Valley soils will be transitional between the margins of the plateau and the current drainage channels. Inspection of the ESMZ in the vicinity of the Kafue Flats shows that the floodplain environment borders the Undulating Plateau directly over some of its boundary, and Dissected Plateau units in the east near Lusaka. In contrast, the map of Trapnell et al. (1947) shows Upper Valley soils around most of the margin of the alluvial and grey clay soils of the Flats, except in the south west where they are bounded by various units of Kalahari Sand soils. This indicates that the geomorphological account that underlies the ESMZ legend units is somewhat different from the interpretation of Trapnell et al. (1947) with the Upper Valley soils corresponding to both Undulating and Dissected plateau units.
3. Grid nodes on Trapnell's Escarpment Hill soils are primarily over the ESMZ's Escarpment Hills, consistent with what we saw on the half-degree national grid.

3.2.1. *Sum of indicator variograms*

For the half-degree national grid, the sums of indicator variograms for the Trapnell soil units and the ESMZ physiographic units are similar (Fig. 3). We note that they all show pronounced transitive behaviour (the SIV increases with lag distance then levels off) indicating that the half-degree grid captures a spatially dependent pattern of soil variation as represented in the different maps. Over most distances the normalized value of $\zeta_{I,J}(\mathbf{h})$ is larger for this pair of legend units than for any other. The sum of indicator variograms is larger for the SAA units than for

