Universidade Federal Rural do Semi-Árido
Pró-Reitoria de Pesquisa e Pós-Graduação
https://periodicos.ufersa.edu.br/index.php/caatinga

ISSN 0100-316X (impresso)
ISSN 1983-2125 (online)
http://dx.doi.org/10.1590/1983-21252022v35n420rc

MANAGEMENT ZONES DESIGN FOR SOYBEAN CROP USING PRINCIPAL COMPONENTS AND GEOSTATISTICS[1]

RICARDO NIEHUES BUSS[2], RAIMUNDA ALVES SILVA[2], OSVALDO GUEDES FILHO[3], GLÉCIO MACHADO SIQUEIRA[2]*

ABSTRACT - In precision agriculture, determining management zones for soil and plant attributes is a complex process that requires knowledge of several variables, which complicates management and decision-making processes. This study evaluated the spatial variability of soybean yield and soil chemical properties using geostatistical and multivariate analyses to define management zones in an Oxisol. The soybean yield and soil chemical properties between 0 to 0.2 and 0.2 to 0.4 m soil depths were sampled at 70 points. Geostatistical and multivariate analyses were then performed on these data. The soil chemical properties showed higher variability at 0.2 to 0.4 m soil depth. The semivariogram parameters of the principal component analysis (PCA) data (PCA 1, PCA 2, and PCA 3) for both depths were more homogeneous than the original data. The maps of soil chemical properties showed high similarity to the soybean yield map. The PCA explained 65.34% (0 to 0.2 m) and 70.50% (0.2 to 0.4 m) of data variability, grouping the soybean yield, organic matter, pH, phosphorous, potassium, calcium, magnesium, and sodium. PCA spatialization allowed for the definition of management zones indicated by PCA 1, PCA 2, and PCA 3 for both depths. The result indicates that the area must be managed using different strategies of soil fertility management to increase soybean yield.

Keywords: Principal components analysis. Semivariogram. Soil chemical properties. Crop yield. Precision agriculture.

DELINEAMENTO DE ZONAS DE MANEJO PARA A CULTURA DA SOJA POR MEIO DE COMPONENTES PRINCIPAIS E GEOESTATÍSTICA

RESUMO - Na agricultura de precisão a determinação de zonas de manejo dos atributos de solo e planta, é um processo complexo que demanda o conhecimento de muitas variáveis, o que dificulta o processo de gestão e tomada de decisão. Este estudo avaliou a variabilidade espacial da produtividade da cultura da soja e de atributos químicos do solo por meio de análise multivariada e geoestatística para a determinação de zonas de manejo específico em um Latossolo. A produtividade de soja e os atributos químicos do solo nas camadas 0-0.2 e 0.2-0.4 m de profundidade foram amostrados em 70 pontos de amostragem. Análises geoestatísticas e multivariadas foram então realizadas. As propriedades químicas do solo apresentaram maior variabilidade na profundidade de 0,2 a 0,4 m. Os parâmetros do semivariograma dos dados da análise de componentes principais (PCA) (PCA 1, PCA 2 e PCA 3) para ambas as profundidades foram mais homogêneos do que os dados originais. Os mapas de propriedades químicas do solo apresentaram alta similaridade com o mapa de produtividade da soja. A ACP explicou 65,34% (0 a 0,2 m) e 70,50% (0,2 a 0,4 m) da variabilidade dos dados, agrupando a produtividade da soja, matéria orgânica, pH, fósforo, potássio, cálcio, magnésio e sódio. A espacialização do PCA permitiu a definição das zonas de manejo indicadas pelo PCA 1, PCA 2 e PCA 3 para ambas as profundidades. O resultado indica que a área deve ser manejada utilizando diferentes estratégias de manejo da fertilidade do solo para aumentar a produtividade da soja.

Palavras-chave: Análise de componentes principais. Semivariograma. Propriedades químicas do solo. Produtividade das culturas. Agricultura de precisão.

*Corresponding author
[1]Received for publication in 02/27/2021; accepted in 06/23/2022.
Paper extracted from the doctoral thesis of the first author.
[2]Postgraduate Program in Biodiversity and Biotechnology in the Amazon, Universidade Federal do Maranhão, São Luís, MA, Brazil; rnbuss@gmail.com - ORCID: 0000-0003-3444-4243, ray-234@hotmail.com – ORCID: 0000-0002-0380-8190, glecio.siqueira@ufma.br – ORCID: 0000-0002-3513-2658.
[3]Soil laboratory, Universidade Federal do Paraná, Jandaia do Sul, PR, Brazil; osvaldoguedes@yahoo.com.br – ORCID: 0000-0001-8550-8505.

