ISSN: 1984-5529

Científica

Jaboticabal

v.48, n.1, p.25–35, 2020

http://dx.doi.org/10.15361/1984-5529.2020v48n1p25-35

Using multivariate analysis to design management zones

Uso de análise multivariada no delineamento de zonas de manejo

Calisto Manuel MÁQUINA[1]; Nerilson Terra SANTOS[2]; Marcelo Marques COSTA[3]; Samuel de Assis SILVA[4]

[1] Correspondent Author; Assistant Professor, Rovuma University, Department of Statistics and Information Management, Nampula, Mozambique. Email: calisto.calmaqna.maquina@gmail.com
[2] Full Professor, Federal University of Viçosa, Department of Statistics, Viçosa, MG-Brazil; E-mail: nsantos@ufv.br
[3] Professor, Federal University of Jataí, Special Academic Unit of Agrarian Sciences, Jataí, GO-Brazil; E-mail: marcelo.marques.costa@gmail.com
[4] Professor, Federal University of Espírito Santo, Rural Engineering Department, ES-Brazil; E-mail: samuel.assilva@gmail.com

Recebido em: 12-10-2019; Aceito em: 20-12-2019

Abstract

Soil chemical and physical attributes are important in any agricultural cropping system, but in precision agriculture they are more relevant due to the possibility of application using different management practices along a production field. However, the correlation between these attributes has been little explored in the delineation of management zones. This work aims to maximize the use of joint spatial variability for soil attributes. Its secondary objectives were 1) reduction of spatial variability dimensionality among all attributes and 2) assessment of agreement between univariate and multivariate management zones. The management zones resulting from the interpolation of attribute values, as well as from the scores of each of the three main components, were delineated using the Fuzzy c-means algorithm. The fuzzy performance and modified partition entropy indexes were used to determine the optimal number of management zones. The Kappa index was used to evaluate the agreement of management zones obtained from attributes with those obtained from principal components. By using principal component analysis, it was possible to reduce the dimensionality of the number of variables that contribute to the joint spatial variability existing in the study area. There was no complete agreement between the uni- and multivariate management zones outlined, which is why further studies on the subject are needed.

Additional keywords: Fuzzy c-means algorithm; geostatistics; principal component; spatial variability.

Resumo

Os atributos químicos e físicos do solo são importantes em qualquer sistema de cultivo agrícola, porém na agricultura de precisão eles recebem maior atenção devido à possibilidade de aplicação em práticas de manejo diferenciadas ao longo de um campo de produção. Todavia, a correlação entre esses atributos tem sido pouco explorada no delineamento de zonas de manejo. Com este trabalho, objetivou-se maximizar o uso da variabilidade espacial conjunta entre os atributos de solo. Seus objetivos secundários foram: 1) redução da dimensionalidade da variabilidade espacial entre todos os atributos e 2) avaliação da concordância entre as zonas de manejo univariadas e as multivariadas. As zonas de manejo resultantes da interpolação dos valores dos atributos, assim como a partir dos escores de cada uma das três componentes principais, foram delineadas com o uso do algoritmo *Fuzzy c-means.* Os índices de performance *Fuzzy* e de partição da entropia modificada foram utilizados para determinar o número ótimo de zonas de manejo. O índice Kappa foi empregado para avaliar a concordância das zonas de manejo obtidas a partir dos atributos com aquelas obtidas a partir das componentes principais. Com o uso da análise de componentes principais foi possível reduzir a dimensionalidade do número de variáveis que contribuem para a variabilidade espacial conjunta existente na área em estudo. E não houve concordância completa entre as zonas de manejo uni e multivariadas delineadas, razão por que mais estudos sobre o assunto são necessários.

Palavras-chave adicionais: algoritmo Fuzzy c-means; componente principal; geoestatística; variabilidade espacial.

Introduction

Precision agriculture (PA) is defined by Inamasu & Bernardi (2014) as an agricultural management system based on spatial and temporal variability of agricultural production fields and aims to increase economic returns and sustainability and minimize effects on the environment. For the development of this activity, it is essential to know the attributes within these management zones. Several researches focusing on PA have suggested the evaluation of spatial variability of soil attributes.

A study by Rodrigues & Corá (2015) showed that the preliminary evaluation of spatial variability of soil physical and chemical attributes is essential for a good crop management. Based on the information of these attributes, it is possible to map the patterns of variability of production factors. This mapping enables the establishment of zones where cultural management should be homogeneous, which allows the application of the amount of fertilizer effectively needed in each point. Therefore, localized input application techniques are very important for a sustainable agricultural production, that is, a production that seeks to obtain high-quality products, using techniques that guarantee sustainability and preserve the soil and the environment.

According to Sylvester-Bradley et al. (1999), the relation between cost and quality in precision agriculture depends mainly on the delineation of management zones within potential fields for agricultural practice. Identifying these management zones is an important step towards good practices in precision agriculture. To this end, account should be taken of this identification of spatial variability of crop productivity and the soil physical and chemical properties, which play a fundamental role in agricultural production.

Several authors have argued that the evaluation of spatial variability of soil physical and chemical properties within specific zones is also important for the evaluation of potential use of variable rates (Beckett, 1971; Stafford, 1996). Several studies, such as those by Silva et al. (2010c) and Silva & Lima (2012), have pointed out that an irregular distribution of chemical and physical properties has been a major challenge for precision agriculture. This is especially true when such distribution occurs in zones with higher concentrations of attributes that, in many situations, are correlated. Multivariate analysis techniques could be used for the outlining of management zones.

Multivariate data analysis procedures have been highlighted in studies related to many spatially correlated attributes. Principal component analysis (PCA) is a multivariate technique in which, from p correlated variables, a number of variables is obtained $q<p$, so that, such q independent variables, explain much of the variability contained in p original variables (Junior, 2012; Mingoti, 2005). The principal components are obtained in such a way that the former explains the largest variance present in p variables (Ribeiro Júnior, 2012; Li et al. 2007; Mingoti, 2005; Johnson & Wichern, 2002). PCA has been used in precision agriculture works, as in the study by Silva et al. (2010a), in which the objective was to analyze the spatial variability of chemical attributes of a humic Red-Yellow Latosol, and in the study by Li et al. (2007), in which one of the objectives was to characterize the spatial variability of the soil and the landscape attributes that affect crop productivity.

This work aims to maximize the use of joint spatial variability for soil attributes. Its secondary objectives were 1) reduction of spatial variability dimensionality among all attributes and 2) assessment of agreement between univariate and multivariate management zones.