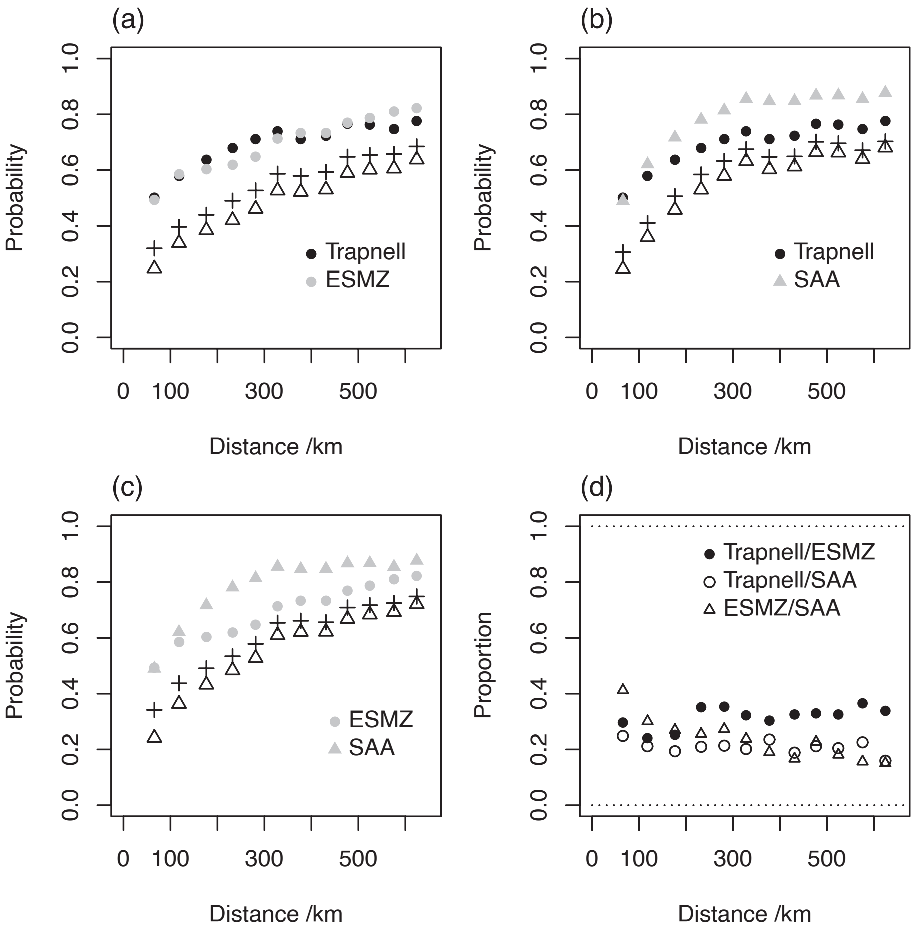

Fig. 3. Sum of indicator variogram estimates for legend units on the half-degree national grid. Plots (a) to (b) show estimates of $\zeta_I(\mathbf{h})$ for each of a pair of legends, with the estimate of $\zeta_{I,J}(\mathbf{h})$ (cross) and its lower bound, $\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})$ shown as an open triangle. In plot (d) is shown, for each pair of legends, the value of $\zeta_{I,J}(\mathbf{h})$ normalized over the range of its bounds at each lag distance.

either the Trapnell soil units or the ESMZ physiographic soil units, showing that, over this grid, the pattern of soil variation described by these units is the most intricate, in the sense that more heterogeneity in the delineated classes is seen over a particular distance.

On the Kafue Flats grid the sums of indicator variograms are very similar for all three sets of legend units (Fig. 4). Over most distances the normalized value of $\zeta_{I,J}(\mathbf{h})$ takes values between 0.2 and 0.4, and is largest for the Trapnell and SAA legend units.

4. Discussion

It was notable that the soil legend units based on the WRB or FAO-Unesco soil classes (SAA RSGs and EZMZ soil classes) accounted for significant variation in measured soil properties (with the exception of soil organic carbon). The proportion of variation accounted for is substantial — 40 to 50% of the variance of bulk density and sand content at between location scales, and over 70% of the variance of sand content — which is surprising given the small scale of the corresponding maps. The maps could be used as a basis for coarse, regional-scale modelling where properties such as soil texture or bulk density are relevant (e.g. for modelling broad trends in groundwater recharge), and certainly provide a framework for more detailed survey of soil properties. By contrast, differences among these map units did not account for variation in soil organic carbon. This may not be surprising, given the importance of land use and management in controlling this variable, a factor which may be expected to vary at much finer scales than are captured by such soil maps. Note that, while the fact that legacy soil surveys can be shown to account for substantial variation in legacy data on soil properties, this does not necessarily mean that the values of those properties can be assumed to be predictive of current conditions, particularly for variables prone to change. However, they do suggest that the maps account for variation in soil factors of formation and so might still provide a framework to improve sampling and modelling of, and prediction from, new measurements on the soil.

The physiographic units of the ESMZ accounted for significant variation in the bulk density of soil, but accounted for a notably smaller proportion of variation than did the soil classes (nested within the physiographic units in the legend structure), and it is clear that the variation of soil texture within these physiographic units is substantial, such that the contribution of between-unit differences was not significant. Similarly, Trapnell's soil units, which are essentially broad physiographic associations of soils, did not account for significant variation of the soil properties. There were not sufficient data on soil properties available for it to be feasible to compare the vegetation–soil units of Trapnell et al. (1947) in the same way. Further research would be needed with new sampling, to determine whether the differing vegetation classes which Trapnell and colleagues recognized and mapped reflect variations in underlying soil conditions which still persist.

Whilst the results for the physiographic units of ESMZ suggest that such units are too broad to account for substantial soil variation, it must also be accepted that other sources of error may reduce the value of Trapnell's soil units for this purpose. Trapnell et al. (1947) acknowledge that, over most of Zambia where no airphotography was available at the time, the prediction of map units away from the routes of the original