INTRODUCTION

In Brazil, agribusiness makes a significant contribution to the national economy, with soybeans as the main agricultural commodity. The mean soybean yield in Brazil is 3,000 kg ha^{-1}, cultivated in an area of 41 million hectares. However, in the state of Maranhão alone, the mean yield is 3,300 kg ha^{-1} cultivated in 1 million hectares (CONAB, 2022).

Several studies have examined the spatial variability in soybean yield (CAMBARDELLA et al., 1994; SILVA et al., 2010; VIEIRA et al., 2010; LIMA et al., 2013; SIQUEIRA et al., 2015a; GAVIOLI et al., 2016; BUTTAFUOCO et al., 2017; FREDDI et al., 2017; BUSS et al., 2019; JIANSHU, 2019). According to Vieira (2000), some soil natural variability always exists but may be altered by soil use and management. Siqueira et al. (2015a) stated that fertilizer application in agriculture increases the variability of soil properties and that management zone definition requires the use of methods involving a greater number of properties that vary in time and space.

GuedesFilho et al. (2010) found a common standard of spatial variability when evaluating crop yield maps in long-term experiments. However, in some years, the spatial variability of crop yield was aleatory. Vieira et al. (2010) reported that the spatial variability of nutrient export by plants is not homogeneous, although soil fertility is also managed. However, it is necessary to understand the spatial variability dynamics of crop yields and soil chemical properties using mathematical models that can integrate a greater number of variables.

Multivariate analysis describes the variance and covariance structure of variable groups by constructing linear combinations and reducing the number of dimensions (JEFFERS, 1978; SILVA et al., 2010; LIMA et al., 2013; BUSS et al., 2019). Therefore, studies involving the application of geostatistical and multivariate techniques provide a multidimensional understanding and higher accuracy in defining management zones. Alarcón-Jiménez et al. (2015) used a multivariate analysis of soil physical properties and corn yield to establish management zones. Córdoba et al. (2016) presented a geostatistical and multivariate analysis protocol to determine management zones using soil and crop yield data.

Multiple approaches (geostatistical and multivariate analysis) have been used by researchers to describe the spatial variability in plants and soils (SILVA et al., 2010; SILVA; LIMA, 2012; LIMA et al., 2013; GAVIOLI et al., 2016; BUTTAFUOCO et al., 2017; FREDDI et al., 2017; MASOUD et al., 2018; OUMENSKOU et al, 2018; URIBEETXEBARRIA et al., 2018; BUSS et al., 2019), demonstrating its importance.

The objective of this study was to evaluate the spatial variability in soybean yield and soil chemical properties using geostatistical and multivariate analyses to define management zones.

MATERIAL AND METHODS

The study area is in Mata Roma municipality, Maranhão State, Brazil (3º 70’ 80.88″ S e 43º 18’ 71.27″ W), with a median altitude of 103 m. The regional climatic classification is Aw, warm and humid, with two well-defined seasons: rainy (December–May) and dry (June–November). The annual mean precipitation is 1,835 mm, and the mean temperatures in summer and winter are 26.5 and 28 °C, respectively (Figure 1a).

The soil area is Oxisol (SANTOS, et al., 2018), which has a clayey texture. The soil characteristics are presented in Table 1. The area has been cultivated with soybean (*Glycine max* L.) and corn (*Zea mays* L.) in crop rotation without irrigation under no-tillage since 2008. The total area is 44.75 ha (Figure 1b). The crop yield and soil were sampled after soybean harvest (2015/2016) at 70 regular sampling areas of 100 × 35 m. Sampling points were referenced using a static GPS with post-processed differential correction (static DGPS) (Figure 1b).

