



Multivariate split moving windows and magnetic susceptibility for locating soil boundaries of São Paulo, Brazil



João Fernandes da Silva Júnior^{a,*}, Diego Silva Siqueira^b, Daniel De Bortoli Teixeira^c, Alan Rodrigo Panosso^d, José Marques Júnior^b, Gener Tadeu Pereira^d

^a Campus of Capanema, Federal Rural University of Amazon (UFRA), Research Group GEOP — Geotechnologies and Pedometrics, Capanema, Para, Brazil

^b Department of Agricultural Sciences, São Paulo State University (Unesp), Research Group CSME — Soil Characterization for Specific Management, Jaboticabal, São Paulo, Brazil

^c Center of Agrarian Sciences, University of Marília (UNIMAR), Marília, São Paulo, Brazil

^d Department of Engineering and Exact Sciences, São Paulo State University (Unesp), Research Group CSME — Soil Characterization for Specific Management, Jaboticabal, São Paulo, Brazil

ARTICLE INFO

Keywords:

Pedological cartography
Mahalanobis distance
Pedology
Pedometrics
Digital soil mapping
Soil survey
Geomorphology
Tropical soils

ABSTRACT

Multivariate split moving window (MSMW) is a tool to automate soil mapping and to assess uncertainty in soil boundaries. In this paper, we propose a new approach to locate soil boundaries. We investigated the potential of (i) MSMW associated with Mahalanobis D^2 and (ii) split moving window (SMW) associated with magnetic susceptibility (MS) as tools for validation of soil boundaries. A transect was lined in Guatapará city (Sao Paulo state, Brazil), and 172 soil samples were collected from 86 locations, at a depth of 0–25 and 25–50 cm, for physical and chemical analyses. Additionally, categorical properties – geology, land use, and altitude – were assessed at the same sampling points. All the data were organized into four groups of properties (G1, G2, G3, and G4) and analyzed by principal component analysis, MSMW analysis – to delineate map units using D^2 –, and SMW analysis using MS. MSMW and SMW were compared regarding their potential to locate soil boundaries along the transect. The MS peaks in SMW presented a correlation with peaks of t-statistics and D^2 in MSMW: ($r = 0.56$; $p < 0.01$ – $r = 0.69$; $p < 0.01$) and ($r = 0.55$; $p < 0.01$ – $r = 0.64$; $p < 0.01$) for both depth intervals (0–25 cm and 25–50 cm), respectively. MSMW was more sensitive than SMW in the detection of soil boundaries in areas with lower clay content. Compared to conventional soil surveys, MSMW considerably improved the prediction of boundaries in mapping units. We concluded that the proposed method is a promising strategy for soil surveyors and can be used to assist pedological cartography.

1. Introduction

The delineation of the boundaries of a soil map unit is one of the main difficulties during soil survey for accurate soil maps. Due to the subjectivity or uncertainty of traditional methods in soil survey, there has been a great demand for training in quantitative methods. It is a requirement for professionals in soil and land resources survey so that more multidisciplinary teams can support operations of predictive modelling in a digital soil mapping (DSM) approach (Bui et al., 2020).

Conventional soil survey is the delineation of soil boundaries following a knowledge-based soil-landscape empirical model (Hudson, 1992). When conventional methods are combined with quantitative methods (e.g. geostatistical and multivariate), they are called hybrid

techniques for soil survey and allocation into pre-existing soils classes (McBratney et al., 2000). Despite their advances, current soil survey maps contain many soil complexes, rather than consociations. Base maps and resources available today are limitations to delineate soil boundaries on a detailed scale (Miller and Schaetzl, 2015).

To obtain knowledge about some limitations of traditional soil surveys, studies have applied knowledge-based techniques and fuzzy logic concepts as a predictive approach, for instance, the software SoLIM (Zhu et al., 2003) and automated predictive maps (MacMillan et al., 2007). Predictive soil mapping approaches are an alternative to locate soil boundaries. They are also called pedometrics (Webster, 1994; McBratney et al., 2003) in reference to quantitative research in the field of pedology. One of those approaches is the soil numerical classification

* Corresponding author.

E-mail address: joao.fernandes@ufra.edu.br (J.F. Silva Júnior).

using fuzzy logic (McBratney and Odeh, 1997; McBratney et al., 2000; Hughes et al., 2014) and taxonomic distances (Minasny and McBratney, 2007; Rossiter et al., 2017; Massawe et al., 2018). As a rule, these approaches use soil and landform information as a basis for classifying landform units (Iticha and Takele, 2018).

The association of quantitative methods with landform models is recommended to mitigate uncertainty to a satisfactory degree – e.g., split moving window analysis (SMW), which can be applied at a global scale. Siqueira et al. (2015) used such approach for a detailed mapping unit design based on landscape.

Several pedologists have rationalized and focused their thinking on a search for quantitative explanations in the delineation of soil boundaries. The taxonomic distance can be incorporated into a numerical classification routine for accuracy assessment of soil classes using the following criteria: a regional correlator's opinion, distance between classes in a numerical taxonomy assessment, distance within taxonomic hierarchy, and an error loss function (Minasny and McBratney, 2007; Rossiter et al., 2017). The OSACA algorithm had the lowest uncertainty in the allocation of soil profiles or creation of a new classification of soil profiles (Carré and Jacobson, 2009). The soil–landscape relationship has helped to categorize highly variable soils, and they are the basis for delineating soil mapping units (Iticha and Takele, 2018). The use of geostatistical analyses to delineate mapping units based on the spatial uncertainty of magnetic susceptibility (MS) and clay content allows the identification of appropriate regions using the modal pedon (Teixeira et al., 2018).

Due to the subjectivity of delineating soil mapping units, some authors (e.g. McBratney et al., 2000; Silva Júnior et al., 2012; Hughes et al., 2018) have proposed the use of mathematical tools to develop methods of soil classification, and these new approaches have stimulated the application of quantitative techniques in soil science (McBratney et al., 2003; Beaudette et al., 2013; Rossiter, 2018). Consequently, soil scientists have investigated the possibility of applying these new technologies and covariates to soil mapping to identify and accurately locate boundaries between units containing different soil properties.

Quantitative topofunctions have been recognized as essential tools in soil survey and digital soil mapping (Minasny and McBratney, 2016; Zhang et al., 2017). For example, Barrios et al. (2012), Siqueira et al. (2015), and Ramos et al. (2017) found a strong relationship between the soil management unit defined by SMW and boundaries identified based on Dalrymple et al.'s (1968) landscape models using MS and physical-chemical soil attributes. Siqueira et al. (2015) and Ramos et al. (2017) found that landform models associated with field observations of transects are efficient to delineate soil mapping units.

All methods described above take a univariate approach and map soil properties individually. It is a limitation if we consider Norris's assumptions, who suggest the application of a multivariate analysis using a limited and defined range of different soil properties in soil survey to delineate soil mapping units, instead of using only an univariate approach (Norris, 1971).

In order to increase the accuracy of spatial information, one can use numerical classification tools, such as SMW, but it requires a large number of samples, which precludes its implementation in research and land use planning. For that reason, another approach, called multivariate split moving window (MSMW), has recently been tested to detect soil map unit boundaries for multivariate data, using vectors of principal components (PCs) to estimate the Mahalanobis distance (D^2) (Webster, 1978; Rossiter, 2012).

D^2 can be used in soil mapping, classification and pedological modelling due to large scale differences of input variables. This distance has been tested for pedological distance between soil profiles, and its great potential has been demonstrated in applications such as digital soil mapping and numerical classification (Carré and Jacobson, 2009). Compared to the methods mentioned previously, MSMW using D^2 is an innovation because it delineates soil mapping units using PC scores from

the linear combination of optimally-weighted variables. This is an interesting linear transformation technique that helps soil mappers to understand how many data are really significant.

Another difficulty in soil survey is to identify suitable covariates for predicting soil properties. These covariates should be easily determined, non-invasive, accurate, and economically feasible. Some studies have indicated MS as an indirect quantification alternative (Siqueira et al., 2010; Barrios et al., 2012; Matias et al., 2013; Siqueira et al., 2015). In this line, Ramos et al. (2017) assessed the use of SMW based on MS values and reported its effectiveness in outlining representative compartments of pedogenic environments in Southern Brazil. Siqueira et al. (2015) employed a SMW analysis using MS based on the soil-landform relationship to design detailed mapping units, but they suggested that a MSMW analysis would be more effective. Based on this suggestion by Siqueira et al. (2015) and the use of D^2 recommended by Carré and Jacobson (2009), we hypothesized that MSMW can be associated with D^2 as a hybrid multivariate method to delineate soil mapping units.

In this study, we compared the quantitative techniques MSMW and SMW using MS to locate soil boundaries along the landscape. We hypothesized that MSMW and SMW are effective and complementary techniques to locate soil boundaries. When used with a topographic model, these techniques may help mappers to better place soil mapping unit boundaries at locations where defendable landscape breaks exist.

This study had two goals: (i) to investigate whether MSMW can be applied consistently and objectively through hillslope positions as an improved tool to delineate soil boundaries and (ii) to investigate whether MSMW is more robust than SMW in locating soil boundaries.

2. Material and methods

2.1. Description of the area and sampling

The study area had 380 ha and was located in the city of Guatapará, a small town in the northeasters of the state of São Paulo, Brazil. The geographical coordinates of the central position of the area are $21^\circ 28' N$ and $48^\circ 01' W$, and the maximum altitude is of 600 m (Fig. 1A, B). The area is located in the geomorphic province of the western plateau of the state of São Paulo, where the source material is related to the basalt transition, São Bento group, Serra Geral formation, colluvial-eluvial and alluvial deposits (IPT, 1981).

The local soils are classified as Oxisols in mapping units: eutroferric Red Latosol (LVef), dystroferric Red Latosol (LVdf), dystrophic Red Latosol (LVd), and dystrophic Red-Yellow Latosol (LVAd) (Fig. 1) according to the Brazilian System of Soil Classification (SiBCS) (Santos et al., 2018) and Soil Taxonomy (Soil Survey Staff, 2014). According to Thornthwaite (1948), the local climate can be defined as B1rB'4a', which means a humid mesothermal climate with small water deficit and summer evapotranspiration lower than 48% of the annual evapotranspiration.

The area has been used for sugarcane cultivation with mechanized harvesting for more than 10 years, and the varieties grown are SP87-365, SP83-2847, and RB 925345. Based on the local soil classes, climate, and field-grown sugarcane varieties, three land uses were identified: A – expected productivity >95 tons of stalks ha^{-1} ; B – expected productivity between 90 and 95 tons of stalks ha^{-1} ; and C – expected productivity between 85 and 89 tons of stalks ha^{-1} .

In the field work, a planialtimetric survey provided a base-map with an overview of the local topography. A digital elevation model (DEM) was used along with data from Shuttle Radar Topography Mission's active sensor (SRTM), with horizontal resolution of 90 m and landform. Next, a 2508-m slope was identified from the hilltop, across the crest to the foothills.