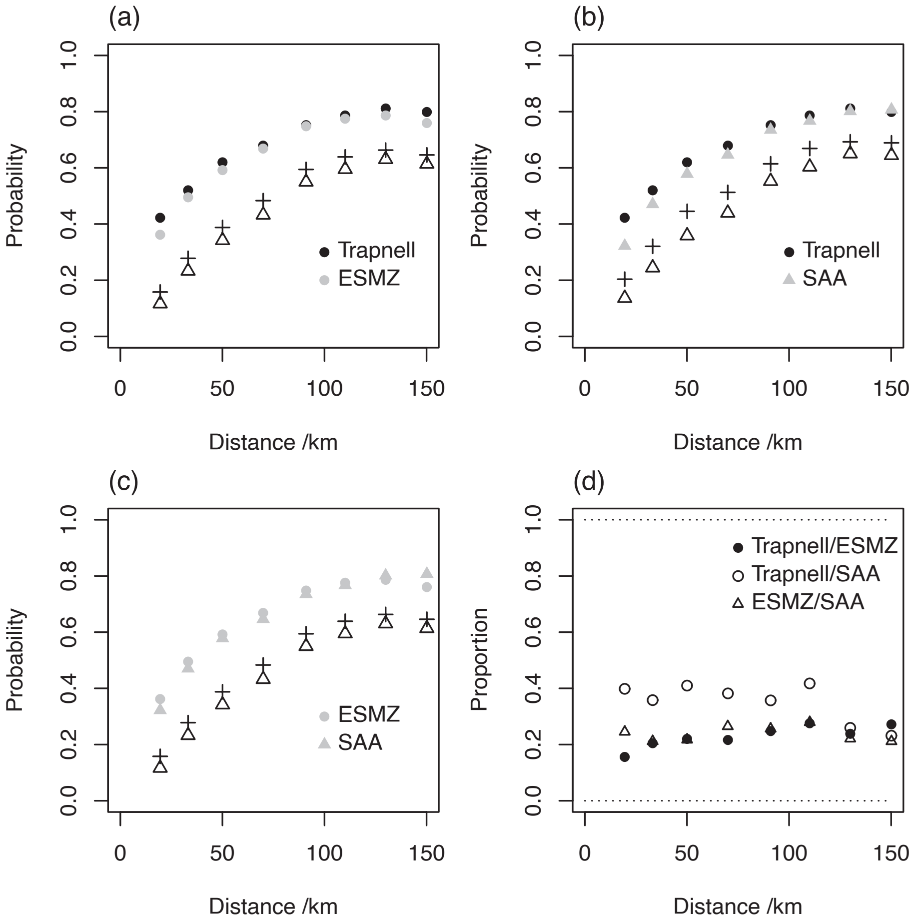

Fig. 4. Sum of indicator variogram estimates for legend units on the Kafue Flats grid. Plots (a) to (b) show estimates of $\zeta_I(\mathbf{h})$ for each of a pair of legends, with the estimate of $\zeta_{I,J}(\mathbf{h})$ (cross) and its lower bound, $\zeta_I(\mathbf{h}) \times \zeta_J(\mathbf{h})$ shown as an open triangle. In plot (d) is shown, for each pair of legends, the value of $\zeta_{I,J}(\mathbf{h})$ normalized over the range of its bounds at each lag distance.

traverses is speculative.

At national scale (on the half-degree grid), the sums of indicator variograms and the cross-function, $\zeta_{I,J}(\mathbf{h})$ in Fig. 3 indicated that the physiographic units of ESMZ and the soil units of Trapnell et al. (1947) represent spatial variation at comparable scales and with units that are related to each other. As seen in the cross-tabulation there are clear relations between, for example, Trapnell's Plateau soils and the undulating and dissected plateau units, Pu and Pd, of the ESZM. Similarly, Trapnell's units over the Kalahari sands (K) and Transitional Sands (Ks) correspond predominantly to the Sedimentary plains (Es) of ESMZ.