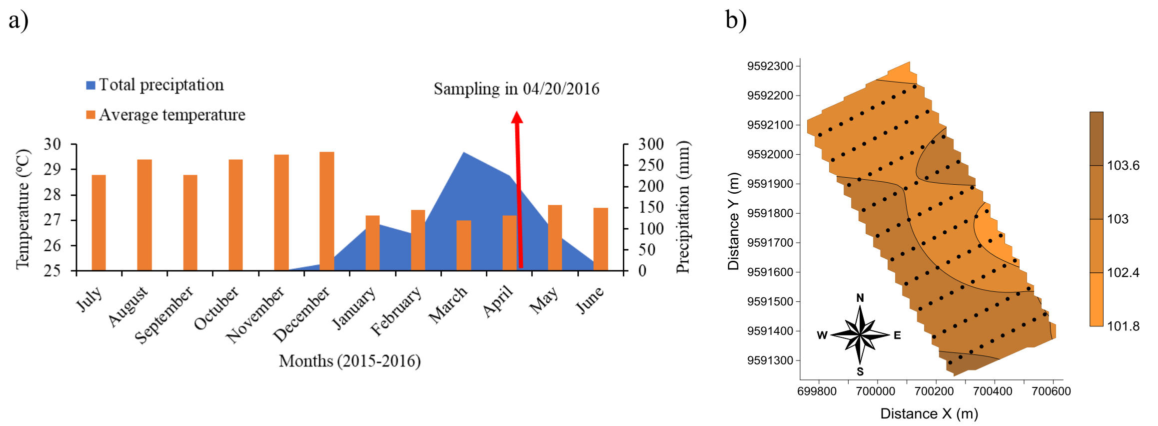

Figure 1. Climate parameters (a) and map with level difference and location of sampling points in area (b).

Table 1. Physical and chemical characterization of the soil of the area cultivated with soybean under no-tillage.

0–0.2 m													
Sand	Silt	Clay	BD	Macro	Micro	TP	OM	pH	P	K	Ca	Mg	CEC
-------- g kg^{-1} -----------			Mg gm^{-3}	--------- m^3 m^{-3} ---------			g dm^{-3}		mg dm^{-3}	---------- mmol$_c$ dm^{-3} ----------			
745.58	138.21	117.14	1.27	0.17	0.38	0.55	22	5	49	0.7	18	3	46.7
0.2–0.4 m													
Sand	Silt	Clay	SD	Macro	Micro	TP	OM	pH	P	K	Ca	Mg	CEC
-------- g kg^{-1} ---------			Mg m^{-3}	--------- m^3 m^{-3} ----------			g dm^{-3}		mg dm^{-3}	---------- mmol$_c$ dm^{-3} ----------			
737.77	141.70	120.63	1.29	0.16	0.37	0.53	19	4.7	47	0.5	17	3	45.6

BD, bulk density; Macro, macroporosity; Micro, microporosity; TP, total porosity; OM, organic matter; CEC, cation exchange capacity.

Soybean yield (kg ha^{-1}) was determined on April 20, 2016 in plots of 18 m^2. After harvest, the grains were oven-dried at 65 °C and weighed after attaining a constant mass.

Disturbed soil was sampled at soil depths between 0 to 0.2 and 0.2 to 0.4 m to determine the following chemical properties: organic matter (OM, g dm^{-3}), pH $CaCl_2$, potential acidity (H+Al, mmol$_c$ dm^{-3}), phosphorus (P, mg dm^{-3}), potassium (K, mmol$_c$ dm^{-3}), calcium (Ca, mmol$_c$ dm^{-3}), magnesium (Mg, mmol$_c$ dm^{-3}), sodium (Na, mmol$_c$ dm^{-3}), cation exchange capacity (CEC, mmol$_c$ dm^{-3}), base sum (BS, mmol$_c$ dm^{-3}), base saturation (V%), copper (Cu, mg kg^{-1}), iron (Fe, mg kg^{-1}), manganese (Mn, mg kg^{-1}), and cadmium (Cd, mg kg^{-1}), following procedures mentioned in Raij et al. (2001).

Data were analyzed using descriptive statistics with the help of R 3.3.1 (R CORE TEAM, 2018), and the following measures were determined: mean, variance, standard deviation, coefficient of variation, asymmetry, kurtosis, and D (maximum deviation in relation to normal distribution, using the Kolmogorov-Smirnov test with error probability of 0.01).

The assumptions of the intrinsic hypothesis of geostatistics were considered for modelling and adjusting the experimental semivariogram according to Vieira (2000), obtaining the following parameters using the jackknifing technique: C_0 (nugget effect), $C_0 + C_1$ (sill), and a (range, m).

Scaled semivariograms were adjusted to evaluate the spatial variability pattern of pairs of variable variances (Equation 1) (VIEIRA et al., 1997). This allows for the overlap of variables with different scalar magnitudes by the standardization of semivariance pairs, as described by Siqueira et al. (2015b).

$$y^{sc}(h) = \frac{y(h)}{Var\ (z)} \qquad (1)$$

where:
$y^{sc}(h)$ is the scaled semivariogram;
$y(h)$ is the original semivariogram;
$Var(h)$ is the data variance.