A transect was lined along the crest of the terrain, on which 86 sampling points were set at regular intervals of 30 m (Fig. 1C). The field, along the transect, was divided in units based on Dalrymple et al.'s (1968) conceptual model (Fig. 1C). Soil samples were collected from two

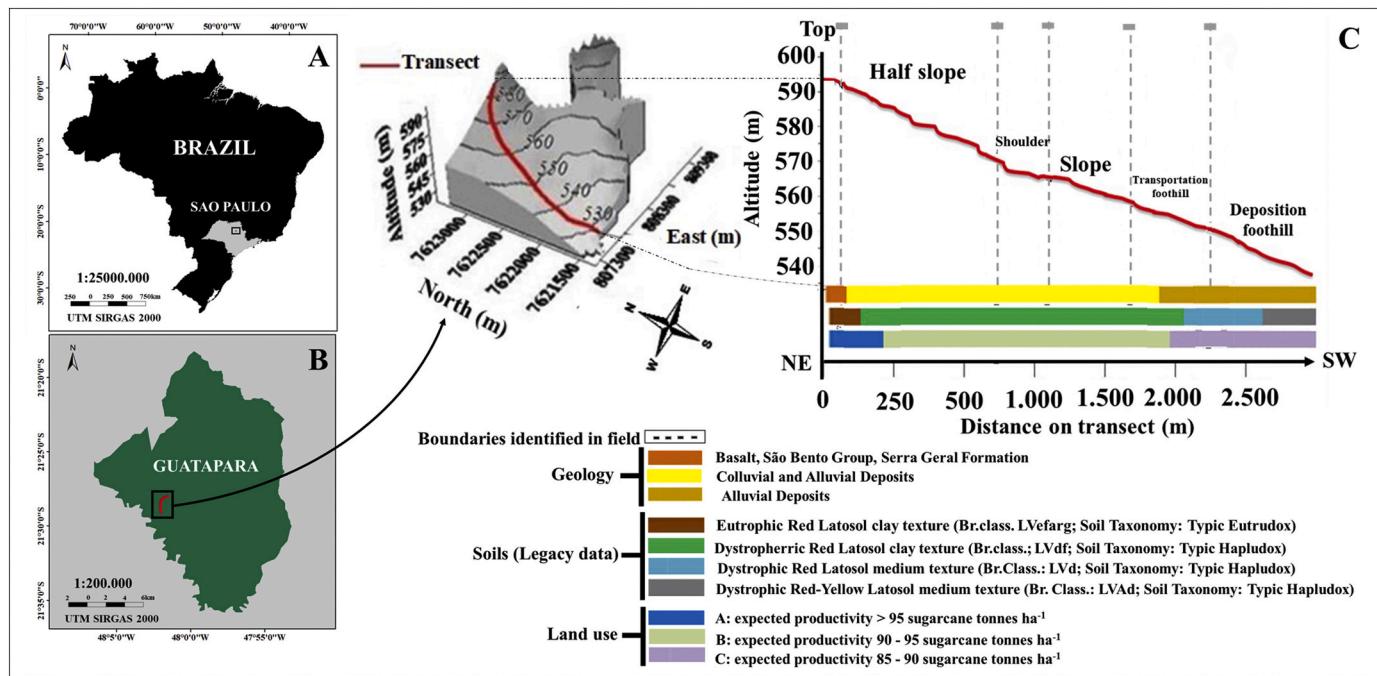


Fig. 1. Characterization of the study area. Location of the sampling area along the transect (A) and (B); digital elevation model and geomorphologic units based on Dalrymple et al. (1968) (C).

depth ranges (0–25 cm and 25–50 cm) – totaling 172 samples –, which are currently used for soil management in the sugar-alcohol industry of the state of São Paulo.

2.2. Laboratory analyses

Soil samples were analyzed for texture by the pipette method, using a 0.1 mol L⁻¹ NaOH solution as dispersing agent and a mechanical stirring apparatus at low speed for 16 h, following the method recommended by Embrapa (1997). The sum of bases (SB), cation exchange capacity (CEC), and base saturation (BS%) were calculated using the content of the bases: Ca²⁺, Mg²⁺, K⁺, and H + Al.

Then, air-dried fine earth (ADFE) fractions of the 172 soil samples were used to determine MS values. The amount of MS per mass unit was determined at low frequency (0.47 kHz), using a Bartington sensor (MS2 System) (Dearing, 1999; Costa et al., 1999). This frequency is suggested for precise results (Dearing, 1999).

Munsell color codes were obtained for each soil sample using a spectrophotometer with an 80-mm integrating sphere, each 1 nm, using a Lambda 950 device for evaluation by diffuse reflectance spectrum. Next, the XYZ tristimulus values were determined according to the International Commission on Illumination (CIE) (Wyszecki and Stiles, 1982), after the Munsell parameters hue, value, and chroma were deducted by Munsell Conversion, version 6.4, according to Viscarra Rossel et al. (2010).

2.3. Data analysis

2.3.1. Principal Component Analysis (PCA) using groups of properties

PCA was applied to a data set to extract interrelationships between several variables and find the best linear combination of individual variables to separate groups. We followed these steps: 1) calculate correlation between properties; 2) standardize data with a mean of zero and variance of one; 3) calculate the covariance matrix for all data using a correlation matrix; 4) calculate eigenvectors and eigenvalues from the covariance matrix to identify PCs, which were extracted and adjusted using varimax rotation (Kaiser, 1958); 5) identify their importance level,

discarding those of minor importance; 6) choose the PCs to be used in the MSMW analysis with eigenvalue ≥ 1 . These PCs were considered significant, since they explained the overall variability and reduced the size and collinearity of the original properties (Kaiser, 1960). PCA was performed using the R software (R Development Core Team, 2018).

2.3.2. Autocorrelation Function (ACF) using groups of properties and MS

Initially, changes of properties along the transect were analyzed by the autocorrelation method (Webster and Cuanalo, 1975). The autocorrelation function (ACF) reveals how the correlation between any two values of the signal changes as the distance between them changes. ACF measures the maximum range of autocorrelation and reveals information about the window width that one can use in the SMS and MSMW analyses. The plot of ACF with respect to distance is called an autocorrelogram. In this study, the autocorrelation values (window width) shows a high correlation between (X_i) and (X_{i+h}) , and covariance decreases gradually until it became zero. No correlation or spatial dependence exists between these two series considering a confidence limit (-0.2 and 0.2), which represents a confidence interval (CI = 95%). Eq. (1) was used to calculate the autocorrelation function (Davis, 1973):

$$r(h) = \frac{[(n-h)(\sum X_{i+h}X_i) - \sum(X_{i+h})\sum(X_i)]/(n-h)(n-h-1)}{\left[n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2\right]/n(n-1)} \quad (1)$$

where $r(h)$ is the autocorrelation function value and n is the number of observations.

If the autocorrelogram shows a high correlation between (X_i) and (X_{i+h}) , the observations are dependent.

An ACF value near to zero indicates that the process is completely uncorrelated (Nounou and Bakshi, 1999), that is, this position indicates the approximate boundaries of the soil mapping unit in the landscape.

2.3.3. SMW analysis using MS

After identifying the landform component boundaries in the field using the conceptual model proposed by Dalrymple et al. (1968), the SMW analysis was performed to validate these map units (Ludwig and Cornelius, 1987; Cornelius and Reynolds, 1991; Siqueira et al., 2015).

First, we divided the 172 samples with MS values in two parts, called windows, with their width defined by ACF. Then, these two windows were moved along the transect. At each new position, the window was divided into two parts (or two windows), and the mean of the two windows was calculated and compared. These comparisons were performed using the Student's *t*-test.

The highest peaks along the transect were expected to correspond to the mathematical limits between two areas with distinct variations in MS.

The Student's *t*-statistic was defined by Eq. (2):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2)$$

where: \bar{x}_1 and \bar{x}_2 are the sample means of two windows (n_1 and n_2); S_p is a combined estimate of the standard deviation of the windows by Eq. (3):

$$S_p = \frac{(n-1)S_1 + (n_2-1)S_2}{n_1 + n_2 - 2} \quad (3)$$

where: S_1 , S_2 , n_1 , and n_2 are standard deviations and sample sizes of the two windows previously established, respectively.

MS was assessed using the SMW analysis at both depth ranges (0–25 cm and 25–50 cm), totaling 172 observations. We applied the Mullion index equal to zero for major discrimination of soil boundaries based on the results of autocorrelograms (Rossiter, 2012).

2.3.4. MSMW using groups of properties

The MSMW analysis was developed by Webster and Wong (1969) to improve the interpretation of soil boundaries in air photographs. They state that the MSMW analysis computes PCs to reduce the dimensionality of data.

Both MSMW and SMW methods are based on the concept of two moving windows and an interval between them, which is called Mullion index. An advantage of MSMW is the possibility that include the number major of covariates assessed, the test of D^2 as a discriminant of soil boundaries, and the use of PCA vectors for covariates.

However, while SMW uses the *t*-statistics of MS values of soil to delineate boundaries (Siqueira et al., 2015), MSMW applies the D^2 of PC scores from PCA, because there is a connection between the Mahalanobis distance and principal components analysis. According to Brereton (2015), D^2 is equal to the sum of squares of all non-zero standardized PCs scores.

The Mahalanobis distance was first proposed by the Indian statistician P.C. Mahanobis in 1936. Traditional univariate statistics usually calculate the number of standard deviations an observation is from the center of a dataset, and they use this value to determine various statistics about it. Extending to the multivariate situation, Mahalanobis proposed a distance D^2 from the center of the data.

The detailed description of the Mahalanobis distance acquisition can be seen in Gath and Hayes (2006). In this study, we will present a generalization for the Mahalanobis distance in the special case of PCs. D^2 provides a robust diagnosis when considering several properties along the transect (Fig. 4). In this study, PCs scores were used to calculate D^2 at any window position, and the traditional mathematical definition was given by Eq. (4):

$$D^2 = (\bar{X}_1 - \bar{X}_2)^T W^{-1} (\bar{X}_1 - \bar{X}_2) \quad (4)$$

where: \bar{X}_1 and \bar{X}_2 are the mean vectors of PC scores of the mean of two windows, and W is the variance-covariance matrix of the mean of windows.

The total window width used to calculate D^2 was two-thirds of the average value of autocorrelogram results of the first PCs obtained for the four groups. We measured the window width based on the following

criteria: the window width was the smallest of PCs between lower confidence limit (-0.2) and upper confidence limit (0.2) (confidence interval, CI = 95%); and the maximum window width should be positive and smaller than 1/2 of the transect length, i.e. (86/2 = 43). In general, this range was defined when autocorrelation coefficients ceased to decay, showing a minimal dependence and positioned close to zero (Webster and Cuanalo, 1975; Webster, 1978). Also, Mullion values of zero points were used in the four groups.

We proposed a data clustering methodology to help soil mappers delineate soil mapping units and calibrate soil legacy maps by an association between statistic tests and tacit knowledge related to soil formation factors. Many clustering techniques identify natural groupings (Omran et al., 2007). Omran et al. (2007) presented an overview on the key clustering methods and approached clustering techniques and algorithms that automatically determine the optimum number of clusters and simultaneously cluster the data set with minimal user interference.

Based on this proposal, we presented the MSMW analysis, which is a non-parametric method to assess how PCs change along transects. Overlapped moving windows followed by PCA enable the extraction of data from groups of soil properties and the application of these data to detect the boundaries of a hillslope position by means of a one-dimensional segmentation. These segments are usually termed moving windows (MWs), since they are understood as a range of points that move along the scan properties by incorporating new points at the window front and dropping off old ones at the rear end.