Although the number of soil units (RSG) occuring on SAA on the half-degree national grid (11) is not much larger than the number of Trapnell's soil units (9, see Table 4a) it is clear from the sums of indicator variograms that the former show greater spatial variation over the grid, and capture variation of soil conditions within the larger soil units of Trapnell et al. (1947). In particular, three RSG are mapped over the Plateau unit of Trapnell. Acrisols, soils with a pronounced texture contrast between the lighter topsoil and more clay-enriched subsoil are found almost exclusively at sites in the Plateau unit on the half-degree grid, as are Ferralsols – deeply weathered red or yellow soils with a large content of sesquioxides and low-activity clays. Trapnell et al. (1947) describe variation within the Plateau soils unit primarily in terms of colour and texture, noting the presence of light-coloured sandy loams, pallid grey to white sandy soils over a 'buff'-coloured subsoil (central plateau regions) and yellow to orange clay soils mainly on the northern plateau. The Ferralsols on SAA are mapped in the northern plateau, and the Acrisols in the southern and central plateau, suggesting that these correspond at least in part to within-Plateau soil variation observed by Trapnell and colleagues. Furthermore, the Leptosols, the third RSG which corresponds significantly to Trapnell's Plateau soils, are plausibly equivalent to shallow stony soils that Trapnell et al. (1947) describe as occuring in the Plateau soil group, truncated soils on the plateau, and similar soils on valley flanks.

It is interesting that when the vegetation units over the Plateau soils on the half-degree grid were examined more closely (Table 4b), it was seen that these differed with respect to the proportions of contrasting units from SAA, and that Acrisols were more common on the Plateau soils of the central plateau under *Isoberlinia paniculata* – *Brachystegia* woodland, whereas Ferralsols predominated under the three contrasting vegetation units of the Northern plateau. The soil units of the SAA appear to reflect, at least to some extent, differences in soil between vegetation units which Trapnell and colleagues had recognized, and our results from the linear mixed models for soil properties suggest that this variation, reflected by the units of SAA, accounts for substantial variations in soil properties.

The denser grid centred around the Kafue Flats region encompasses contrasting soil landscapes. However, the spatial variation reflected by the Trapnell soil units, ESMZ physiographic units and RSG of the SAA as reflected in their sums of indicator variograms (Fig. 4) is of similar magnitude and spatial scale. The similarity between spatial patterns of these units, as measured by the cross-function $\zeta_{I,J}(\mathbf{h})$ is greatest for the Trapnell soil units and RSG of the SAA (Fig. 4b). The Plateau unit of Trapnell et al. (1947) corresponds primarily to Acrisols on the SAA in this grid, although with some Luvisols, and the two vegetation units over

Plateau soils do not appear to differ markedly from each other with respect to the proportions of RSG (Table 6a and 6b). Trapnell's Upper Valley soil unit in this grid corresponds primarily to Luvisols of the SAA, and there is some evidence for a difference between the two vegetation units of Trapnell with respect to RSG units.

In summary, much of the variation of soil properties (texture and bulk density) in our legacy soil data from Zambia occurs within the physiographically-defined soil units of Trapnell et al. (1947), and the related physiographic units of ESMZ. A surprising amount of this variation is accounted for by differences between soil units mapped at national scale (ESMZ at 1:1 million; SAA at 1:3 million). The largest soil unit in the map of Trapnell et al. (1947), the Plateau soils, corresponds to three dominant SRG of the World Reference Base as mapped in the SAA, and there is evidence that the within-Plateau variation that the SAA map units represent is associated with vegetation units that Trapnell and colleagues recognized and mapped as distinct vegetation–soil units. On the basis of this observation we suggest that further work, with a larger set of observations of soil properties in the field, is needed to compare the variation of soil properties between the soil units of SAA and the ESMZ and the vegetation–soil units of Trapnell et al. (1947), and to evaluate these alternative representations of soil variation in Zambia.