Materials and methods

Description of the study area and data collection

The experiment was carried out in an area located at the coordinates 16º28'20" S and 49º00'32" N, in the municipality of Goianápolis, Goiás (GO) state, Brazil. The soil of the area was classified as Red--Yellow Latosol according to the Brazilian Soil Classification System (Santos et al., 2017).

Data collection was performed on a 150-point sampling grid over an area of 75 hectares. The sample points were georeferenced using a Global Navigation Satellite System (GNSS) receiver. This database originated from the work of Costa (2011).

In each of the 150 points, 13 chemical attributes and one physical attribute of the soil were measured. The chemical attributes measured were potential of hydrogen (pH in water), phosphorus (P), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), copper (Cu), iron (Fe), manganese (Mn), zinc (Zn), organic matter (OM), sulfur (S), boron (B), and electrical conductivity of the saturation extract (Ec). The measured physical attribute was clay.

The first stage of data analysis consisted of descriptive analysis of chemical and physical attributes with the objective of statistically describing their distributions. The Shapiro & Wilk (1968) normality test was also performed to verify whether the probability distribution associated with each chemical and physical attribute approximates a normal distribution.

Statistical analysis

Univariate analysis of attributes

Initially, descriptive statistics were obtained for each attribute Z_i (so that $i = 1, \cdots, 14$) to obtain distributions, dispersions and relations between the attributes. Multivariate analysis methodologies assume that the variables (attributes) under analysis are related.

Multivariate analysis of attributes

The PCA was used to reduce the dimensionality of spatial variability to outline management zones. Considering the correlation between the $p = 14$ attributes (13 chemical and one physical), the Y_j principal components were selected so that $j = 1, \cdots, q$ and $q<p$. There are several criteria for determining the optimal number of principal components. In this paper, we used the criterion presented by Kaiser (1960), which considers for analysis all principal components that have eigenvalues greater than 1. This criterion was also described by Dunteman (1989) and Johnson & Wichern (2002), according to whom the principal components with high eigenvalues are considered better since they explain a greater variance of p original variables.

Analysis of spatial variability

For each Z_i attribute and principal component Y_j, the theoretical variogram model that best fitted the empirical variogram was selected. For each of the 14 attributes, semivariance values (Webster & Oliver, 1990) were obtained by:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{s=1}^{N(h)} [Z_{si}(\alpha) \ - Z_{si}(\alpha + h)\ ]^2$$

wherein $[Z_{si}(\alpha);\ Z_{si}(\alpha + h)]$ represents each of the s value pairs of the attribute $Z_i(\alpha)$ at positions *α* and $\alpha + h$, and $N(h)$ is the number of value pairs that satisfy the separation distance *h*. However, for each principal component, the semivariance values were obtained by

$$\gamma(h) = \frac{1}{2N(h)} \sum_{s=1}^{N(h)} \left[Y_{sj}(\alpha) \ - Y_{sj}(\alpha + h)\ \right]^2$$

where $\left[Y_{sj}(\alpha);\ Y_{sj}(\alpha + h)\right]$ represents each of the s value pairs of the principal component $Y_j(\alpha)$ at positions α and $\alpha + h$, and $N(h)$ is the number of value pairs that satisfy the separation distance h.

From each of the selected variogram theoretical models, ordinary kriging interpolated maps were plotted (Isaaks & Srivastava, 1989; Wackernagel, 2003; Webster & Oliver, 2007). The cross-validation statistics was used in both the selection of the variogram theoretical model and in the definition of the search neighborhood to perform the ordinary kriging for each attribute.

From each of the maps plotted, management zones were defined by Fuzzy c-means clustering, a method which, according to Hartigan (1975), aims to partition all observations into *c* groups, in which each observation is automatically assigned to a group closer to the mean. The optimal number of partitions was defined by the joint evaluation of two indexes, the fuzzy performance index (FPI), and the modified partition entropy index (MPE), according to Gorsevski et al. (2003). Several studies have used FPI and MPE indexes in outlining management zones, such as Li et al. (2007) and Rodrigues et al. (2015). Six clusters were considered as the maximum number of management zones for each PC_j, and scenarios with two, three, four, five and six classes were simulated. The optimal number of management zones for each PC_j was determined by the point from which FPI and MPE values were minimum.

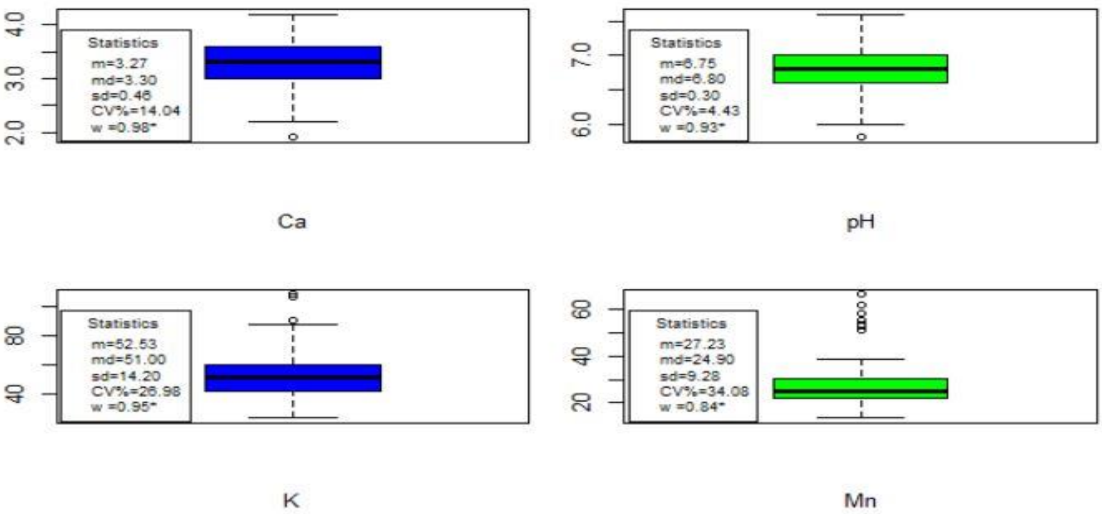
This number of clusters is commonly used in several works on precision agriculture. It aims to determine the optimal number of management zones (Li et al., 2007; Rodrigues et al., 2015; Oldoni & Bassoi, 2016).

The Kappa index, according to Congalton & Mead (1986), was used to evaluate the agreement of management zones defined by attributes and those defined by main components.