In this study, we reduced the subsets of original covariates to two dimensions represented by the first PCs. We used standardized subsets determined by auto values >1.0, because these are significant values and capture most of variation according to Kaiser (1960). After calculating the differences between subsets, two main criteria were used to identify soil boundaries: location along the transect and the highest differences between MWs. Groups could then be identified in the graph by significant peaks and the distance between windows along the transect. This study considered the following properties: the land use of sugarcane, altitude, and internal factors of soil—texture, color, geology, chemical attributes and MS. All data were normalized (standardized) to a mean of zero and a standard deviation of one.

The data used for PCA were organized into four groups according to the set of properties evaluated: G1: base saturation (BS%), clay content, hue, value, and chroma at the depth range of 0–25 cm, altitude, E = east, N = north in UTM; G2: base saturation (BS%), clay content (depth 25–50 cm), altitude, E = east, N = north in UTM; G3: base saturation (BS %), clay content, hue, value, chroma (depth 0–25 cm), base saturation (BS%), clay content (depth 25–50 cm), altitude, E = east, N = north in UTM; and G4: base saturation (BS%), clay content, hue, value, chroma (depth 0–25 cm), base saturation (BS%), clay content (depth 25–50 cm), altitude, land use, geology, E = east, N = north in UTM.

The above-described groups were organized according to the approach proposed by Webster and Wong (1969), who employed a technique in multivariate statistics to delineate soil boundaries. This technique transforms a set of correlated variates – as clay content of surface horizon, clay content of second horizon, depth to mottling – into a new set of mutually uncorrelated variates (PCs). We hypothesize that the dissimilarity between MWs plotted in the lines, or points, may be used as standards to compare with intuitively recognized boundaries.

2.3.5. Comparison of results based on *t*-statistics and the D^2

We compared the results of the MSMW analysis using Pearson's linear correlation coefficients (r) (Pearson, 1920). We assessed the linear relationship between *t*-statistics and D^2 by measuring the intensity (weak or strong) and direction of positive or negative correlations. The calculation of r is defined in Eq. (5).

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[\sum x^2 - (\sum x)^2] \ln \sum y - (\sum y)^2}} \quad (5)$$

where: y is values on the vertical axis; and x is values on the horizontal axis; n is number of periods observed; r is the correlation coefficient.

The statistical BIAS is a feature of a statistical technique (mean of the residuals) showing the average deviation of the variable (D^2) of the model in relation to a variable (t-statistics). It is expressed by the difference between (P_n) and (O_n) as shown in Eq. (6):

$$BIAS = \frac{1}{N} \sum_{n=1}^N (P_n - O_n) \quad (6)$$

where: P_n is the value predictor, and O_n is the value of reference or value observed. The nearer to zero, the better the result.

The statistics used to test the model quality included Lin's concordance correlation coefficient (LCCC) (Lin, 1989) in R (3.2.2). The

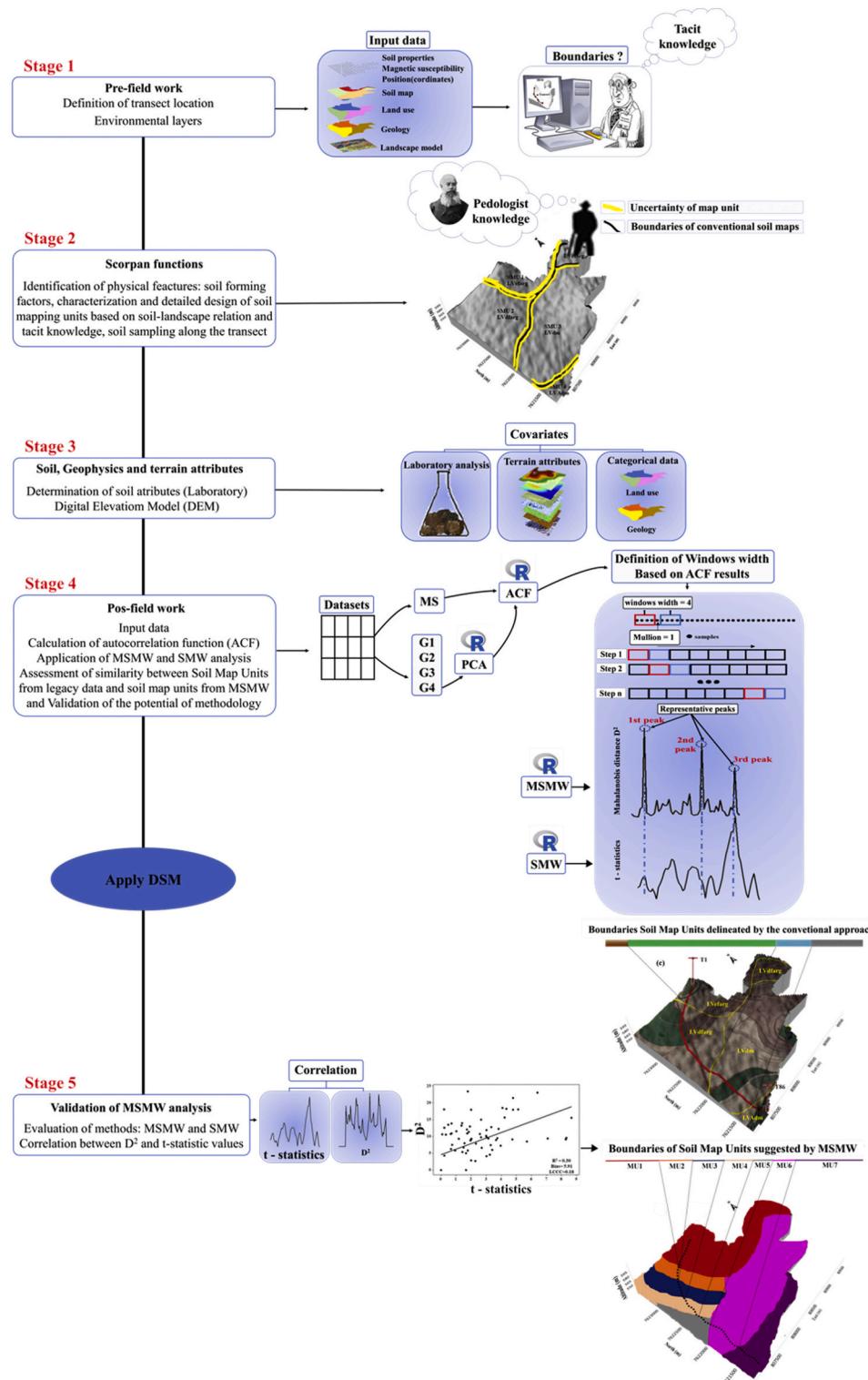


Fig. 2. Flowchart of the integrated method.

formula is shown in Eq. (7):

$$LCCC = \frac{2\pi\partial_x\partial_y}{\partial_x^2 + \partial_y^2 + (\bar{x} + \bar{y})^2} \quad (7)$$

where: x and y represent the predicted value and measured value at point i ; \bar{x} and \bar{y} are the mean of predicted and measured values; ∂_x and ∂_y represent the variance of predicted and measured values.

Our results were presented in graphs, where peaks of t-statistics and D^2 represented the sites where significant segmentation of landform occurred. However, if t-statistics and D^2 values were similar to the units of landform in the conceptual model by Dalrymple et al. (1968) (Fig. 4), the hypothesis is that MSMW can be used to delineate soil boundaries with any of the four groups of D^2 values using PC scores, being better than peaks of t-statistics with MS found by Siqueira et al. (2015). This proposal is illustrated in the flowchart shown in Fig. 2.

3. Results

3.1. Correlation and PCA

Table 1 shows that PCA extracted the best properties that can be used to delineate soil boundaries. The highest cumulative explained variance using PCs was found in each group as follows: G3 (88.57%) > G4 (87.49%) > G1 (81.77%) > G2 (70.00%). This result suggests that the groups' covariates have higher representative PC loadings between groups by correlation.

In G1, altitude, Coord_E, Coord_N, and clay⁽¹⁾ had high loadings of PC1. These covariates (altitude, Coord_E, Coord_N, and clay⁽¹⁾) were negatively correlated with PC1, while chroma was negatively correlated with PC2. Hue was positively correlated with PC2. PC3 had a positive correlation with value only.

G2 was represented only by PC1. Due to the high loadings of PC1, altitude, Coord_E, Coord_N, and clay⁽²⁾ were negatively correlated, indicating their strong association with PC1.

G3 was represented by four PCs – altitude, Coord_E, Coord_N and clay^(1,2) – due to the high loadings of PC1. These covariates were negatively correlated with PC1. Hue was negatively correlated with

PC2, and chroma was positively correlated with PC2. In PC3, correlations were low, and, in PC4, the covariate value was positively correlated.

G4 was represented by covariates more important: altitude, Coord_E, Coord_N, and clay^(1,2), land use and geology due to the high loadings of PC1. These covariates were negatively correlated with PC1, except land use and geology, which were positively correlated. Hue and BS%⁽²⁾ were positively correlated with PC2, and chroma was negatively correlated. In PC3, correlations were low, and, in PC4, the covariate value was negatively correlated. All correlations described above were considered as strong associations.

In all four groups, the sum of PCs represents more than 70% of the explained variance, indicating that PCs had a higher cumulative explained variance. In the MSMW analysis with D^2 , PCs revealed that this technique was more efficient than the SMW peak analysis to delineate soil boundaries along the transect on the geomorphologic unit half slope (Fig. 4).

We believe that there exists a similarity to regional approaches for most local geomorphologic units outlined in the Dalrymple et al. (1968) model and the peaks of G1, G2, G3, and G4 (see Fig. 4), except in the transition top/half-slope. To better quantify the relationships between variables and identify covariates, we observed that clay content at depths 0–25 cm and 25–50 cm, altitude, coordinates E and N, land use, geology were the most important properties, as indicated by the correlation between PCs and properties between the four groups. The soil color properties had secondary importance, as indicated by PC2 and PC3. All sets of properties have a high discriminating power in delineating soil boundaries because of the high explained variance, above 70% (Table 1).

Therefore, the PCA analysis indicated the most important covariates to detect of soil boundaries using MSMW.

3.2. ACF

Fig. 3 shows that autocorrelation decreased to close to zero in a window of 23 at 690 m in G1, which was the smallest window width of the three PCs and confidence limit (−0.2 and 0.2) in blue lines. G1

Table 1
Correlation coefficients between soil properties and each of the principal components (PCs); scores from the four groups and soil properties.