The spatial dependence ratio (SDR) among the samples was determined as previously described by Cambardella et al. (1994), which is classified as low (75–100%), medium (25–75%), and high (0–25%).

Multivariate analysis (principal component analysis, PCA) was used to analyze soil and plant data, taking into account the null mean and unitary variance, using the software Statistica 12.0 (STATSOFT, 2015). Data were initially standardized (mean = 0 and standard deviation = 1) to the original data of the same magnitude, once all evaluated variables had different orders of magnitude. With this standardization, it was possible to perform the PCA.

PCA was performed using the correlation matrix between standardized variables (JEFFERS, 1978), allowing collinearity determination, which contributes to reducing variable dimensionalities (BUSS et al., 2019) by the orthogonal linear recombination of variables (JEFFERS, 1978). From the correlation matrix between standardized variables, the place of each variable with collinearity was determined in a covariance matrix.

For the PCA biplot graph, a set of eigenvectors (PC 1, PC 2,..., PC h) was included, which explained more than 60% of the data variability, and the variance of each main component was calculated using Equation 2:

$$CP_h = \frac{\lambda_h}{(C)} 100 \qquad (2)$$

where:
CP_h = principal component h;
λ_h = eigenvalue h; and,
C = covariance matrix; trace $(C) = \lambda_1 + \lambda_2 + ... + \lambda_h$.

For each eigenvalue (PC1, PC2, ..., PC$_h$) that explained more than 60% of the data variability, outliers were checked using the Hotelling T^2

technique. The validation was performed considering the quadratic sum of the forecast errors [SPE(Q)]. After validating the PCA, the scores of each eigenvalue were determined to further examine spatial variability using the experimental semivariogram and scaled semivariogram adjustment. Maps of spatial variability were obtained using kriging interpolation and performed on SURFER 12 (GOLDEN SOFTWARE, 2014).

RESULTS AND DISCUSSION

The mean yield in the area was 3,770.01 kg ha^{-1} (Table 2), and is 13.21% above the mean yield of Maranhão (3,330 kg ha^{-1}) and 25.66% above that of the national mean yield (3,000 kg ha^{-1}) (CONAB, 2022). Our results were higher than those of Freddi et al. (2017), who found a soybean yield of 3,280 kg ha^{-1} in an Oxisol located in the Cerrado and Amazônian Forest ecotones.

The organic matter content was higher in the surface layer (12.64 g dm^{-3}) than in the subsurface (11.08 g dm^{-3}). In no-tillage, the high values of OM in the surface layer are attributed to the absence of mobilization and accumulation of crop residues during cultivation. Our OM results corroborate those found by Lima et al. (2013), Buttafuoco et al. (2017), Freddi et al. (2017), Silva and Siqueira (2020), and Siqueira et al. (2022).

Table 2. Descriptive statistics of soil chemical attributes at 0.0–0.2 m and 0.2–0.4 m depth cultivated with soybean.