Results and discussion

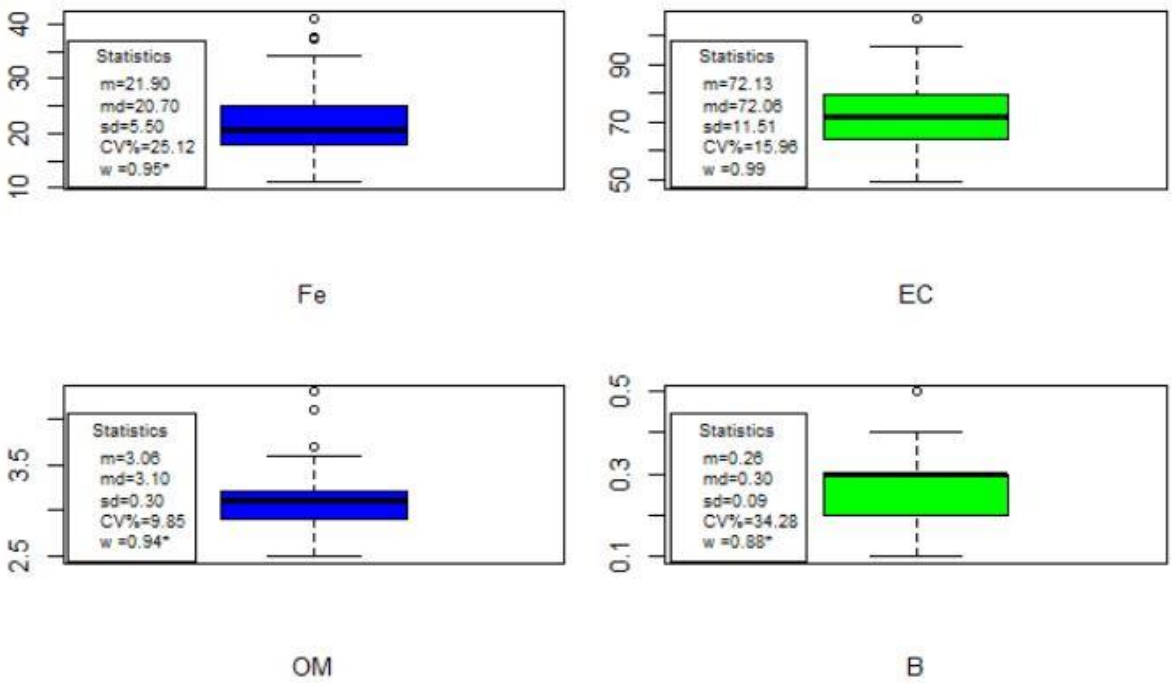
Descriptive statistics of attributes

Figures 1 to 4 show the boxplots and the statistical values (mean, median, standard deviation, percentage of coefficient of variation, and Shapiro-Wilk test for normality) of all attributes considered in this research.



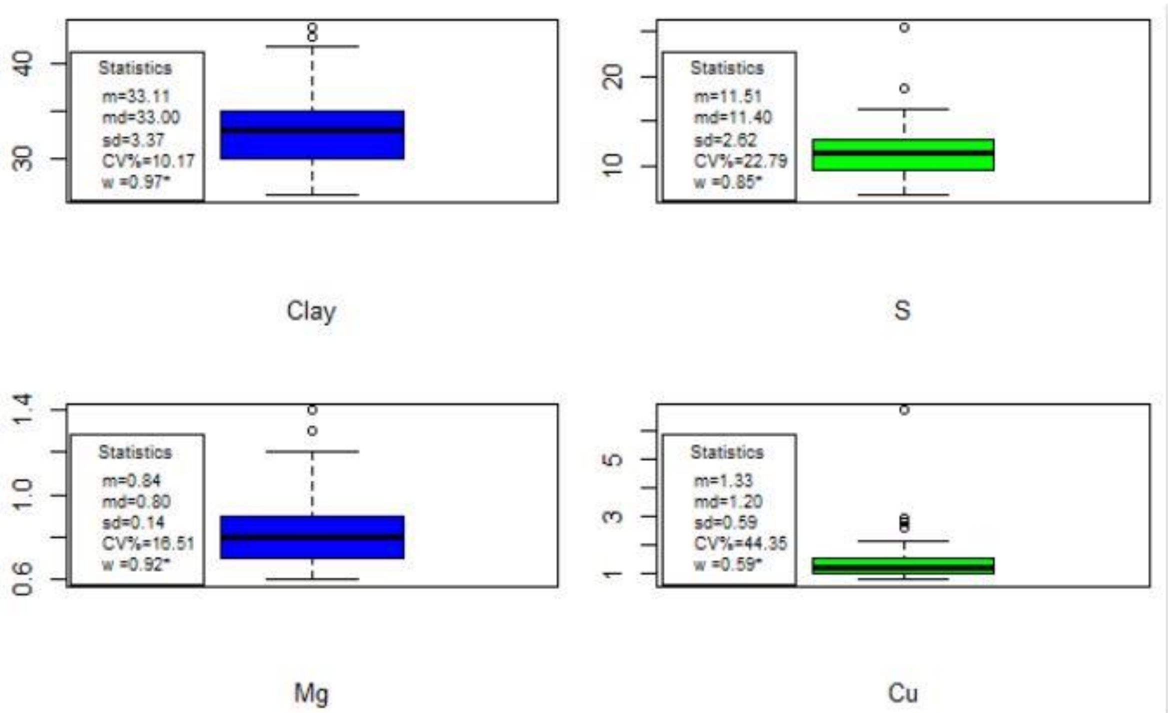
m - Mean; md - Median; sd - Standard deviation; CV% - Coefficient of variation (%); w - Shapiro-Wilk statistics for normality test; and * - Normality test significant at 5% probability.

Figure 1 - Boxplots of the chemical attributes: Ca - Calcium, pH - Potential of hydrogen, K - Potassium, and Mn - Manganese.



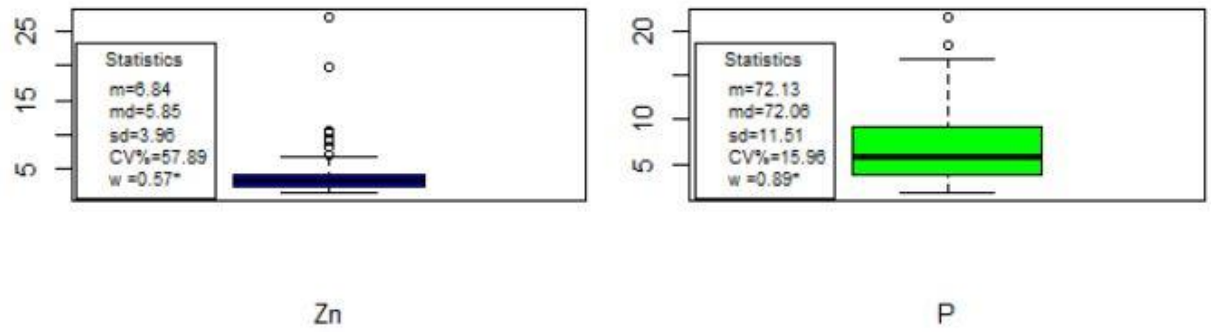
m - Mean; md - Median; sd - Standard deviation; CV% - Coefficient of variation (%); w - Shapiro-Wilk statistics for normality test; and * - Normality test significant at 5%.