Principal components	(Group 1)			(Group 2)			(Group 3)			(Group 4)		
	PC1	PC2	PC3	PC1	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
Eigenvalues	3.64	1.87	1.06	3.50	4.56	2.15	1.09	1.06	5.86	2.47	1.10	1.07
Cumulative variance (%)		81.77		70.00		88.57				87.49		
Explained variance (%)	45.50	22.97	13.30	70.00	45.57	21.50	10.88	10.63	48.81	20.60	9.17	8.90
Correlations												
BS% ⁽¹⁾	0.57	0.43	-0.07	-	0.61	-0.52	0.47	-0.14	0.50	0.62	-0.51	0.03
Clay ⁽¹⁾	-0.82	-0.09	0.04	-	-0.85	-0.10	0.37	-0.19	-0.82	-0.13	-0.44	0.10
Hue ⁽¹⁾	0.22	0.86	-0.04	-	0.30	-0.74	-0.40	0.18	0.14	0.78	0.38	-0.06
Value ⁽¹⁾	0.07	0.24	-0.94	-	0.10	-0.23	0.27	0.90	0.04	0.26	-0.05	-0.93
Chroma ⁽¹⁾	-0.18	-0.82	-0.39	-	-0.24	0.74	0.40	0.28	-0.05	-0.78	-0.28	-0.39
BS% ⁽²⁾	-	-		0.38	0.50	-0.63	0.44	-0.15	0.37	0.71	-0.48	0.04
Clay ⁽²⁾	-	-		-0.82	-0.85	-0.16	0.33	-0.17	-0.84	-0.07	-0.40	0.10
Altitude	-0.95	0.26	-0.01	-0.98	-0.91	-0.34	-0.09	0.06	-0.97	0.14	0.06	-0.02
Coord_E	-0.88	0.24	-0.13	-0.91	-0.84	-0.30	-0.09	0.19	-0.88	0.10	0.09	-0.15
Coord_N	-0.93	0.23	0.04	-0.95	-0.88	-0.31	-0.12	0.02	-0.95	0.12	0.10	0.01
Land use	-	-	-	-	-	-	-	-	0.89	-0.27	-0.01	-0.02
Geology	-	-	-	-	-	-	-	-	0.82	-0.39	-0.11	0.01
	Clay ⁽¹⁾⁺	Color	Color	Clay ⁽²⁾⁺	Clay ⁽²⁾⁽¹⁾⁺	Color	Color	Color	Clay ⁽¹⁾⁽²⁾⁺	Color	Color	
	Altitude+			Altitude+	Altitude+				Altitude+			
	Position			Position	Position				Position			
									Land use+			
									Geology			

⁽¹⁾ and ⁽²⁾ are the depth ranges 0–25 cm and 25–50 cm, respectively. G1: Base saturation (BS%), clay content, hue, value, and chroma at the depth range of 0–25 cm, altitude, E = east, N = north in UTM; G2: Base saturation (BS%), clay content (depth 25–50 cm), altitude, E = east, N = north in UTM; G3: Base saturation (BS%), clay content, hue, value, chroma (depth 0–25 cm), base saturation (BS%), clay content (depth 25–50 cm), altitude, E = east, N = north in UTM; and G4: Base saturation (BS%), clay content, hue, value, chroma (depth 0–25 cm), base saturation (BS%), clay content (depth 25–50 cm), altitude, land use, geology, E = east, N = north in UTM. (Loadings ≥ 0.70 are printed with bold characters.), Land use.

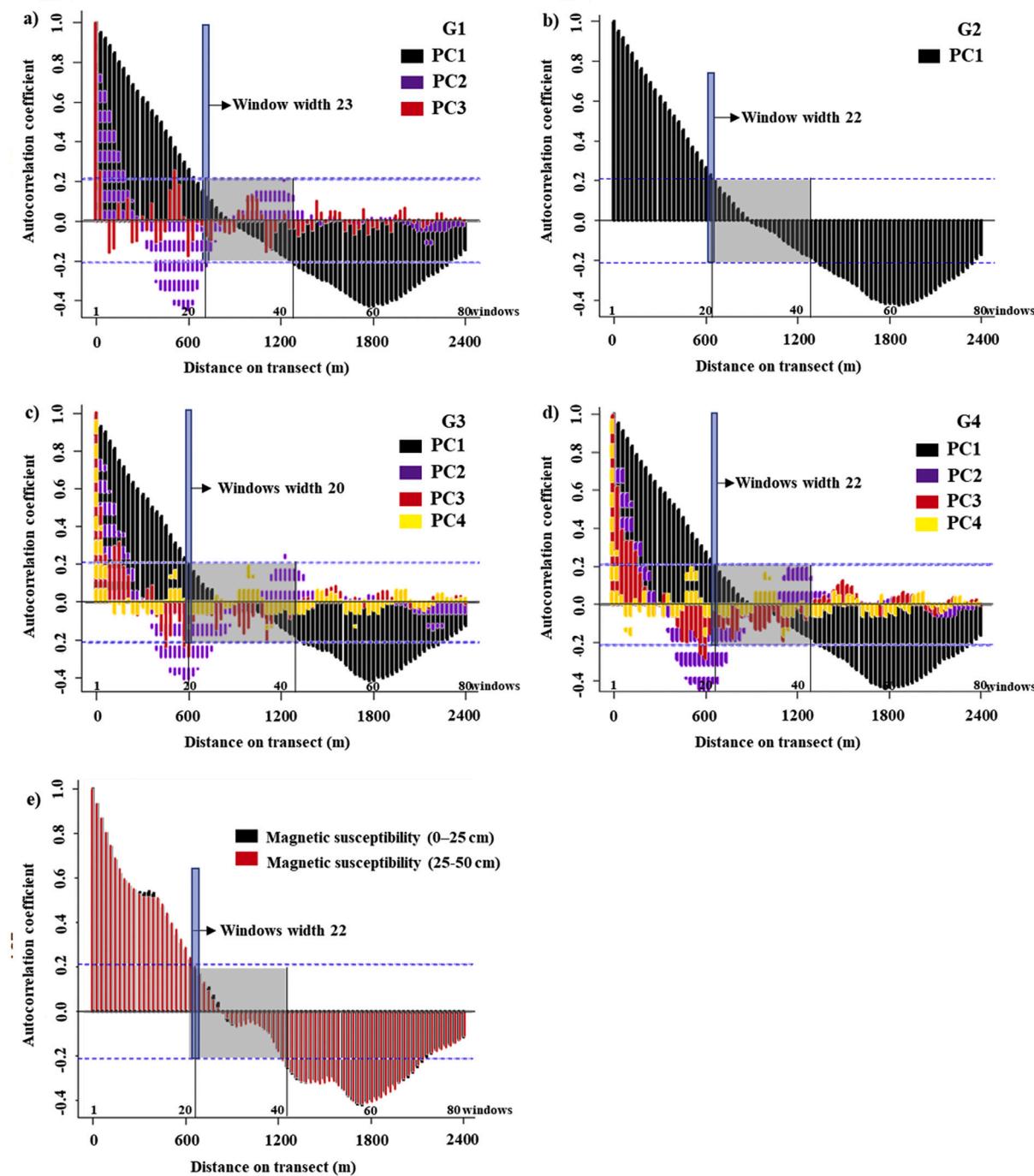


Fig. 3. Correlograms of the principal components (PCs): G1: BS%, clay, hue, value, and chroma (0–25 cm); G2: BS%, clay (25–50 cm), and altitude; G3: BS%, clay, hue, value, and chroma (0–25 cm); G4: soil classes, land use, and geology; MS: magnetic susceptibility (0–25 cm and 25–50 cm).

suggests a soil boundary at a distance of 690 m. Autocorrelation in G2 decreased to close to zero, with a minimum window width of PC1 of 22, suggesting a soil boundary at a distance of 660 m. This window width of 22 should be used with the group properties. G3 lost its significant autocorrelation with a stationarity near zero, indicating the use of a window width of 20, suggesting a soil boundary at 600 m. G4 presented similar results to those of G2, suggesting a soil boundary at a distance of 660 m along the transect.

The autocorrelation of MS at depths of 0–25 cm and 25–50 cm reached close to zero in a window of 22 considering the confidence limit (-0.2 and 0.2) in blue lines. It suggests a soil boundary at a distance of 660 m.

3.3. Detection of boundaries with SMW using *t*-statistics by MS

Fig. 4 shows that MS indicated five different soil boundaries based on its respective peaks: 1st, 2nd, 3rd, 4th, and 5th. These boundaries are consistent with the different hillslope positions defined by the [Dalrymple et al. \(1968\)](#) model: top, half-slope, shoulder, slope, transportation foothill, and deposition foothill. The SMW model suggested boundaries in the landscape separated into six series of soils, showing a complete distinction between the five soil boundaries.

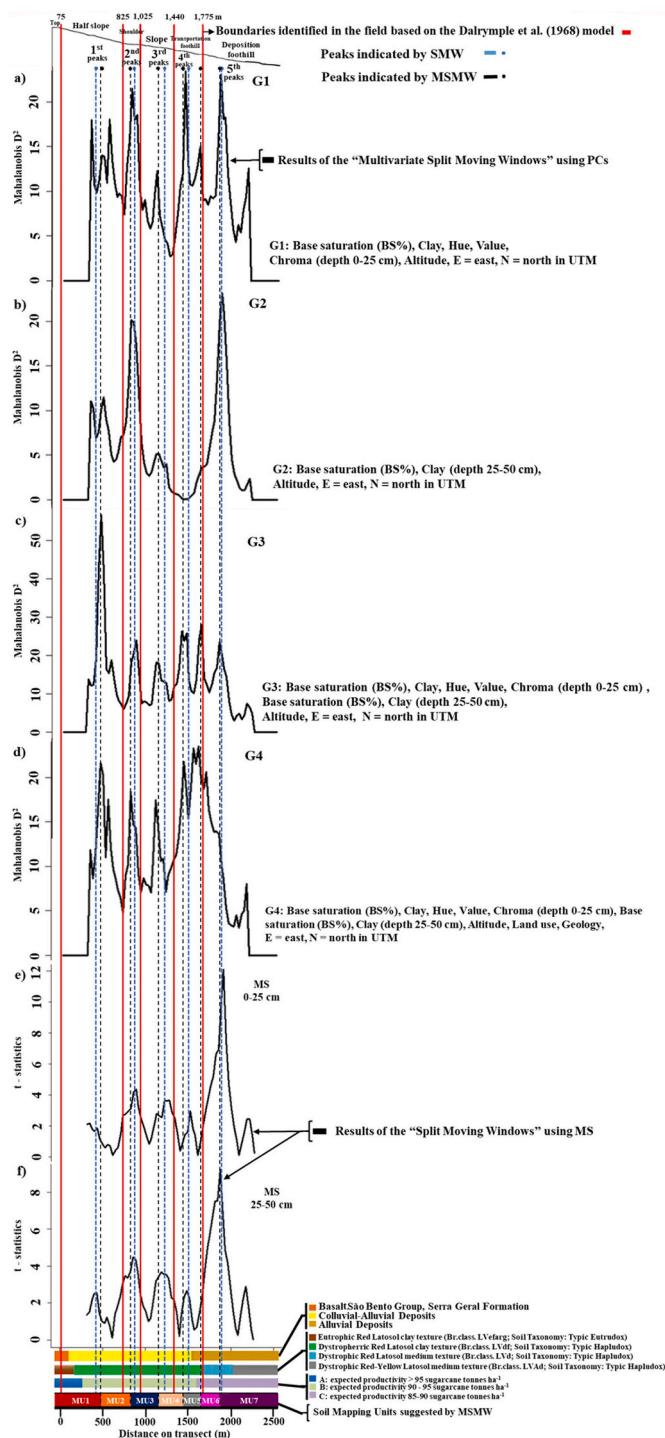


Fig. 4. Results of the Multivariate Split Moving Windows (MSMW) and Split Moving Windows (SMW) analysis and similarity among t-statistics values and Mahalanobis distance D^2 for the groups G1, G2, G3, and G4 and magnetic susceptibility (MS), along the transect, with 30-m spaced points.