5. Conclusions

In conclusion, we have shown that differences between soil map units of two small-scale soil maps of Zambia, the Exploratory Soil Map of Zambia (ESMZ; Soil Survey Section, 1991) and the Soil Atlas of Africa (SAA; Jones et al., 2013) account for a significant and substantial proportion of variation in sand and clay content and bulk density of the soil. Physiographic units of the ESMZ, and the comparable soil units of the map of Trapnell et al. (1947) encompass more soil variation, and the only significant evidence for a map-unit effect was for bulk density between units of the ESMZ. The variability of the soil within the broad soil units of Trapnell et al. (1947) had been recognized by those authors.

Comparing the different legend units of these three maps on two sample grids showed the similarity between the physiographic units of the ESMZ and Trapnell's map, although these do not directly correspond. The soil units of the SAA show greater spatial variability. It was apparent that the variation of the soil within the largest unit of Trapnell et al. (1947), the Plateau soils, as expressed by contrasting map units of SAA, was associated to a significant extent with differences in the vegetation cover which Trapnell et al. (1947) had recognized and mapped. Further work is needed to examine directly the extent to which the different vegetation units, mapped in Zambia in the 1930s, reflect underlying soil variation which is of contemporary significance for agriculture and the water cycle. The fact that it was possible to interpret the SAA mapping of Acrisols, Ferralsols and Leptosols over the unit designated Plateau Soils by Trapnell et al. (1947) in terms of variation within the Plateau soils described by Trapnell and Clothier (1937), Trapnell (1943) and Trapnell et al. (1947) suggests that integrating older legacy maps with more recent surveys can illuminate both, and may allow us to relate the legacy observations on soil, vegetation and land use to contemporary descriptions of the soil.

Finally we note that this study has demonstrated some approaches that can be used to compare and evaluate legacy soil surveys of different age, and to understand better the extent to which they reflect a common underlying pattern of spatial variation of the soil.

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 the Arts and Humanities Research Council [Grant No. AH/T00410X/1] through their Programme *Cultures, Behaviours and Histories of Agriculture, Food, and Nutrition*, Global Challenges Research Fund. The project is entitled *Towards transdisciplinary understanding of inherited soil surveys: an exploratory case study in Zambia*. The collation of the soil physical data was undertaken as part of the Natural Environment Research Council-funded project *CEPHaS: Strengthening Capacity in Environmental Physics, Hydrology and Statistics for Conservation Agriculture Research* which is part of the GROW programme, Global Challenges Research Fund [Grant No. NE/P02095X/1].

The Soil Atlas of Africa digital shapefiles were used under licence (No. 34032) from European Soil Data Centre.

The authors are grateful to Ms. Mirriam Makungwe for extracting the soil map unit delineations from the scanned version of the Exploratory Soil Map of Zambia, and to Ms. Anna Lark for extracting the map unit delineations from the half-degree nodes on the map of Trapnell et al. (1947).

References

Allan, W., 1965. The African husbandman. Oliver & Boyd, Edinburgh.

Batjes, N.H., Ribeiro, E., van Oostrum, A., Leenaars, J., de Mendes, J. 2016. Standardised soil profile data for the world (WoSIS, July 2016 snapshot), https://doi.org/10.17027/isric-wdcsoils.20160003.

Batjes, N.H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., de Mendes, J., 2017. WoSIS: Providing standardised soil profile data for the world. Earth Syst. Sci. Data 9, 1–14. https://doi.org/10.5194/essd-9-1-2017.

Brys, G., Hubert, M., Struyf, A. 2003. A comparison of some new measures of skewness. In: Developments in robust statistics, eds R. Dutter P. Filzmoser, U. Gather, P.J. Rousseeuw. Physica-Verlag Heidelberg pp 98–113.

Cheatle, R.J. 1980 A possible Nitosol at York Farm, Makeni, near Lusaka, Zambia. Zambia Department of Agriculture, Soil Bulletin No. 4, Mt Makulu, Zambia.

Chileshe L. 1988. Soils of the Copperbelt province, memoir accompanying the 1: 1,000,000 soil map. Soil Survey Report No. 153. Soil Survey Unit. Chilanga. Lusaka.