Figure 2 - Boxplots of the chemical attributes: Fe - Iron, EC - electrical conductivity of saturation extract, OM - organic matter, and B - boron.



m - Mean; md - Median; sd - Standard deviation; CV% - Coefficient of variation (%); w - Shapiro-Wilk statistics for normality test; and * - Normality test significant at 5%.

Figure 3 - Boxplots of the physical (Clay), and chemical attributes: S - Sulphur, Mg - Magnesium and Cu - Cooper.



m - Mean; md - Median; sd - Standard deviation; CV% - Coefficient of variation (%); w - Shapiro-Wilk statistics for normality test; and * - Normality test significant at 5%.

Figure 4 - Boxplots of the chemical attributes: Zn - Zinc and P - Phosphorus.

The chemical attribute electrical conductivity of saturation extract (EC) was the only one following a distribution with an asymmetry coefficient close to zero and test for non-significant normality. The other attributes, besides presenting a significant test for normality, showed negative asymmetry coefficients (hydrogen potential, calcium and manganese) or positive asymmetry coefficients (phosphorus, potassium, magnesium, copper, iron, zinc, organic matter, sulfur, boron, and clay). However, interpolation was performed for all attributes since the assumption of normality is not a requirement for ordinary kriging.

The presence of outliers is a common characteristic between chemical attributes and the physical attribute considered in this work and may be a factor explaining the lack of symmetry observed. A relative similarity was observed in the dispersion of chemical and physical attributes, according to Figures 1 to 4, since the chemical attribute (zinc) has characteristics that make it a high dispersion attribute due to the presence of higher values of outliers.

Warrick & Nielsen (1980) proposed dispersion classes according to the value of the coefficient of variation expressed as a percentage (CV%). According to these authors, low dispersion occurs when CV%<12, medium dispersion when 12≤CV%≤60, and high dispersion when CV%>60. Therefore, a low dispersion was observed (Figures 1 to 4) for potential of hydrogen and clay and high dispersion was observed for zinc. The other attributes presented an average dispersion. Santos et al. (2017), evaluating the fertility of a soil cultivated with cacao, found similar values as this work. According to the authors, the soil chemical attributes tend to vary more than physical and physical-chemical ones, which helps to understand the soil system and adopt localized management practices. Similar results were also reported by Costa (2011) in a study aiming to evaluate the spatial and temporal variability of the apparent soil electrical conductivity.

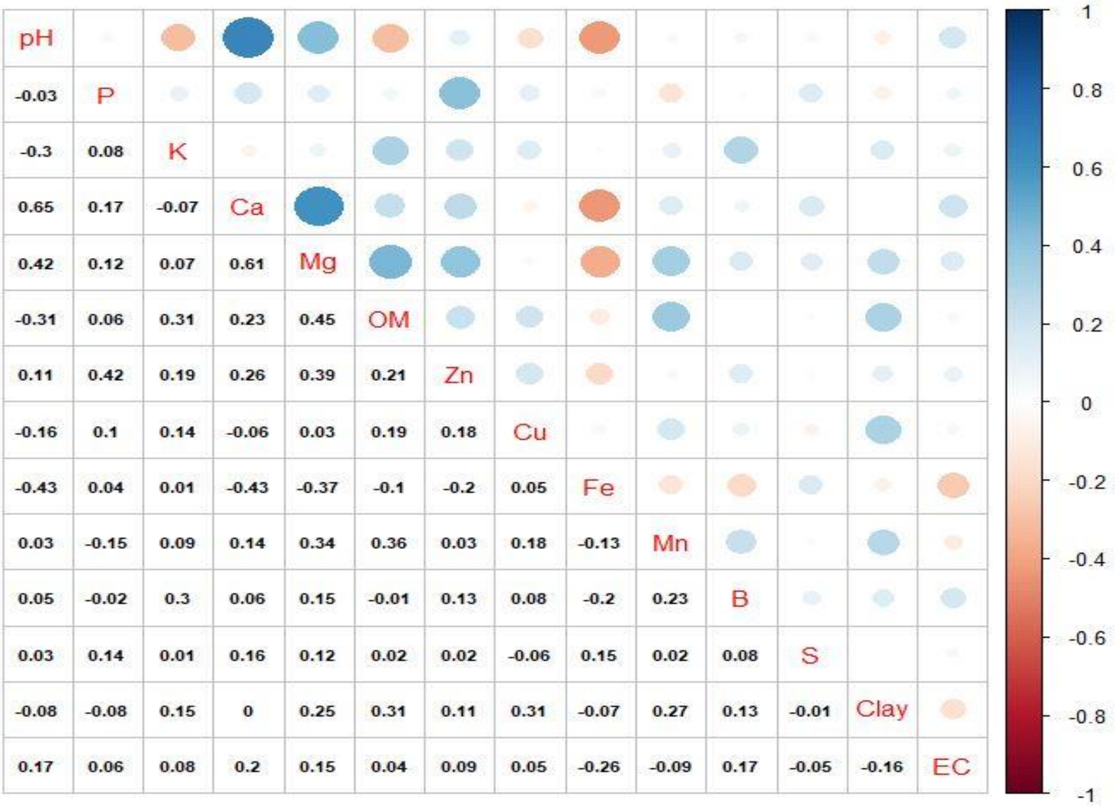
In general, according to the results of Figures 1 to 4, the soil of the studied area presents fertility restrictions for the development and production of different agricultural crops. The lack of natural fertility is a particular feature of Brazilian Latosols, characterized by depth of the diagnostic horizon, high weathering, and low natural fertility (Santos et al., 2017). Despite the natural characteristics of the soil, agronomic management practices can improve its fertility conditions, making the soils more fertile and allowing the exploitation of the productive potential of crops (Silva & Lima, 2014).

One of the assumptions for performing a multivariate analysis is the correlation between the attributes. Figure 5 shows the Pearson correlation matrix among the attributes evaluated. It shows a considerable number of chemical attribute pairs with moderate (0.40≤|r|≤0.59) and strong (0.60≤ |r|≤0.89) correlations. Figure 5 also shows that there are three chemical attributes with at least a moderate correlation with the calcium attribute, namely potential of hydrogen, magnesium, and iron. The correlation of Ca^{2+} with different chemical attributes shows that this is the most abundant cation in the soil solution, dominating the loads in the soil assortment complex. However, according to Albuquerque (2004), an at least moderately negative correlation between Ca^{2+} and Fe may be evidence of iron chlorosis in the study area, which occurs when there is iron deficiency in the soil.