3.4. Detection of boundaries with MSMW using D^2

The geomorphologic units in this study were identified in the field as top/half-slope at 75 m, half-slope/shoulder at 825 m, shoulder/slope at 1025 m, slope/transportation foothill at 1440 m, and transportation foothill/deposition foothill at 1775 m, according to the conceptual model by Dalrymple et al. (1968). Fig. 4 shows that all six boundaries identified in the MSMW analysis were similar to those defined by SMW

along the transect, except for G2 at transportation foothill. The 1st, 2nd, 3rd, 4th, and 5th peaks of D^2 close to the geomorphological units of the Dalrymple et al. (1968) model were almost constant along the transect (Fig. 4).

The peaks of G1 and G2 had similar patterns, except for transportation foothill, which had a smooth curve. G3 stood out for presenting the highest explained variance (about 88.57%) in PCA. This result suggests that the boundaries of the soil series were clear and similar to those found by MS (Fig. 4c). The boundaries defined by MSMW in G3 and G4 were similar, so this method allows soil boundary delineation and easy interpretation of peaks.

We observed a similarity in the delineation of soil boundaries between the results obtained by SMW, MSMW, and the geomorphologic units of the conceptual model of Dalrymple et al. (1968) (Fig. 4). There is a movement to the right of 1st, 2nd, 3rd, 4th, and 5th peaks of D^2 and t-statistics in relation to the geomorphologic units in red lines.

The MSMW technique suggests possible new boundaries of mapping units, the fact that representative peaks are located at 481 m and 1891 m along transect. Non-parametric methods are more robust and, when combined with spatial variability, t-statistics, and D^2 , they become a powerful tool for identifying boundaries in the soil-landscape relationship.

3.5. Assessment of similarity between D^2 and t-statistic values

We confirmed our hypothesis that the MSMW technique is complementary to SMW based on Pearson's linear correlation coefficients between D^2 and t-statistics. Table 2 compares t-statistics and D^2 values and shows they had similar results. The coefficient r ranged from 0.36 to 0.69 and from 0.38 to 0.64 in the depth ranges 0–25 cm and 25–50 cm, respectively. BIAS ranged from 2.77 to 6.86 and from 2.68 to 9.08 in the depth ranges 0–25 cm and 25–50 cm, respectively. LCCC ranged from 0.08 to 0.40 and from 0.08 to 0.35 in the depth ranges 0–25 cm and 25–50 cm, respectively.

G2 showed that MSMW is a good technique, as its results were the most similar to those found in SMW, with good agreement according to the following values: $r = 0.69$, $r = 0.69$; BIAS = 2.77, BIAS = 2.68; and LCCC = 0.40, LCCC = 0.35 in the depth ranges 0–25 cm and 25–50 cm, respectively.

3.6. Comparison between boundaries by conventional and statistical methods

The results obtained by MSMW produced prominent peaks in the D^2 graph (Fig. 4). We tested the boundaries indicated by these peaks in legacy maps (Table 3) and found similar boundaries using the field survey technique, but the comparison was not exact. Most of the sharp boundaries recognized in the field were represented by prominent peaks in the D^2 graph. The boundaries obtained by MSMW were comparable in the initial region and areas of flat relief. This method suggested four new mapping units in the taxonomic class Dystroferric Red Latosol. The

Table 2

Pearson's correlation coefficient (r), Mean Bias Error (BIAS), Lin's Concordance Correlation Coefficient (LCCC) between the values of Mahalanobis distances D^2 in the MSMW analysis in the four groups and the t-statistic values in the SMW analysis using magnetic susceptibility (MS) in the depth ranges 0–25 cm and 25–50 cm.

D^2 by MSMW	t-statistics by SMW					
	MS 0–25 cm depth			MS 25–50 cm depth		
	r	Bias	LCCC	r	Bias	LCCC
G1	0.56	6.00	0.20	0.55	5.91	0.18
G2	0.69	2.77	0.40	0.64	2.68	0.35
G3	0.36	6.59	0.08	0.38	9.08	0.08
G4	0.44	6.86	0.14	0.51	6.77	0.15

Table 3

Evaluation of agreement between boundaries obtained by MSMW methods and those obtained by conventional methods along the transect.

MSMW Mapping units	Range (m)	Conventional methods		Range (m)
		Classification		
MU1	0-480	Eutroferric Red Latosol clay texture (LVef)		0-230
MU2	481-840			231--
MU3	841-1,140			---
MU4	1141-1470	Dystroferric Red Latosol clay texture (LVdf)		---
MU5	1471-1650			--1620
MU6	1651-1890	Dystrophic Red Latosol medium texture (LVd)		1621-1980
MU7	1891-2500	Dystrophic Red-Yellow Latosol medium texture (LVAd)		1981-2500

expressive peaks of D² suggested additional boundaries at 840, 1140, 1470 m, and 1650 m (Table 3).

4. Discussion

4.1. Assessment of covariate importance

PCs were controlled by clay⁽¹⁾, Coord_E, Coord_N, altitude, hue, chroma, and value. These variables were selected as appropriate covariates to delineate soil boundaries because they are more susceptible to environmental changes when there is an approach of the landscape and subjectivity of compartment nomination (soil boundaries) (Table 1). However, these soil properties and environmental covariates can be readily formulated using mathematical or statistical models for predicting soil properties and soil classes (Ma et al., 2019), in addition to diagnosing the soil type and pedogenic processes (e.g., soil erosion and clay illuviation). In this paper, the four groups presented representative, easily measurable covariates with a highly discriminating power.

Descriptive knowledge on soil-environmental relationships can also be used to delineate lateral boundaries, as all four groups in this study were representative, which explains more than 70.00% of the total variance. This may have several important implications. For example, the use of better covariates increases accuracy in soil mapping. In addition, PCA is cheaper and faster to delineate soil boundaries in the field. Therefore, it can help to understand covariates more clearly, as the use of legacy data can be hampered by a number of problems like unavailability of numeric data, lack of harmonization and imprecision of soil descriptions, imprecise georeferencing, non-optimal location of soil data as reported in the literature (McBratney et al., 2003; Lagacherie, 2008). Based on our results, we believe it is possible to develop models of local and regional pedogeomorphological evolution based on numerical classification to delineate soil boundaries in similar landscapes.

G3 was considered the best discriminating group compared to the other groups, with four PCs representing 88.57% of total variance. This can be explained by the influence of G3's main covariates (clay, position, altitude), which are soil-forming factors previously discussed by the literature (McBratney et al., 2003; Ma et al., 2019). In general, subsurface clay contents, position, landform morphology, and terrain elevation are deemed as the most relevant properties for the spatial distribution of soil units (Jafari et al., 2013).

G4 provided the second-best group of covariates, because of its high explained variance of 87.49% of total variance (Table 1). G4's covariates were correlated and considered important (altitude, Coord_E, Coord_N, clay^(1, 2), soil color, BS%, land use, and geology). However, these covariates can also be used to delineate soil boundaries in São Paulo due to their discriminating power in loading PCs. Based on these results, we confirmed the potential of PCA to help pedologists that are not experts in DSM to select the best covariates. Peng et al. (2020) concluded that selecting a proper set of covariates is one of the most important factors that influence the accuracy of DSM.

In Brazil, the lack of detailed maps (e.g. high resolution) of categorical environmental covariates may limit studies in this field. Carvalho Júnior et al. (2014), found that the lack of discriminating power of these covariates may be due to the absence of high-resolution, accurate covariates. Indeed, PCA's potential increases with the increase in the number of covariates in the group. In consequence, if the goal is to increase the power of PCA to apply in MSMW and improve accuracy in predicting soil boundaries, we recommend the use of DEM, with high resolution and accuracy, and a large dataset. Ngunjiri et al. (2019) concluded that input datasets with higher spatial resolutions, for example, DEM with higher spatial resolution, result in more detailed soil maps.

4.2. ACF

According to the autocorrelation coefficients (r) of PC1 and PC2 in G1, G2, G3, G4, and MS, measurements declined sharply in the beginning of the analysis until the mark of 600 m in the transect (Fig. 3). This indicates probable boundaries of soil map units. A previous work (Siqueira et al., 2015) also indicated that autocorrelograms can represent a similar pattern to mapping soil types where soil-landscape relationships are established based on terrain attributes. Autocorrelation coefficients (r) using PCs scores also improve the interpretation of soil-landscapes relationships, indicating positions where there is great variation in relief and attributes to quantitatively delineate soils with similar patterns. Webster and Cuanalo (1975) reported a similar pattern in autocorrelation graphs, which significantly decreased within shorter distances (Fig. 3). However, in our study, we used MS data and observations already collected from a previous application of SMW (Siqueira et al., 2015) as an initial reference based on a quantitative univariate approach. But the use of SMW with MS alone is not mandatory to effectively delineate soil boundaries.

This study indicated a significant similar pattern along the transect, with a correlation structure of stationary variables in the interval between 20 and 43 of window width for PC1. PC2 and PC4 had a sharp initial decrease, assuming that it consists of independent variables in the spatial scale evaluated (Fig. 3); that is, they had greater variability within shorter distances. Such large-scale variation is related to PC2, PC3, and PC4, represented by soil color properties and BS%, as these properties are highly susceptible to changes due to the exposure to processes such as surface water flow, soil erosion, sediment deposition, weathering, and anthropic action (Brown et al., 2017).

A preliminary boundary analysis was based on window width in Fig. 3. Autocorrelogram values began to stabilize near 660 m on the transect. The boundaries indicated in Fig. 4 were identified near 440 m along the transect. This result confirmed the hypothesis by Webster and Cuanalo (1975) and Rossiter (2012), where soil boundaries are around two-thirds of the expected distance along the transect, and it corresponds to two-thirds of the mean autocorrelation value. Thus, this preliminary information can be used by mappers when employing the SMW

and MSMW techniques. However, before applying MSMW, soil surveyors can interpret ACF as an initial analysis, because it indicates positions approached of mapping units by spatial structures, but it demands calibrating window widths and including soil-landscape relationships.

The major limitation of the ACF techniques with PCs is the great variability between PCs. The use of these autocorrelograms as metrics for pedological interpretation is not ease. According to Wadoux et al. (2020), due to the increase in the size of soil databases, soil mappers have faced an increasing complexity in modelling soil with soil data and covariates. Thus, integrating pedological knowledge and model interpretation is essential.

4.3. Detection of boundaries by MSMW and SMW

Based on the classification criterion described in Section 2.1, the DEM resulted in landforms based on Dalrymple et al.'s model (1968) prior to MSMW and SMW, because it is a straightforward way to increase interpretability of soil-landscape relationships. In our study, these geomorphologic units were our first step to delineate some soil boundaries (Fig. 4). Kramm et al. (2017) confirmed this procedure because it is correlated with geomorpho-pedological processes. However, our analysis of geomorphologic units in the results of boundaries had great influence of expert soil tacit knowledge in DSM, which was a challenging exercise.