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2015. Rescue and renewal of legacy soil resource inventories: a case study of the Limpopo National Park, Mozambique. Catena 125, 169–182.

Dalal-Clayton, B.C. 1974 Detailed soil survey of Chisamba Farms (Ntendere, Nsadzu and Penyaonse Blocks). Soil Survey Unit, Land use services Division, Ministry of Rural Development, Soil Survey Report No. 14, Mt Makulu, Zambia.

Dalal-Clayton, B.C. 1980 Investigations into the nature and distribution of sandveldt soils of the Central Province of Zambia with observations on their landuse potential. Soil Survey Unit, Land use Branch, Department of Agriculture, Soil Bulletin No. 6, Mt Makulu, Zambia.

Dalal-Clayton, D.B., English c., Williams G.J. and Spaargaren O.C. 1985. The Geomorphic Legend of Zambia. Technical Guide 15. Soil Survey Unit. Chilanga, Lusaka.

Dewitte, O., Jones, A., Spaargaren, O., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Gallali, T., Hallett, S., Jones, R., Kilasara, M., Le Roux, P., Micheli, E., Montanarella, L., Thiombiano, L., Van Ranst, E., Yemefack, M., Zougmore, R., 2013. Harmonisation of the soil map of Africa at the continental scale. Geoderma 211–212, 138–153.

FAO. 1985. The Soil Map of the World. Revised Legend World Soil Resources Report. Revised third draft. Rome.

FAO-Unesco, 1974. Legend of the Soil Map of the World. Food and Agriculture Organization of the United Nations, Rome, Italy.

FAO, IIASA, ISRIC, ISSCAS, JRC. 2012. Harmonized World Soil Database (version 1.2). Food and Agriculture Organization of the United Nations, Rome, Italy and IIASA, Laxenburg, Austria.

Forbes, T.R., Rossiter, D.G., Van Wambeke, A., 1982. Guidelines for evaluating the adequacy of soil resource inventories. Cornell University Department of Agronomy, New York.

Goovaerts, P., Webster, R., 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur. J. Soil Sci. 45, 79–95.

Hijmans, R.J. (2019). geosphere: Spherical Trigonometry. R package version 1.5-10. https://CRAN.R-project.org/package=geosphere.

Hodge, J.M., 2007. Triumph of the Expert. Agrarian doctrines of development and the legacy of British colonialism. Ohio University Press, Athens, OH.

Jones, A., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Dewitte, O., Hallett, S., Jones, R., Kilasara, M., Le Roux, P., Micheli, E., Montanarella, L., Spaargaren, O., Tahar, G., Thiombiano, L., Van Ranst, E., Yemefack, M., Zougmore, R. (Eds.), 2013. Soil Atlas of Africa. European Commission. Publication Office of the European Union, Luxembourg https://esdac.jrc.ec.europa.eu/content/soil-map-soil-atlas-africa.

Journel, A.G., 1983. Non-parametric estimation of spatial distributions. Math. Geol. 15, 445–468.

Landon, J.R., 1991. Booker Tropical Soil Manual: A Handbook for Soil Survey and Agricultural Land Evaluation in the Tropics and Subtropics. Routledge, London.

Lark, R.M., Beckett, P.H.T., 1998. A geostatistical descriptor of the spatial distribution of soil classes, and its use in predicting the purity of possible soil map units. Geoderma 83, 243–267.

Lee, C.A. 1968. Detailed soil survey Mount Makulu Central Research Station. Research Branch, Department of Agriculture, Zambia, Soil Survey Report No. 4, Mt Makulu, Zambia.

Maclean, A.H. 1970. Moisture characteristics of some Zambian soils. Soil Science Research Section Soil Survey Report No. 6. Department of Agriculture, Mt Makulu, Zambia.