At least moderate significant correlations between attributes indicate that the respective pair of attributes will have similar weights in the same main component.

Principal component analysis

The first five principal components (PC_j to $j=1,\cdots,5$), retained by the Kaiser criterion presented previously, explain 64.35% of the total variability of the 14 attributes (Table 1). The first principal component explains individually 21.32% of total variability, while the fifth explains only 8.01%. The coefficients associated with each of the 14 attributes in each PC_j, as well as the correlation of each coefficient with their attributes, are presented in Table 1.



pH - Potential of hydrogen; P - Phosphorus; K - Potassium; Ca - Calcium; Mg - Magnesium; OM - Organic matter; Zn - Zinc; Cu - Cooper; Fe - Iron; Mn - Manganese; B - Boron; S - Sulphur; and EC - Electrical conductivity of saturation extract.

Figure 5 - Matrix of correlations between chemical attributes and physical attribute measured at 150 points in the study area.

Table 1 - Eigenvector estimates and correlation of PC_j for $j = 1, \cdots, 5$, with soil chemical and physical attributes.

Attribute	Eigenvector					Correlation				
	PC_1	PC_2	PC_3	PC_4	PC_5	PC_1	PC_2	PC_3	PC_4	PC_5
pH	-0.30	0.48	-0.14	0.04	0.01	-0.22	0.38	-0.13	-0.02	0.32
P	-0.12	-0.03	0.65	0.14	0.09	-0.15	-0.03	0.62	-0.02	0.05
K	-0.10	-0.39	0.18	-0.34	-0.23	-0.42	-0.84	0.40	-0.59	-0.77
Ca	-0.44	0.26	0.02	0.17	-0.03	-0.43	0.13	0.02	-0.12	0.10
Mg	-0.49	0.00	-0.04	0.19	-0.02	-0.56	-0.13	-0.08	-0.12	-0.04
OM	-0.26	-0.40	-0.03	0.17	0.08	-0.42	-0.43	-0.07	-0.15	-0.27
Zn	-0.31	-0.10	0.43	0.01	0.20	-0.45	-0.18	0.38	-0.16	-0.07
Cu	-0.09	-0.34	0.04	-0.06	0.33	-0.21	-0.26	0.00	-0.07	-0.03
Fe	0.34	-0.20	0.17	0.28	-0.17	0.66	-0.18	0.15	0.35	-0.30
Mn	-0.23	-0.25	-0.41	0.11	-0.15	-0.54	-0.45	-0.66	0.12	-0.36
B	-0.18	-0.13	-0.05	-0.49	-0.47	0.39	-0.26	0.01	-0.29	-0.26
S	-0.06	0.01	0.19	0.36	-0.69	0.01	-0.04	0.13	0.18	-0.47
Clay	-0.16	-0.36	-0.26	0.13	0.13	-0.27	-0.42	-0.29	0.09	-0.07
EC	-0.17	0.14	0.16	-0.52	0.03	-0.51	-0.27	0.38	-0.81	0.13
Eigenvector	2.99	2.19	1.50	1.29	1.12					
s^2%	21.32	15.68	10.73	9.21	8.01					
s_j^2%	21.32	36.40	47.13	56.34	64.35					

PC_j - j-th principal component, so that j = 1, ···, 5; pH - potential of hydrogen; P - Phosphorus; K - Potassium; Ca - Calcium; Mg - Magnesium; OM - Organic matter; Zn - Zinc; Cu - Copper; Fe - Iron; Mn - Manganese; B - Boron; S - Sulfur; EC - Electrical conductivity of saturation extract; s^2% - Percentage of the proportion of variance explained by PC_j; and s_j^2% - Percentage of the proportion of explained and cumulative variance up to PC_j.

As expected, attributes with a high absolute value in a given PC_j also presented a large absolute value for the Pearson coefficient *r.* Similar results were obtained by Silva & Lima (2012) and Jimenez-Espinosa et al. (1993). In these two studies, the authors found that the attributes presented, in large part, both scores and correlations of great magnitude in the first principal component.

A large correlation value ($r > 0.5$) indicates the importance of a given attribute on the variability explained by that principal component. Thus, the signal and magnitude of these correlations allow us to group the attributes and interpret their contribution in PC_j. Thus, sample points that show a great magnitude for magnesium, manganese and electrical conductivity of the saturation extract contribute to decrease PC_1 scores. However, points of great magnitude, such as iron, help to increase the PC_1 score. The interpolation map of this component could, in principle, be used for the management of these elements in the focus area.

Higher scores of PC_3 indicate high phosphorus and low manganese soils. Interpolation maps plotted using this component can facilitate phosphorus fertilization and manganese management since in soils with predominance of low-crystallinity Mn oxides, manganic oxides tend to adsorb phosphorus (Gonçalves et al. 2011).

Electrical conductivity and potassium were the only attributes that showed $r > 0.5$ in PC_4 and, therefore, are responsible for explaining much of the variability summarized by this component. The attribute potassium was the only one to have $r > 0.5$ in components PC_2 and PC_5. Thus, as the variability explained by the potassium attribute will already be considered when obtaining management zones from component PC_4, management zones for components PC_2 and PC_5 will not be obtained, as they depend basically on the potassium attribute.

As mentioned earlier, each PC_j $j = 1,\ldots,5$, represents a group of chemical and physical attributes in the area. The PCA aims to capture the maximum variability of these attributes using a number of PC_j lower than the number of attributes. The components PC_j thus obtained can facilitate management by defining management zones using PC_j. The use of PCA from a geostatistical point of view was initially made by Davis & Greenes (1983) and later applied by other authors (Li et al., 2007; Silva et al., 2010a; Silva et al., 2010b; Barnett & Deutsch, 2012; Silva & Lima, 2012).

Analysis of spatial variability

The selected variogram models, as well as the cross-validation statistics for each PC_j and also for each attribute that presented $r > 0.5$ with the coefficients in each PC_j, are presented in Table 2.

For all attributes and for all PC_j, it was possible to fit a theoretical variogram model (spherical or exponential). This result indicates that the reduction in the dimensionality of the number of attributes performed by principal component analysis did not result in the impossibility of characterizing spatial variability. This indicates the potential of this methodology for simplifying analysis considering many attributes. However, it is noteworthy that the variogram model selected for a given PC_j did not always coincide with the model chosen for each of its attributes with the highest contribution. The results obtained in this research corroborate those found by Silva et al. (2010a) and Silva & Lima (2012), who also found spatial variability for principal components fitted using soil chemical, physical and physical-chemical attributes.