	Mean	Variance	SD	CV (%)	Skew	Kurtosis	D*
				0.0–0.2 m			
Soybean yield	3770.71	189447	435.25	11.54	0.11	−0.50	0.065n
OM	12.64	35.36	5.95	0.47	1.15	1.43	0.194Ln
pH	5.12	0.34	0.58	0.11	−0.09	−0.05	0.074n
P	10.87	55.21	7.43	0.68	2.21	6.80	0.193n
K	1.88	2.01	1.42	0.75	1.42	0.51	0.329Ln
Ca	14.51	27.08	5.20	0.36	2.44	9.53	0.191n
Mg	4.94	9.56	3.09	0.63	0.59	−0.38	0.185n
H+Al	20.77	16.50	4.06	0.20	−0.15	−0.46	0.101n
Na	3.53	0.50	0.71	0.20	0.46	−0.62	0.101n
CEC	45.63	69.92	8.36	0.18	1.74	4.34	0.161n
BS	24.86	64.71	8.04	0.32	1.36	2.77	0.121n
V%	53.70	93.67	9.68	0.18	0.02	−0.52	0.054n
Cu	0.11	0.01	0.08	0.71	0.80	−0.15	0.154n
Fe	15.15	136.33	11.68	0.77	1.51	4.25	0.103n
Mn	0.44	0.11	0.33	0.75	0.72	−0.19	0.129n
Cd	0.01	0.00	0.01	0.74	0.66	−0.22	0.141n
				0.2–0.4 m			
OM	11.08	15.83	3.98	0.36	1.45	1.53	0.189n
pH	4.76	0.18	0.43	0.09	0.02	−0.14	0.123n
P	11.34	190.98	13.82	1.22	3.58	14.87	0.295Ln
K	1.31	0.22	0.47	0.36	4.68	29.93	0.224n
Ca	13.40	22.42	4.73	0.35	0.78	0.25	0.125n
Mg	4.35	6.08	2.47	0.57	0.37	−0.45	0.135n
H+Al	23.24	35.88	5.99	0.26	0.59	−0.32	0.091n
Na	3.61	0.47	0.69	0.19	0.15	−0.90	0.099n
CEC	45.78	50.19	7.08	0.15	0.62	−0.19	0.111n
BS	22.54	37.88	6.15	0.27	0.79	0.42	0.105n
V%	49.15	108.10	10.40	0.21	0.27	−1.27	0.140n
Cu	0.18	0.02	0.15	0.83	4.24	23.67	0.265Ln
Fe	33.83	200.92	14.17	0.42	1.36	2.00	0.183n
Mn	0.34	0.09	0.30	0.87	3.07	12.51	0.212n
Cd	0.01	0.00	0.01	0.71	0.59	−0.39	0.087n

OM: organic matter (g dm^{-3}); pH in $CaCl_2$; P: phosphorus (mg dm^{-3}); Cu: copper (mg dm^{-3}); Fe: iron (mg dm^{-3}); Mn: manganese (mg dm^{-3}); Cd: cadmium (mg dm^{-3}); K: potassium (mmol$_c$ dm^{-3}); Ca: calcium (mmol$_c$ dm^{-3}); Mg: magnesium (mmol$_c$ dm^{-3}); H+Al: hydrogenj + aluminum (mmol$_c$ dm^{-3}); Na: sodium (mmol$_c$ dm^{-3}); CEC: cation exchange capacity (mmol$_c$ dm^{-3}); BS: base sum (mmol$_c$ dm^{-3}); V%: base saturation (%); n: normal frequency distribution and Ln: log-normal frequency distribution, according to the Kolmogorov-Smirnov test at $p \leq 0.01$ (D).

The pH and macro and micronutrients presented high values in surface layer than in subsurface (Table 2), which reflects the concentration of fertilizers and correctives in this layer in a no-tillage system (LIMA et al., 2013). The CEC values for both depths indicated the area with media/high fertility, according to Sobral et al. (2015) (CEC > 40%).

The chemical attributes showed low coefficient of variation (CV < 12%) (WARRICK; NIELSEN, 1980) indicating a low variation in the area, which does not mean that there is no spatial variability in the area. Geostatistical analysis allows us to describe the scales of variability that are not detected by classic statistics. Asymmetry, kurtosis, and Kolmogorov-Smirnov tests indicated that chemical attributes (both depths) and yield had a generally normal frequency distribution.

The variables H+Al, CEC, and Cd at 0–0.2 m soil depth and P, H+Al, BS, V%, Cu Mn and Cd at 0.2–0.4 m soil depth presented a pure nugget effect (PNE) (Table 3), which means that the spatial variability of these variables occurred at a lower scale than the smaller distance between the sampling points. The spherical model was adjusted for soybean yield, OM, Ca, and Fe at the 0–0.2 m soil depth, while pH, Na, V%, and Mn were adjusted to the exponential model. The other variables were adjusted using a Gaussian model. At the 0.2–0.4 m soil depth, the exponential model was adjusted to OM, K, Ca, Na, CEC, and Fe, whereas the Gaussian model was adjusted to pH and Mg. Similar results were reported by Lima et al. (2013), Tripathi et al. (2015), and Bitencourt et al. (2016).

Table 3. Semivariogram adjustment parameters for soybean yield and soil chemical properties at 0–0.2 m and 0.2–0.4 m soil depth under no-tillage system.