For mapping purposes, the similarity between landforms was relatively consistent and should be considered for improving soil maps (Fig. 4). These results were reported by Siqueira et al. (2015), who demonstrated that the proximity of soil boundaries to the SMW peak analysis indicate that the conceptual model described by Dalrymple et al. (1968) may be used as a tool to identify boundaries in areas with different variation patterns. Bourennane et al. (2014) found good results using multivariate methods in developing landform structure classification models. Moreover, our findings support a statement by Rizzo et al. (2016): when PCs have covariates with greater performance as predictors and capture most of explained variance, they are major indicators of mapping unit boundaries. Our SMW and MSMW analyses provided strong and consistent evidence to define properly the mapping unit boundaries based on the groups of properties and MS (Fig. 4a–f). Vincent et al. (2018) incorporated specialized soil-landform relationships to establish soil mapping units, which was crucial to develop a consistent soil map. However, the protocol developed in this study demonstrated that the use of SMW and MSMW associated with geomorphological units and tacit knowledge are potential tools for locating soil boundaries.

Our findings highlight the necessity to reassess the number of covariates to use MSMW. In general, first we need to section the covariates' major potential to stratigraphy of landscape to be used in PCA described in 3.1. Samuel-Rosa et al. (2015) concluded that a more detailed covariate has a greater potential to improve prediction accuracy in soil maps. Next, we calculated a potential spatial autocorrelation, which can bias the test of significance due to assumption and is especially problematic on a regional scale. According to the literature, this approach of assessing the "analysis scale" (the spatial pattern of covariates) is common in DSM (Miller and Schaetzl, 2014).

Our results show that MSMW captures better the short-range spatial variation in comparison to SMW. This occurs because higher local pedodiversity reduces the correlation between covariates and soil properties. In contrast, SMW only provides this measure based on the MS of clay fraction of the soil. One interesting observation is that MSMW (half-slope and shoulder) delineated the mapping unit more accurately and expressively than SMW (1st and 2nd peaks) in the highest parts of the landscape, near the unit boundaries identified in the field (Fig. 4). These differences may be explained by soil texture and topography—and a more stable surface—, because the landform genesis of Latosol (Oxisol) may have a greater influence on determining soil properties than the

other factors of soil formation. For example, this soil type frequently occurs in Brazil. There are great differences between latosols, and they can be discriminated by their attributes from lithogenetic or pedogenetic and magnetic origin.

Because of the taxonomic diversity and soil mappers' difficulty in delineating boundaries, our results could have several important implications now and in the future. Jordanova et al. (2013) found an absence of direct correlation between mass specific MS and clay content, which is a widespread phenomenon in well aerated soil types. This absence may be due to the lack of pedogenic strongly magnetic iron minerals in the clay fraction (Jordanova et al., 2010). Teixeira et al. (2017) studied sample planning for quantifying and mapping soil properties, and they found significant differences in the classification of clayey soils within latosols.

However, the use of MSMW has a major potential in comparison with SMW in delineating soil boundaries (half-slope and shoulder) by identifying micro-features as landforms that may not capture functional soil-landscape relationships. SMW's 5th peak showed the most significant change in the landform after about 1880 m along the transect. This is a lower altitude region that has latosol (Oxisol) with sandy texture; there is also a change in the geomorphological units for deposition (deposition foothill). These changes may be due to alterations in the source material, and consequently textural classes—from clayey (35% to 60% clay) to medium texture (<35% clay and >15% sand).

Our study found limitations in applying SMW with MS in this area, because it depends on clay content and type, quantity, size and form of ferromagnetic minerals present in the soil. Campbell and Edmonds (1984) reported how efficient these attributes are to describe spatial variations but rather effective to understand their origins and influence land use. These zones can be stratified with values of MS and clay content as a function of soil mapping units identified in the area (Teixeira et al., 2017).

We also observed the occurrence of the 1st to the 5th peaks at 450 to 1850 m in the SMW analysis with MS. In SMW and MSMW, peaks are dislocated to the right, indicating that the boundaries can be in this site (Fig. 4a–f). In our study, this happened by indication of clay translocation and MS to southwest. The ranges of the terrain factors can significantly influence the products magnetic nanoparticles by an incompletely understood set of organic and inorganic crystal-growth and transformation processes in transitional soils (Maxbauer et al., 2017). In addition, in top landform, clay has an incipient movement to southwest that supports this interpretation, with stronger expression of peaks from MSMW. This agrees with findings by Sarmast et al. (2017).

The peaks from MSMW are powerful for identifying soil boundaries in the landscape, even in positions where soil texture did not change along the transect (e.g. 1st to 4th peak). This can be explained by the covariate position used, if we consider the approach of geomorphologists and stratigraphy, which view the soil-stratigraphic record with emphasis on the fast increment stream energy, indicates multiple episodes of landscape stability and pedogenesis (Holliday et al., 2011). This contrasts with the only expressive peak in SMW, the 5th. It indicated an alluvial region (Fig. 4). These results support our hypothesis that MSMW has a stronger potential to discriminate independent boundaries more efficiently than the univariate SMW approach for soils with gradual variation. Yet, SMW is still indicated in cases in which change of boundaries is sometimes abrupt, as the case we analyzed. This site is on an alluvial plain that has received material from higher parts during different phases. Bousbih et al. (2019) described that alluvial plains are formed mainly by alluvial deposits, with clay and coarse sand formations, so they are characterized by a high spatial variability of texture. Overall, both methods can be used in the same area by soil mappers, but we suggest SMW for more "continuous" (stable) zones and MSMW for more "discontinuous" (unstable) zones.

4.4. Mapping protocol validation

The evaluation parameters of mapping accuracy indicated the good performance of MSMW in comparison to SMW using MS, as already confirmed by Siqueira et al. (2015). We performed Pearson's correlation to correlate t-statistics data from SMW using MS at both depth ranges (0–25 cm and 25–50 cm) with data of D^2 from MSMW using the four groups, as shown in Table 2. We found medium correlations between t-statistics values and those of D^2 in G1 and G2, with $r = 0.56$ and 0.69 (depth range = 0–25 cm) and $r = 0.55$ and 0.64 (depth range = 25–50 cm), respectively. Therefore, MS can be used as an efficient covariate property to delineate mapping units of this landform at both depths. Teixeira et al. (2018) reported the incorporation of information from uncertainty maps of MS and clay content to delineate mapping units.

In G2, we found the smallest values of BIAS and LCCC: BIAS = 2.77 and LCCC = 0.40 for the depth range of 0–25 cm, and BIAS = 2.68 and LCCC = 0.35 for 25–50 cm, respectively. This result confirms that BIAS was small, so the statistical methods (SMW and MSMW) have a similar performance to delineate mapping units. G3 and G4 presented a similar pattern regarding the correlation coefficient: $r = 0.36$ and 0.44 for the depth range 0–25 cm, and $r = 0.38$ and 0.51 for 25–50 cm, respectively. These results of small correlation coefficient values and the high BIAS values indicate that the use of MSMW with this covariate differed in relation to SMW using MS. Despite presenting similar patterns, D^2 and t-statistics differed by more than 60% on the transect. This does not mean that G3 and G4 are not valid; it means they are representative in places where peaks of SMS are inexpressive, such as in the first soil boundaries delineated on the transect. This can be explained by the fact that MS depends on the existence of a contrast of clay or iron oxide to have expressive peaks (Souza Junior et al., 2010). While using MSMW with multiple covariates, the dependence of SMW by MS is overcome in the initial part of the slope, for example in half-slope, shoulder and slope. Samuel-Rosa et al. (2015) evaluated the accuracy in digital soil mapping as a function of covariates and concluded that a more detailed covariate has a greater potential to improve prediction accuracy.

The similarity between MSMW and SMW (see Table 2) accurately indicated the location of soil boundaries. Similar data was successfully used by Siqueira et al. (2015). However, to test such a hypothesis, the collection of data to understand the spatial variability is an essential step to select good quality covariates and automate the design of landform units.

4.5. Comparing methodologies to locate soil boundaries

When comparing the new soil map generated through this study with the current available soil boundaries, we can highlight the boundaries defined by MSMW. We found higher similarity between boundaries along the transect near positions 1620 m to 1980 m (Table 3). It is caused by changes of soil texture (clayey to medium) identified by MSMW and conventional methods. This variability may originate from colluvial-alluvial soils that have received and accumulated eroded soil material by water flow in the landscape. Mosleh et al. (2016) reported that terrain attributes were the main spatial predictors of some soil properties and may give a better insight about the performance of DSM approaches over low-relief areas.

The new soil boundaries defined by MSMW at 231 to 1620 m (see Table 3) are non-agreement between boundaries defined by conventional methods, and this is considered common. There is always a deviation in prediction when using DSM and conventional survey (Ma et al., 2019). The new soil boundaries at positions 480 to 1650 m (see Table 3) can indicate an ungrouping of mapping units at a more detailed level, such as variants, taxadjuncts, or other map unit with similar characteristics. According to studies on the planning of soil surveys and land development, these differences in boundaries are common in transitional soil environments (Vink, 1963). This is a very important finding for pedological mapping applications in Brazilian soils, because

there is a gap regarding tools to delineate soil boundaries in Brazil. Most of soil surveys are still made with inference of soil boundaries based on small-scale cartographic resources, a number of modal profiles, and auxiliary sampling with auger, resulting in little detailed measurements.

Soils grouped into the same cluster (soils between peaks) had similarities in the MSMW analysis. They combined more than four map units at positions 480 to 1650 m compared to one single map unit for LVdf in the conventional method, and these new boundaries can be incorporated in the pedological knowledge. Sarmento et al. (2017) concluded that it is possible to successfully delineate individual soils within map units using expert tacit knowledge and common DSM techniques.

We described the soil legend in Table 2. The major difference was in the mapped area of LVdf. This occurs because of half slope, shoulder and slope positions, and colluvial-alluvial deposits. These landscape patterns demonstrate an increase in erosive and pedogenetic processes toward the Southwest (deposition foothill), where younger soils predominate. Similarly, Poppiet et al. (2019) verified the occurrence of soil classes variability due to the pedomorphogeological relationships in a mapped area of Latosol.

Table 3 illustrates the results of peaks by MSMW, which may be divided into four new mapping units (MU2, MU3, MU4 and MU5). The first LVdf unit was found at approximately 480 m on the transect. But, for the classifier, this new taxonomy unit needs further investigation or an improved representation of the resource space to avoid unwanted extrapolations (Minasny and McBratney, 2006). Mapping this uncertainty is an advantage of MSMW. In soil survey, it may prevent the emergence of subjective elements inherent to the field mapping process. Adhikari et al. (2014) reported the correlation between boundaries obtained by conventional methods and statistical methods; they obtained a good agreement with statistical methods in detailed soil surveys.

Finally, some limitations may hinder the use of MSMW in case predictions can be potentially and sufficiently improved given the extra costs involved in using more detailed covariates. In fact, MSMW may be more efficient with more covariates. But if the extra budget in deriving more detailed covariates is an uncertainty at the moment, we suggest that one focus on substantially improving the detail of covariates' influencing degree. However, other means to calibrate MSMW should be considered.