The analysis of the cross-validation results of the models fitted to PC_j was performed according to Webster & Oliver (1990), and consisted of selecting the models that present a mean standard error (MSE) close to zero and a root mean square of standardized errors (RMSSE) close to 1. We also analyzed the intercept ($\hat{\beta}_0$) and the slope ($\hat{\beta}_1$) estimates of the first-degree linear regression model adjusted to the values observed in function of values predicted, whose ideal values are 0 and 1, respectively, according to Vieira (2000). All models selected and presented in Table 2 met these cross-validation criteria. However, the attributes manganese and electrical conductivity of the saturation extract were $\hat{\beta}_0$ far from zero.

Several researches conducted mainly on precision agriculture used these cross-validation measurements. Burak et al. (2010) obtained results very close to those obtained in this study. Chen et al. (2018) also used MSE and RMSSE in a work aiming to map the spatial variability of soil chemical and physical attributes.

Outlining of management zones

The outlining of management zones is one of the most important phases in precision agriculture (Molin et al., 2015), as it allows a differentiated management at each point of the production area and optimizes productivity. Management zones were defined for each PC_j using the interpolated map of their scores.

For each class number, the values of two validation functions were calculated, namely FPI and MPE (Figure 6). The minimum FPI and MPE values for the principal components PC_1, PC_3 and PC_4 were 6, 3 and 3, respectively.

Using the Fuzzy c-means algorithm, management zones of each PC_j were outlined. They are presented in Figure 7.

Table 2 - Theoretical models, experimental semivariance parameter estimates, respective cross-validations of each of the main interpretable components, PC_j, j = 1, 3, 4, and the six chemical and physical attributes that showed the greatest contribution in PC_j.

PC_j or Attribute	Model	Parameters of semivariance			Cross-validation			
		a (m)	C0	C0+C1	MSE	RMSSE	$\hat{\beta}_0$	$\hat{\beta}_1$
PC_1	Spherical	460.00	8.00	24.00	0.00	0.90	1.99	1.07
PC_3	Spherical	349.00	18.00	45.00	0.00	1.01	-0.11	1.01
PC_4	Exponential	600.00	25.00	60.00	0.00	1.10	-0.86	0.98
Mg	Spherical	300.00	0.01	0.02	0.00	1.04	0.40	0.60
Fe	Spherical	500.00	7.00	30.00	0.00	1.00	-0.94	1.04
Mn	Spherical	373.00	30.00	90.00	0.00	0.72	-2.55	1.01
EC	Spherical	220.00	45.00	145.00	0.00	1.03	9.72	0.87
K	Spherical	600.00	80.00	220.00	0.00	1.23	0.74	0.99
P	Spherical	100.00	6.00	15.00	0.00	0.17	-0.29	1.04

PC_j - Principal component j, so that $j = 1, 3, 4$; Mg - Magnesium, Fe - Iron, Mn - Manganese, EC - Electrical conductivity of saturation extract, K - Potassium, P - Phosphorus, a - Range in meters, C0 - Nugget effect, C0 + C1 - threshold, MSE - mean standard errors, RMSSE - root mean square of standard errors; $\hat{\beta}_0, \hat{\beta}_1$ - Regression coefficients between the observed values and the values predicted by ordinary kriging.

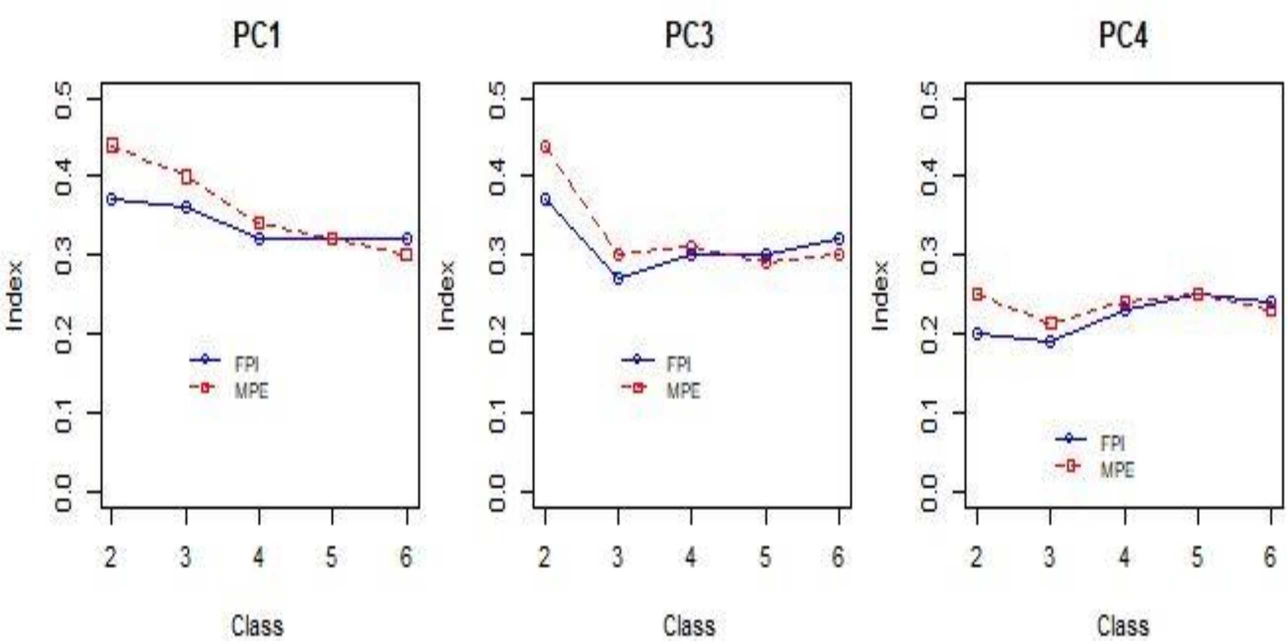


Figure 6 - Fuzzy performance index and modified partition entropy for the three interpretable main components PC_j, $j = 1, 3, 4$; FPI - Fuzzy performance index; MPE - Modified partition entropy.

A hypothesis test for the Kappa index was performed to assess agreement and significance (Congalton & Mead, 1986) among management zones obtained from PC_j and zones obtained from attributes with the highest contributions in PC_j (Table 3).

This agreement was observed among the management zones obtained for the chemical attributes iron, manganese and conductivity of the saturation extract and those obtained for PC_1. However, it is noteworthy that the attribute magnesium did not show such agreement. The management zones obtained for PC_3 had an agreement with zones obtained for manganese and non-agreement for the attribute phosphorus. Both attributes with the highest contribution in PC_4 had no agreement with the zones delineated for that principal component. Results similar as those of this work were obtained by Carvalho (2016) regarding the Kappa index variation. The author obtained Kappa indexes with a classification ranging from not significant to significant.