Variables	Model	C_0	C_0+C_1	a	r^2	RSS	SDR
			0–0.2 m				
Soybean yield	Spherical	145000	250160	200	0.724	35.6	57.96
OM	Spherical	7.1	39.65	265	0.818	293	17.91
pH	Exponential	0.0369	0.338	220	0.243	0.752	10.92
P	Gaussian	0.100	48.88	51	0.752	391	0.20
K	Gaussian	0.263	2.046	52	0.714	0.653	12.85
Ca	Spherical	0.01	31.42	72	0.220	564	0.03
Mg	Gaussian	0.710	9.68	46	0.772	8.73	7.33
H+Al			Pure nugget effect				
Na	Exponential	0.001	0.534	50	0.772	0.028	0.19
CEC			Pure nugget effect				
BS	Gaussian	0.100	67.02	46	0.614	1175	0.15
V%	Exponential	12.2	95.1	20	0.089	1939	12.83
Cu	Gaussian	0.00001	0.0056	66	0.865	3.94E-06	0.18
Fe	Spherical	8.8	147.7	194	0.832	2844	5.96
Mn	Exponential	0.0001	0.124	117	0.927	8.460E-04	0.08
Cd			Pure nugget effect				
			0.2–0.4 m				
MO	Exponential	1.14	17.63	109	0.878	37.9	6.47
pH	Gaussian	0.0001	0.179	47	0.787	3.03E-03	0.06
P			Pure nugget effect				
K	Exponential	0.0001	0.179	49.9	0.406	0.0105	0.06
Ca	Exponential	2.31	22.22	36.1	0.814	20.4	10.40
Mg	Gaussian	0.49	5.705	57	0.456	4.16	8.59
H+Al			Pure nugget effect				
Na	Exponential	0.184	0.491	63	0.635	0.0382	37.47
CTC	Exponential	0.5000	50.52	49	0.876	258	0.99
SB			Pure nugget effect				
V%			Pure nugget effect				
Cu			Pure nugget effect				
Fe	Exponential	22	201.9	28	0.247	7480	10.90
Mn			Pure nugget effect				
Cd			Pure nugget effect				

OM: organic matter (g dm^{-3}), pH $CaCl_2$, H+Al: potential acidity ($mmol_c$ dm^{-3}), P: phosphorous (mg dm^{-3}), K: potassium ($mmol_c$ dm^{-3}), Ca: calcium ($mmol_c$ dm^{-3}), Mg: magnesium ($mmol_c$ dm^{-3}), Na: sodium ($mmol_c$ dm^{-3}), CEC: cationic exchange capacity ($mmol_c$ dm^{-3}), BS: base sum ($mmol_c$ dm^{-3}), V: base saturation (%), Cu: copper (mg kg^{-1}), Fe: iron (mg kg^{-1}), Mn: manganese (mg kg^{-1}), Cd: cadmium (mg kg^{-1}), C_0: nugget effect, C_0+C_1: sill, a: range (m), r^2: coefficient of regression, RSS: residual sums of squares; SDR: spatial dependence ratio (%).

Soybean yield had a range of 200 m, while OM, pH, P and Mn at 0–0.2 m soil depth, and OM at 0.2–0.4 m soil depth presented a range of 109–265 m. The other variables showed a range of < 72 m. The range represents the distance at which sampling points are spatially dependent on each other (VIEIRA, 2000). The variables exhibited higher data variability at 0.2–0.4 m soil depth than at 0–0.2 m. The soil chemical properties showed high soybean yield media spatial dependence ratios (Table 3). Bitencourt et al. (2016) found similar values of the spatial dependence ratio when studying the chemical and physical properties of soil.

PCA at the 0–0.2 m soil depth had three components that explained 65.34% of the total variability of the data (PCA 1 = 31.16%, PCA 2 = 20.87%, and PCA 3 = 13.30%), and were correlated with the following eight variables: soybean yield, OM, pH, P, K, Ca, Mg, and Na (Table 4). At a 0.2-0.4 m soil depth, the three components grouped (PCA 1 = 37.87%, PCA 2 = 23.61%, and PCA 3 = 18.01%) explained 70.50% of the total variability of the data, and it correlated the following seven variables: soybean yield, OM, pH, P, K, Ca, and Na. These results are corroborated by Tripathi et al. (2015) and Jianshu (2019), who described the importance of multivariate analysis to group soil properties into principal components that consider only the properties that correlate. Therefore, PCA allows the grouping of variables even if they have presented a pure nugget effect in the preliminary geostatistical analysis, and variables related to soil natural properties or those affected by soil management (JIANSHU, 2019).