Another limitation is the speculation as to the validity of recorded field observations in relation to the soil boundaries delineated by MSMW. MSMW uses remote sensing by pixelization of images with a different resolution in order to capture and understand lateral soil boundaries and mapping units. We suggest testing MSMW on a larger number of representative transects on different scales. Additionally, we recommend the test of other covariates, such as curvature profile, topographic index, slope, topographic factor (LS), different relief dissections, land use, vegetation, and geology. In consequence, we suggest the assessment of accuracy with field reference data to evaluate and relate the data to other factors, such as covariate precision and cost. Some deficiencies are when is applied MSMW using Mahalanobis distance in a big dataset of original variables highly correlated according to (Varmuza and Filzmoser, 2016). However, this can be solved using the D^2 of PC scores from PCA. With MSMW, assessing the capability to separate homogeneous and taxonomically pure map units is to consistently define or map them on only one hillslope position, as described in Miller and Schaetzl (2015). This approach demonstrated a high potential to assist pedological cartography, but pedological knowledge should still be considered to increase the interpretability of MSMW. Based on our work, MSMW's application in other regions is a hypothesis that can be tested in future research.

5. Conclusions

The combination of MSMW analysis and D^2 with field observations is a promising tool to identify the best position for potential boundaries of

detailed soil mapping units. The MSMW analysis showed to be a useful diagnostic tool to evaluate the reliability of predictions made by conceptual landform models, thus providing additional tools for soil mapping. Therefore, we can automate the identification of mapping units based on landform-soil relationships.

The MSMW analysis can also be combined with tacit pedological knowledge to construct new mapping protocols, as well as to develop new or revised soil-landform relationships. This technique can expand the pedological vision during conventional and digital soil mappings; therefore, it should be used by surveyors to validate their tacit knowledge on soil-landform relationships.

The SMW and MSMW analyses associated with pedological knowledge can also be used to refine boundaries in soil maps. The use of MS as a covariate, at the depth ranges of 0–25 cm and 25–50 cm, proved to be efficient for the spatial distribution of landform mapping units. MS can be an economical, efficient, and non-invasive covariate in soil surveys by means of the landform transect approach to design detailed mapping units.

In conclusion, our findings suggest that soil mappers can benefit from MSMW using D² information by applying new knowledge about soil boundaries along the transect, which can be used to recalibrate mappers' tacit knowledge of soil-landscape relationships. In Brazil, this represents an opportunity to apply the experience of several "digital soil mappers" or "conventional soil mappers" to produce better soil maps at a low cost.

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.

Acknowledgments

The authors would like to thank the São Paulo Research Foundation (FAPESP) for granting a Doctorate scholarship to the second author (No. 13/17552-6) and Brazilian National Council for Scientific and Technological Development (CNPq) for the scholarship granted to the fourth, fifth and sixth authors.

References

- Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H., 2014. Constructing a soil class map of Denmark based on the FAO legend using digital techniques. *Geoderma* 214–215, 101–113. <https://doi.org/10.1016/j.geoderma.2013.09.023>.
- Barrios, M.R., Marques Jr., J., Panosso, A.R., Siqueira, D.S., Scala Jr., N.L., 2012. Magnetic susceptibility to identify landscape segment on a detailed scale in the region of Jaboticabal, São Paulo, Brazil. *Rev. Bras. Cienc. Solo* 36, 1073–1082. <https://doi.org/10.1590/S0100-06832012000400002>.
- Beaudette, D.E., Roudier, P., O'Geen, A.T., 2013. Algorithms for quantitative pedology: a toolkit for soil scientists. *Comput. Geosci.* 52, 258–268. <https://doi.org/10.1016/j.cageo.2012.10.020>.
- Bourennane, H., Couturier, A., Pasquier, C., Chartin, C., Hinschberger, F., Macaire, J.J., Salvador-Blanes, S., 2014. Comparative performance of classification algorithms for the development of models of spatial distribution of landscape structures. *Geoderma* 219, 136–144. <https://doi.org/10.1016/j.geoderma.2014.01.001>.
- Bousbih, S., Zribi, M., Pelletier, C., Gorrab, A., Chabaane, Z.L., Baghdadi, N., Aissa, N.B., Mougenot, B., 2019. Soil texture estimation using radar and optical data from Sentinel-1 and Sentinel-2. *Remote Sens.* 11, 1520. <https://doi.org/10.3390/rs11131520>.
- Brereton, R.G., 2015. The Mahalanobis distance and its relationship to principal component scores. *Aust. J. Chem.* 3, 143–145. <https://doi.org/10.1002/cem.2692>.
- Brown, A.G., Tooth, S., Bullard, J.E., Thomas, D.S.G., Chiverrell, R.C., Plater, A.J., Murton, J., Thorndycraft, V.R., Tarolli, P., Rose, J., Wainwright, J., Downs, P., Aalto, R., 2017. The geomorphology of the anthropocene: emergence, status and implications. *Earth Surf. Process. Landforms* 42, 71–90. <https://doi.org/10.1002/esp.3943>.
- Bui, E.N., Searle, R.D., Wilson, P.R., Philip, S.R., Thomas, M., Brough, D., Harms, B., Hill, J.V., Holmes, K., Smolinski, H.J., Gool, D.V., 2020. Soil surveyor knowledge in digital soil mapping and assessment in Australia. *Geoderma Reg.* 22 <https://doi.org/10.1016/j.gedr.2020.e00299>. S2352-0094(20)30048-1.
- Campbell, J.B., Edmonds, W.J., 1984. The missing geographic dimension to soil taxonomy. *Ann. Assoc. Am. Geogr.* 74, 83–97. <https://doi.org/10.1111/j.1467-8306.1984.tb01436.x>.
- Carré, F., Jacobson, M., 2009. Numerical classification of soil profile data using distance metrics. *Geoderma* 148, 336–345. <https://doi.org/10.1016/j.geoderma.2008.11.008>.
- Carvalho Júnior, W., Lagacherie, P., Chagas, C.S., Calderano Filho, B., Bhering, S.B., 2014. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. *Geoderma* 232–234, 479–486. <https://doi.org/10.1016/j.geoderma.2014.06.007>.
- Cornelius, J.M., Reynolds, J.F., 1991. On determining the statistical significance of discontinuities within ordered ecological data. *Ecology* 72, 2057–2070. <https://doi.org/10.2307/1941559>.
- Costa, A.C.S., Bigham, J.M., Rhoton, F.E., Traina, S.J., 1999. Quantification and characterization of maghemite in soils derived from volcanic rocks in southern Brazil. *Clay Clay Miner.* 47, 466–473. <https://doi.org/10.1346/CCMN.1999.0470408>.
- Dalrymple, J.B., Blong, R.J., Conacher, A.J., 1968. A hypothetical nine unit land a surface model. *Z. Geomorphol.* 12, 60–76.
- Davis, J.C., 1973. Statistics and data analysis in geology, 23. John Wiley & Sons, New York, pp. 51–58.
- Dearing, J.A., 1999. Environmental Magnetic Susceptibility: Using the Bartington MS2 System, 2nd ed. Chi Publishing, Kenilworth.
- Embrapa, 1997. Empresa Brasileira de Pesquisa Agropecuária. Centro Nacional de Pesquisa de Solos. Manual Methods of Soil Analysis, Rio de Janeiro (212 pp. (in Portuguese)).
- Gath, E.G., Hayes, K., 2006. Bounds for the largest Mahalanobis distance. *Linear Algebra Appl.* 419, 93–106. <https://doi.org/10.1016/j.laa.2006.04.007>.
- Holliday, V.T., Meltzer, D.J., Mandel, R., 2011. Stratigraphy of the Younger Dryas Chronozone and paleoenvironmental implications: central and southern Great Plains. *Quat. Int.* 242, 520–533. <https://doi.org/10.1016/j.quaint.2011.03.047>.
- Hudson, B.D., 1992. The soil survey as a paradigm-based science. *Soil Sci. Soc. Am. J.* 56, 836–841. <https://doi.org/10.2136/sssaj1992.03615995005600030027x>.
- Hughes, P.A., McBratney, A.B., Minasny, B., Campbell, S., 2014. End members, end points and extragrades in numerical soil classification. *Geoderma* 226–227, 365–375. <https://doi.org/10.1016/j.geoderma.2014.03.010>.
- Hughes, P., McBratney, A.B., Huang, J., Minasny, B., Micheli, E., Hempel, J., 2018. A nomenclature algorithm for a potentially global soil taxonomy. *Geoderma* 322, 56–70. <https://doi.org/10.1016/j.geoderma.2018.02.020>.
- IPT – Instituto de Pesquisas Tecnológicas do Estado de São Paulo, 1981. *Geomorphological Map of the State of São Paulo (94 pp.)*.
- Iticha, B., Takele, C., 2018. Soil-landscape variability: mapping and building detail information for soil management. *Soil Use Manag.* 34, 111–123. <https://doi.org/10.1111/sum.12404>.
- Jafari, A., Ayoubi, S., Khademi, H., Finke, P.A., Toomanian, N., 2013. Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. *Geomorphology* 201, 86–97. <https://doi.org/10.1016/j.geomorph.2013.06.010>.
- Jordanova, D., Jordanova, N., Petrov, P., Tsacheva, T., 2010. Soil development of three Chernozem-like profiles from North Bulgaria revealed by magnetic studies. *Catena* 83 (2–3), 158–169. <https://doi.org/10.1016/j.catena.2010.08.008>.
- Jordanova, D., Jordanova, N., Werban, U., 2013. Environmental significance of magnetic properties of Gley soils near Rosslau (Germany). *Environ. Earth Sci.* 69, 1719–1732. <https://doi.org/10.1007/s12665-012-2006-3>.
- Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200. <https://doi.org/10.1007/BF02289233>.
- Kaiser, H.F., 1960. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20, 141–151.
- Kramm, T., Hoffmeister, D., Curdt, C., Maleki, S., Khormali, F., Kehl, M., 2017. Accuracy assessment of landform classification approaches on different spatial scales for the Iranian loess plateau. *ISPRS Int. J. Geo-Inf.* 6, 366. <https://doi.org/10.3390/ijgi6110366>.
- Lagacherie, P., 2008. Digital soil mapping: a state of the art. In: Hartemink, A.E., McBratney, A., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer, Dordrecht, The Netherlands, pp. 3–14. https://doi.org/10.1007/978-1-4020-8592-5_1.
- Lin, L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45, 255–268. <https://doi.org/10.2307/2532051>.
- Ludwig, J.A., Cornelius, J.M., 1987. Locating discontinuities along ecological gradients. *Ecology* 68, 448–450. <https://doi.org/10.2307/1939277>.
- Ma, Y., Minasny, B., Malone, B.P., McBratney, A.B., 2019. Pedology and digital soil mapping (DSM). *Eur. J. Soil Sci.* 70, 216–235. <https://doi.org/10.1111/ejss.12790>.
- MacMillan, R.A., Moon, D.E., Coupe, R.A., 2007. Automated predictive ecological mapping in a Forest Region of B.C., Canada, 2001–2005. *Geoderma* 140, 353–373. <https://doi.org/10.1016/j.geoderma.2007.04.027>.
- Massawe, B.H.J., Subburayalu, S.K., Kaaya, A.K., Winowiecki, L., Slater, B.K., 2018. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. *Geoderma* 311, 143–148. <https://doi.org/10.1016/j.geoderma.2016.11.020>.
- Matias, S.S.R., Marques Jr., J., Siqueira, D.S., Pereira, G.T., 2013. Modelos de paisagem e susceptibilidade magnética na identificação e caracterização do solo. *Pesqui. Agrop. Trop.* 43, 93–103. <https://doi.org/10.1590/S1983-40632013000100003>.
- Maxbauer, D.P., Feinberg, J.M., Fox, D.L., Nater, E.A., 2017. Response of pedogenic magnetite to changing vegetation in soils developed under uniform climate, topography, and parent material. *Sci. Rep.* 7, 17575. <https://doi.org/10.1038/s41598-017-17722-2>.