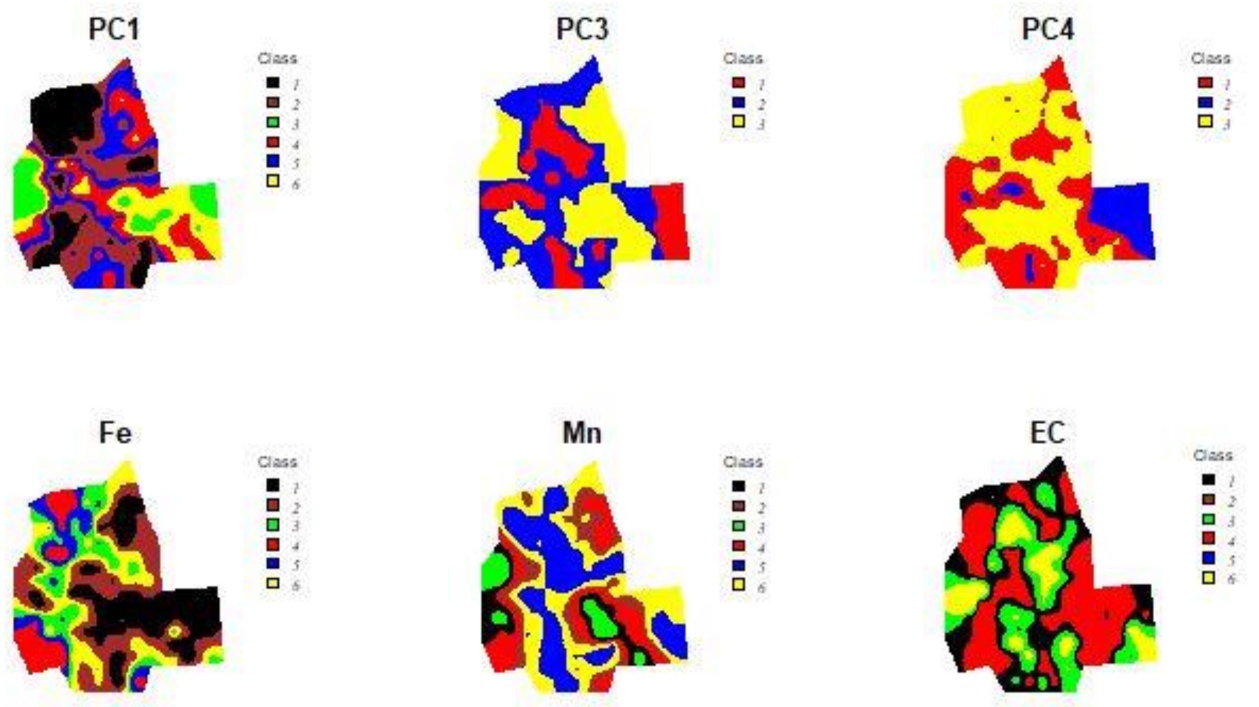
Table 3 - Kappa index of the classification between the three principal components and six chemical and physical attributes that presented the greatest contribution in the *PCj*.

PC_j	Attributes					
	Mg	Fe	Mn	EC	K	P
PC_1	-0.04	0.04*	0.04*	0.04*		
PC_3			0.06*			-0.06
PC_4				-0.03	-0.03	

PC_j: Principal component j, so that $j = 1, 3, 4$; Mg: Magnesium, Fe: Iron; Mn: Manganese; EC: Electrical conductivy of saturation extract; K: Potassium; P: Phosphorus; and *: significant to 5%.

For some attributes with a large contribution in a PC_j, the absence of a significant agreement does not compromise the results of this work. There is a possibility of using this technique in precision agriculture, as has been reported in the literature (Li et al., 2007; Molin & Castro, 2008). These authors considered a great correlation that existed between the attributes and their respective PC_j as one of the factors that could explain the greater interest in PC for the outlining of management zones.

According to Fridgen et al. (2004), the agreement of FPI and MPE indexes in each PC_j is an indication of a good classification. The interpolated maps with their respective management zones defined for each PC_j, $j = 1, 3, 4$, are shown in Figure 7. This Figure shows spatial pattern differences between the management zones outlined for each PC_j, a fact that can be explained by the different contributions of the attributes in each PC_j. The management zones obtained from interpolated maps of PC_1 show a more irregular spatial distribution pattern than those obtained from the maps of PC_3 and PC_4, a fact that can be explained by the higher number of management zones outlined based on the principal component PC_1. In addition, the regularity of the spatial distribution pattern of PC_j management zones was similar as that of the attributes with $r > 0.5$. An example of this result is the similarity of the pattern of spatial distribution of the attribute iron and the PC_1. A similar result was also obtained by Rodrigues & Corá (2015) in a work that aimed to identify management zones using the algorithm of fuzzy c-means clustering based on spatial and temporal variability of soil attributes and corn yield.



Fe - Iron; Mn - Manganese; EC - Electrical conductivy of saturation extract

Figure 7 - Management zones defined by the Fuzzy c-means method from the spatial variability maps of the principal component PC_j, so that $j = 1, 3, 4$, and maps of the three chemical attributes that showed a significant agreement with at least one principal component.

Conclusions

The multivariate analysis of principal components allows reducing the dimensionality of the number of soil chemical attributes, which results in components susceptible to agronomic interpretation.

It is possible to jointly describe the pattern of spatial distribution of chemical, physical and physical-chemical soil attributes using the scores of principal components.

The outlining of management zones using principal components did not present consistent results that allow the immediate use of techniques for precision agriculture.

Although there was no complete agreement between uni- and multivariate management zones outlined, the management of each chemical and physical attribute can in principle be performed within its management zone, as each attribute contributes with a fraction of the variance of the principal component. Therefore, further studies are needed.

Acknowledgements

To the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code-001, for granting financial support for this research and to the ad hoc rapporteurs of this article for their valuable suggestions and corrections.

References

Albuquerque TCSD (2004) Adubação mineral da videira. In: Feira Nacional da Agricultura Irrigada-Fenagri, 2004, Petrolina. Minicursos: apostilas. Petrolina: Embrapa Semi-Árido.

Barnett RM, Deutsch CV (2012) Practical implementation of non-linear transforms for modeling geometallurgical variables. In P. Abrahamsen, R. Hauge, & O. Kolbjornsen (Eds.). Geostatistics 409–422.