The soil chemical properties that correlated with PCA with soybean yield at the 0–0.2 m soil depth were K (PCA 1 = 16.30%), OM (PCA 2 = 15.63%), and P (PCA 3 = 43.27%). At the 0.2–0.4 m soil depth, pH (PCA 1 = 28.18%), Na (PCA 2 = 23.63%), and OM (PCA 3 = 68.64%) contributed more to explaining soybean yield variability. OM was the only property common to both the soil depths. Buttafuoco et al. (2017) stated that OM is one of the predominant variables to define management zones using geostatistical and multivariate analyses, as soil with a higher content of organic matter had normally higher CEC and V%.

Table 4. Principal components for soybean yield and soil chemical properties at the 0–0.2 and 0.2–0.4 m soil depths under a no-tillage system.

	-------------- 0–0.2 m --------------			-------------- 0.2–0.4 m --------------		
	PCA 1	PCA 2	PCA 3	PCA 1	PCA 2	PCA 3
% of variance	31.16	20.87	13.30	37.87	23.61	18.01
Cumulative %	31.16	52.04	65.34	37.87	61.49	70.50
Eigenvalue	21.50	14.40	9.18	26.51	16.52	12.61
Variables contributions (%)*						
Soybean yield	24.87	56.26	4.24	30.31	53.83	0.40
OM	13.97	15.63	0.10	0.00	9.51	68.44
pH	16.30	7.05	18.97	28.18	2.06	9.38
P	1.15	12,51	43.27	2.68	0.04	0.44
K	20.65	3.58	0.61	1.43	7.79	4.04
Ca	0.16	2.32	0.15	17.87	3.11	0.13
Mg	9.20	1.24	5.74	-	-	-
Na	13.65	1.60	26.88	19.50	23.63	17.13

OM, organic matter; pH, $CaCl_2$; P, phosphorus; K, potassium; Ca, calcium; Mg, magnesium; Na, sodium; * Variable contributions based on the correlation matrix of the main components (PCA).

Maps of spatial variability were constructed only for the more representative variables established by PCA (Figure 2). The soybean yield map was split into two zones: upper half of the area with a yield above 2,800 kg ha^{-1}, and lower half with a yield between 2,200 and 2,800 kg ha^{-1} (Figure 2a). The K (Figure 2b) and P (Figure 2d) maps did not show similarity with the soybean yield, which may be attributed to the influence of soil management on these soil chemical properties (JIANSHU, 2019). The maps of the spatial variability of the other soil chemical properties showed high similarity with the soybean yield map. This demonstrated the efficiency of multivariate analysis in grouping soil variables related to yield to define management zones.