- McBratney, A.B., Odeh, I.O.A., 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma* 77, 85–133. [https://doi.org/10.1016/S0016-7061\(97\)00017-7](https://doi.org/10.1016/S0016-7061(97)00017-7).
- McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* 97, 293–327. [https://doi.org/10.1016/S0016-7061\(00\)00043-4](https://doi.org/10.1016/S0016-7061(00)00043-4).
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
- Miller, B.A., Schaetzl, R.J., 2014. The historical role of base maps in soil geography. *Geoderma* 230–231, 229–239. <https://doi.org/10.1016/j.geoderma.2014.04.020>.
- Miller, B.A., Schaetzl, R.J., 2015. Digital classification of hillslope position. *Soil Sci. Soc. Am. J.* 79, 132–145. <https://doi.org/10.2136/sssaj2014.07.0287>.
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* 32, 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>.
- Minasny, B., McBratney, A.B., 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* 142, 285–293. <https://doi.org/10.1016/j.geoderma.2007.08.022>.
- Minasny, B., McBratney, A.B., 2016. Estimating the water retention shape parameter from Digital soil mapping: a brief history and some lessons. *Geoderma* 264, 301–311. <https://doi.org/10.2136/sssaj2006.0298N>.
- Mosleh, Z., Salehi, M.H., Jafari, A., Borujeni, I.E., Mehnatkesh, A., 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environ. Monit. Assess.* 3, 188–195. <https://doi.org/10.1007/s10661-016-5204-8>.
- Ngunjiri, M.W., Libohova, Z., Minai, J.O., Serrem, C., Owens, P.R., Schulze, D.G., 2019. Predicting soil types and soil properties with limited data in the Uasin Gishu Plateau, Kenya. *Geoderma Reg* 16, e00210. <https://doi.org/10.1016/j.geodrs.2019.e00210>.
- Norris, J.M., 1971. The application of multivariate analysis to soil studies: 1. Grouping of soils using different properties. *J. Soil Sci.* 22, 69–80. <https://doi.org/10.1111/j.1365-2389.1971.tb01594.x>.
- Nounou, M.N., Bakshi, B.R., 1999. On-line multiscale filtering of random and gross errors without process models. *AICHE J.* 45, 1041–1058.
- Omran, M.G.H., Engelbrecht, A.P., Salman, A., 2007. An overview of clustering methods. *Intell. Data Anal.* 11, 583–605. <https://doi.org/10.3233/IDA-2007-11602>.
- Pearson, K., 1920. Notes on the History of Correlation. *Biometrika* 13, 25–45. <https://doi.org/10.2307/2331722>.
- Peng, L., Cheng-zhi, Q., A-xing, Z., Zhi-wei, H., Nai-qing, F., Yi-jie, W., 2020. A case-based method of selecting covariates for digital soil mapping. *J. Integr. Agric.* 19, 2127–2136. [https://doi.org/10.1016/S2095-3119\(19\)62857-1](https://doi.org/10.1016/S2095-3119(19)62857-1).
- Poppeli, R.R., Lacerda, M.P.C., Dematté, J.A.M., Oliveira Jr., M.P., Gallo, B.C., Safanelli, J.L., 2019. Pedology and soil class mapping from proximal and remote sensed data. *Geoderma* 384, 189–206. <https://doi.org/10.1016/j.geoderma.2019.04.028>.
- R Development Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. URL: <http://www.R-project.org> [accessed on 08 September 2019].
- Ramos, P.V., Dalmolin, R.S.D., Marques Jr., J., Siqueira, D.S., Almeida, J.A., Moura-Bueno, J.M., 2017. Magnetic susceptibility of soil to differentiate soil environments in southern Brazil. *Rev. Bras. Cienc. Solo* 41, 1–13. <https://doi.org/10.1590/18069657rbcs20160189>.
- Rizzo, R., Dematte, J.A.M., Lepsch, I.F., Gallo, B.C., Fongaro, C.T., 2016. Digital soil mapping at local scale using a multi-depth Vis–NIR spectral library and terrain attributes. *Geoderma* 274, 18–27. <https://doi.org/10.1016/j.geoderma.2016.03.019>.
- Rossiter, D.G., 2012. Technical Note: Optimal Partitioning of Soil Transects with R. Version 2.1. URL: http://www.itc.nl/personal/rossiter/teach/R/R_OptPart.pdf [accessed on 24 October 2020].
- Rossiter, D., 2018. Past, present & future of information technology in pedometrics. *Geoderma* 324, 131–137. <https://doi.org/10.1016/j.geoderma.2018.03.009>.
- Rossiter, D.G., Zeng, R., Zhang, G.L., 2017. Accounting for taxonomic distance in accuracy assessment of soil class predictions. *Geoderma* 292, 118–127. <https://doi.org/10.1016/j.geoderma.2017.01.012>.
- Samuel-Rosa, A., Heuvelink, G.B.M., Vasques, G.M., Anjos, L.H.C., 2015. Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma* 243–244, 214–227. <https://doi.org/10.1016/j.geoderma.2014.12.017>.
- Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Lumbierras, J.F., Coelho, M.R., Almeida, J.A., Araújo Filho, J.C., Oliveira, J.B., Cunha, T.J.F., 2018. Sistema brasileiro de classificação de solos, Fifth ed. Embrapa, Brasília. 530pp.
- Sarmast, M., Farpoor, M.H., Boroujeni, I.E., 2017. Magnetic susceptibility of soils along a lithotoposequence in southeast Iran. *Catena* 156, 252–262. <https://doi.org/10.1016/j.catena.2017.04.019>.
- Sarmiento, E.C., Giasson, E., Weber, E.J., Flores, C.A., Hasenack, H., 2017. Disaggregating conventional soil maps with limited descriptive data: A knowledge-based approach in Serra Gaúcha, Brazil. *Geoderma Reg* 8, 12–23. <https://doi.org/10.1016/j.geodrs.2016.12.004>.
- Silva Júnior, J.F., Siqueira, D.S., Marques Jr., J., Pereira, G.T., 2012. Classificação numérica e modelo digital de elevação na caracterização espacial de atributos dos solos. *Rev. Bras. Eng. Agric. Ambient.* 16, 415–424. <https://doi.org/10.1590/S1415-43662012000400012>.
- Siqueira, D.S., Marques Jr., J., Pereira, G.T., 2010. The use of landforms to predict the variability of soil and orange attributes. *Geoderma* 155, 55–66. <https://doi.org/10.1016/j.geoderma.2009.11.024>.
- Siqueira, D.S., Marques Jr., J., Pereira, G.T., Teixeira, D.B., Vasconcelos, V., Carvalho Júnior, O.A., Martins, E.S., 2015. Detailed mapping unit design based on soil–landscape relation and spatial variability of magnetic susceptibility and soil color. *Geoderma* 135, 149–162. <https://doi.org/10.1016/j.catena.2015.07.010>.
- Soil Survey Staff, 2014. Keys to Soil Taxonomy, twelfth ed. USDA-Natural Resources Conservation Service, Washington. 360pp. https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=stelprdb1252094&ext=pdf. (Accessed 20 October 2020).
- Souza Junior, I.G., Costa, A.C.S., Vilar, C.C., Hoepers, A., 2010. Mineralogia e susceptibilidade magnética dos óxidos de ferro do horizonte B de solos do Estado do Paraná. *Cienc. Rural* 40, 513–519. <https://doi.org/10.1590/S0103-84782010000300003>.
- Teixeira, D.D.B., Marques Jr., J., Siqueira, D.S., Vasconcelos, V., Carvalho Jr., O.A., Carvalho Jr., O.A., Martins, E.S., Pereira, G.T., 2017. Sample planning for quantifying and mapping magnetic susceptibility, clay content, and base saturation using auxiliary information. *Geoderma* 305, 208–218. <https://doi.org/10.1016/j.geoderma.2017.06.001>.
- Teixeira, D.D.B., Marques Jr., J., Siqueira, D.S., Vasconcelos, V., Carvalho Jr., O.A., Martins, E.S., Pereira, G.T., 2018. Mapping units based on spatial uncertainty of magnetic susceptibility and clay content. *Catena* 164, 79–87. <https://doi.org/10.1016/j.catena.2017.12.038>.
- Thornthwaite, C.W., 1948. An approach towards a rational classification of climate. *Geogr. Rev.* 38, 55–94. <https://doi.org/10.2307/210739>.
- Varmuza, K., Filzmoser, P., 2016. *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC press, Boca Raton, FL.
- Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex Soil Map Units at the regional scale based on soil–landscape relationships. *Geoderma* 311, 130–142. <https://doi.org/10.1016/j.geoderma.2016.06.006>.
- Vink, A.P.A., 1963. Soil survey in relation to agriculture productivity. *J. Soil Sci.* 14, 88–101. <https://doi.org/10.1111/j.1365-2389.1963.tb00934.x>.
- Viscarra Rossel, R.A., Bui, E.N., de Caritat, P., McKenzie, N.J., 2010. Mapping iron oxides and the color of Australian soil using visible–near-infrared reflectance spectra. *Case Rep. Med.* 115, 1–13. <https://doi.org/10.1029/2009JF001645>.
- Wadoux, A.M.J.-C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. *Earth Sci. Rev.* 210, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>.
- Webster, R., 1978. Optimally partitioning soil transects. *Eur. J. Soil Sci.* 29, 388–402. <https://doi.org/10.1111/j.1365-2389.1978.tb00789.x>.
- Webster, R., 1994. The development of pedometrics. *Geoderma* 62, 1–15. [https://doi.org/10.1016/0016-7061\(94\)90024-8](https://doi.org/10.1016/0016-7061(94)90024-8).
- Webster, R., Cuanal, DE LAC.H.E., 1975. Soil transect correograms of north Oxfordshire and their interpretation. *Eur. J. Soil Sci.* 26, 176–194. <https://doi.org/10.1111/j.1365-2389.1975.tb01942.x>.
- Webster, R., Wong, I.F.T., 1969. A numerical procedure for testing soil boundaries interpreted from air photographs. *Photogrammetria* 24, 59–72. [https://doi.org/10.1016/0031-8663\(69\)90005-2](https://doi.org/10.1016/0031-8663(69)90005-2).
- Wyszecki, G., Stiles, W.S., 1982. *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. John Wiley & Sons, New York, p. 976.
- Zhang, G.L., Liu, F., Song, X.D., 2017. Recent progress and future prospect of digital soil mapping: A review. *J. Integr. Agric.* 16, 2871–2885. [https://doi.org/10.1016/S2095-3119\(17\)61762-3](https://doi.org/10.1016/S2095-3119(17)61762-3).
- Zhu, A.X., Burt, J.E., Moore, A.C., Smith, M.P., Liu, J., Qi, F., 2003. SoLIM: A New Technology For Soil Mapping Using GIS, Expert Knowledge and Fuzzy Logic. URL: <http://solim.geography.wisc.edu/pubs/Overview2007-02-16.pdf>. Accessed on: sep 04, 2020. [accessed on 26 July 2019].