Beckett PHT, Webster R (1971) Soil variability – a review. Soils and Fertilizers 34:1–15.

Burak DL, Fontes MPF, Santos NT, Monteiro LVS, Martins ES, Becquer T (2010) Geochemistry and spatial distribution of heavy metals in Oxisols in a mineralized region of the Brazilian Central Plateau. Geoderma 160(2):131-142.

Carvalho PSM, Franco LB, Silva AS, Sodré GA, Queiroz DM, Lima JSS (2016) Cacao crop management zones determination based on soil properties and crop yield. Revista Brasileira de Ciência do Solo 40:1-17.

Chen S, Liang Z, Webster R, Zhang G, Zhou Y, Teng H, Hu B, Arrouays D, Shi Z (2018) A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution. Science of The Total Environment 665(10):273-283.

Congalton RG, Mead RA (1986) A review of discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrices. EEE Transactions on Geoscience and Remote Sensing 24:169-174.

Costa MM (2011) Condutibilidade Elétrica Aparente do Solo, como Ferramenta para Agricultura de Precisão em uma área sob Cerrado. Departamento de Engenharia Agrícola, Universidade Federal de Viçosa (Tese do Mestrado em Agronomia).

Davis BM, Greenes KA (1983) Estimating using spatially distributed multivariate data: An example with coal quality Mathematical Geology 15:287–300.

Dunteman GH (1989) Quantitative Applications in the Social Sciences: Principal components analysis. Newbury Park, CA: SAGE Publications.

Fridgen JJ, Kitchen NR, Sudduth KA, Drummond ST, Wiebold WJ, Fraisse CW (2004) Management zone analyst (MZA): software for subfield management zone delineation. Agronomy Journal 96(1): 100-108.

Gonçalves GK, Meurer EJ, Bortolon L, Gonçalves DRN (2011) Relação entre óxidos de ferro e de manganês e a sorção de fósforo em solos no Rio Grande do Sul. Revista Brasileira de Ciência do Solo 35(5):1633-1639.

Gorsevski PV, Gessler PE, Jankowski P (2003) Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. Journal of Geographical Systems 5(3):223-251.

Hartigan JA (1975) Clustering algorithms. United States of America. 369p.

Inamasu RY, Bernardi ACC (2014) Agricultura de precisão: resultados de um novo olhar. Brasília: Embrapa. 600p.

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. New York: Oxford University Press. 561p.

Jimenez-Espinosa R, Sousa AJ, Chica-Olmo M (1993) Identification of geochemical anomalies using principal component analysis and factorial kriging analysis. Journal of Geochemical Exploration 46(3):245-256.

Johnson RA, Wichern DW (2002) Applied multivariate statistical analysis. 5thEdition. New Jersey. Prentice Hall 2002, 767p.

Kaiser HF (1960) The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20:141-151.

Li Y, Shi Z, Li F, Li HY (2007) Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. Computers and Electronics in Agriculture 56(2):174-186.

Mingoti AS (2005) Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada. Belo Horizonte: Editora UFMG, 300p.

Molin JP, Amaral LRD, Colaço AF (2015) Agricultura de Precisão. 1ª Ed. São Paulo: Oficina de Textos, 238p.

Molin JP, Castro CND (2008) Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. Scientia Agricola 65:567-573.

Oldoni H, Bassoi L (2016) Delineation of irrigation management zones in a Quartzipsamment of the Brazilian semiarid region. Pesquisa Agropecuária Brasileira 51(9):1283-1294.

Ribeiro Júnior JI (2012) Métodos Estatísticos Aplicados à Melhoria da Qualidade. Viçosa-MG, 385p.

Rodrigues MS, Corá JE (2015) Management zones using fuzzy clustering based on spatial-temporal variability of soil and corn yield. Engenharia Agrícola 35:470-483.

Rodrigues MS, Santana MC, Uchôa ALP, Menezes AXSMM, Cavalcante IHL, Lima AMN (2015) Delineamento de zonas de manejo baseadas em atributos físicos do solo em uma área de goiabeira irrigada no Semiárido nordestino. Anais do XXXV Congresso Brasileiro de Ciência do Solo. Centro de Convenções-Natal: Sociedade Brasileira de Ciências de Solo. 4p.

Santos RO, Franco LB, Silva AS, Sodré GA, Menezes AA (2017) Spatial variability of soil fertility and its relation with cocoa yield. Revista Brasileira de Engenharia Agrícola e Ambiental 21(2):88-93.

Shapiro SS, Wilk MB (1968) Approximations for the null distribution of the W statistic. Technometrics 10:861-866.

Silva SA, Lima JSS (2014) Spatial estimation of foliar phosphorus in different species of the genus coffea based on soil properties. Revista Brasileira de Ciência do Solo 38(5):1439-1447.

Silva SA, Lima JSS, Souza GS (2010b) Estudo da fertilidade de um Latossolo Vermelho-Amarelo húmico sob cultivo de café arábica por meio de geoestatística. Revista Ceres 57(4):560-567.

Silva SA, Lima JSDS, Souza GS, Oliveira RB, Silva AF (2010a) Variabilidade espacial do fósforo e das frações granulométricas de um Latossolo Vermelho Amarelo. Revista Ciência Agronômica 41(1):1-8.

Silva SA, Lima JSDS (2012) Multivariate analysis and geostatistics of the fertility of a humic rhodic hapludox under coffee cultivation. Revista Brasileira de Ciência do Solo 36(2):467-474.

Silva SA, Lima JSDS, Xavier AC, Teixeira MM (2010c) Variabilidade espacial de atributos químicos de um Latossolo Vermelho-Amarelo húmico cultivado com café. Revista Brasileira de Ciência do Solo 34 (1): 15-22.

Stafford JV, Ambler B, Lark RM, Catt J (1996) Mapping and interpreting yield variation in cereal crops. Computer and Electronics in Agriculture 14:101–119.

Sylvester-Bradley R, Lord E, Sparkes DL, Scott RK, Wiltshire JJJ, Orson J (1999) An analysis of the potential of precision farming in Northern Europe. Soil Use and Management 15:1-8.

Vieira SR (2000) Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F. et al (Ed). Tópicos em ciência do solo 1:1-54.

Wackernagel H (2003) Multivariate geostatistics: An introduction with applications. Springer-verlag berlin heidelberg. 388 p.

Warrick AW, Nielsen DR (1980) Spatial variability of soil physical properties in the field. In: HILLEL, D., ed. Applications of soil physics. p.319-344.

Webster R, Oliver MA (1990) Statistical methods in soil and land resource survey. Oxford University Press. 269p.

Webster R, Oliver MA (2007) Geostatistics for environmental scientists. 2nd, John Wiley & Sons. 330p.