Moore, H.L., Vaughan, M., 1994. Cutting down trees. Gender, nutrition and agricultural change in the Northern Province of Zambia. James Currey Press, London, pp. 1890–1990.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.

Panagos, P., Van Liedekerke, M., Jones, A., Montanarella, L., 2012. European Soil Data Centre: Response to European policy support and public data requirements. Land Use Policy 29, 329–338.

Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Rasaei, Z., Rossiter, D.G., Farshad, A., 2020. Rescue and renewal of legacy soil resource inventories in Iran as an input to digital soil mapping. Geoderma Regional 21, e00262.

Smith, P., Trapnell, C.G. 2001. Ecological survey of Zambia. The traverse records of C.G. Trapnell, 1932–1943. Three volumes. Royal Botanic Gardens, Kew.

SADCC. 1991. Zambia soil map (Draft(. Food Security Programme, Regional inventory of agricultural resource base, Harare, Zimbabwe.

Soil Survey Section, 1991. National Soils Map of Zambia. Ministry of Agriculture, Lusaka. Mapsheets available from ISRIC at https://edepot.wur.nl/486756 and https://edepot.wur.nl/486757.

Soil Survey Unit. 2011. Memoirs of the 1991 Exploratory Soil Map of Zambia. Edited by Sokotela S. (Unpublished Report). Mt Makulu, Zambia.

Stein, M.L., 1999. Interpolation of spatial data: some theory for kriging. Springer, New York.

Thomas, D.S.G. 1987. Discrimination of depositional environments using sedimentary characteristics in the Mega-Kalahari, central southern Africa. In (eds) L.E. Frostick and I. Reid. Desert Sediments: Ancient and Modern. Geological Society Special Publication No. 35, Blackwell Scientific Publications, Oxford. pp 293–308.

Tilley, H., 2011. Africa as a living laboratory. Empire, development and the problem of scientific knowledge 1870–1950. The University of Chicago Press, Chicago.

Ting-Tiang, W. 1987. The soils of North-Western Province. Memoir accompanying the 1: 1,000,000 soil map. Soil Survey Report No. 150, Soil Survey Research Unit, Department of Agriculture, Mt Makulu, Zambia.

Trapnell, C.G. 1935. A vegetational grouping of soils in Northern Rhodesia south of latitude 1530′. In G. Milne (Ed). Proceedings of the second conference of East African agriculltural and soil chemists, held at Zanzibar, August 3–9, 1934. Government Printer, Nairobi. pp. 53–55.

Trapnell, C.G., 1943. The soils, vegetation and agriculture of North-Eastern Rhodesia. Government Printer, Lusaka.

Trapnell, C.G., Clothier, J.N. 1937. The soils, vegetation and agricultural systems of North-Western Rhodesia. Government Printer, Lusaka. (Available from ISRIC at (Memoir) https://edepot.wur.nl/486617, (Sheet 1), http://edepot.wur.nl/486611 and (Sheet 2) http://edepot.wur.nl/486613.

Trapnell, C.G., Martin, J.D., Allan, W., 1947. Vegetation–soil map of Northern Rhodesia. Government Printer, Lusaka.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth Edition. Springer, New York.

Verbeke, G., Molenberghs, G., 2003. The Use of Score Tests for Inference on Variance Components. Biometrics 59, 254–262.

Webster, R., Oliver, M.A., 2007. Geostatistics for environmental scientists, 2nd Edition. John Wiley & Sons, Chichester.

Woode, P.R. (Ed.) 1985. Proceedings of the XIth International Forum on Soil Technology and Agrotechnology Transfer. Part II Excursions and Appendices.

Worthington, E.B., 1938. A review of scientific research relating to tropical and southern Africa. Oxford University Press, Oxford.

World Reference Base, 2006. World reference base for soil resources 2006. FAO, Rome.

Young, A., 2017. Thin on the ground, soil science in the Tropics, 2nd Edition. Land Resources Books, Norwich.