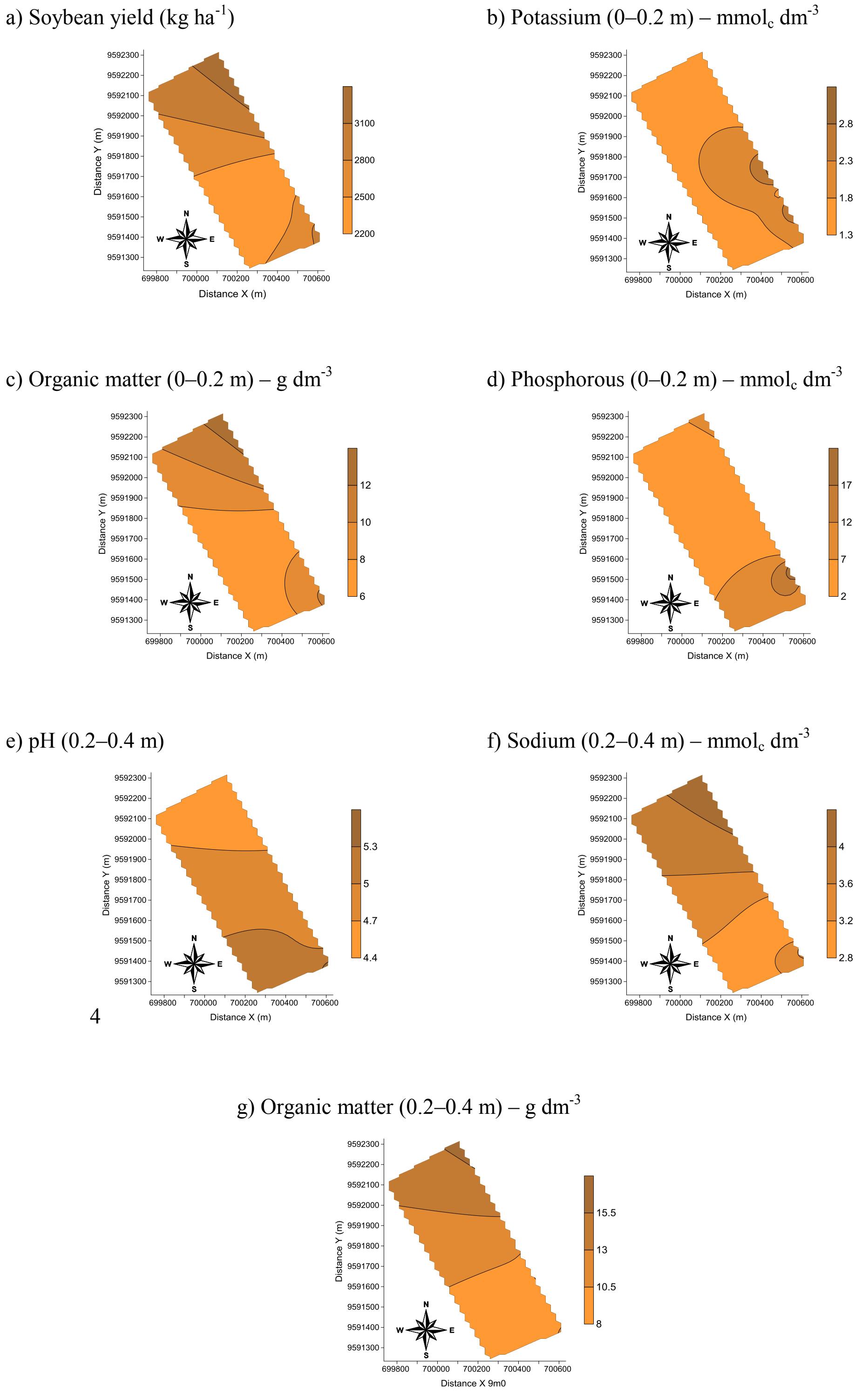

Figure 2. Maps of spatial variability of soybean yield and soil chemical properties at 0—0.2 and 0.2–0.4 m soil depth under a no-tillage system.

The geostatistical analysis of the eigenvalue scores is presented in Table 4. PCA 2 at the 0–0.2 m soil depth showed a PNE, and the OM was the chemical property that explained most of the data variability (15.63%). However, OM was spatially related to soybean yield (Figures 2a and 2c) and presented a range of 265 m in the preliminary geostatistical analysis (Table 3). The other eigenvalue scores had an experimental semivariogram adjusted by the exponential and spherical models with range from 42 m (PCA2 at the 0.2–0.4 m soil depth) up to 156 m (PCA 1 at the 0–0.2 m soil depth). Freddi et al. (2017) also fitted the exponential and spherical models for the principal components and observed the range from 34.2 to 208.1 m when studying soybean yield and soil physical and chemical properties. With regard to the SDR, the variables showed high and medium spatial dependence (Table 5).

The scaled semivariogram for the principal components at both soil depths (Figure 3) confirmed that despite the data semivariograms being fitted to different models (exponential and spherical), there is homogeneity between variance pairs up to 250 m independent of the soil depth (Figure 3). At a 0.2-0.4 m soil depth, the variables were more stable and correlated with each other than at 0-0.2 m soil depth (Figure 3b). This indicates that the maps of spatial variability were more stable and had patterns of similar contour lines.

Table 5. Semivariogram adjustment parameters of the principal components for soybean yield and soil chemical properties at 0–0.2 and 0.2–0.4 m soil depth under no-tillage system.

	Model	C_0	C_0+C_1	a	r^2	RSS	SDR
			0–0.2 m				
PCA 1	Exponential	0.00004	0.00043	156	0.403	9.07E-09	9.30
PCA 2	Pure nugget effect						
PCA 3	Spherical	0.00005	0.00153	213	0.834	3.68E-07	3.27
			0.2–0.4 m				
PCA 1	Exponential	0.00042	0.00091	165	0.863	2.88E-08	46.15
PCA 2	Exponential	0.00012	0.00094	42	0.55	9.25E-08	12.77
PCA 3	Spherical	0.00007	0.00125	82	0.586	1.22E-07	5.60

C_0: nugget effect; C_0+C_1: sill; a: range (m); r^2: coefficient of regression; RSS: residual sum of squares; SDR: spatial dependence ratio (%).

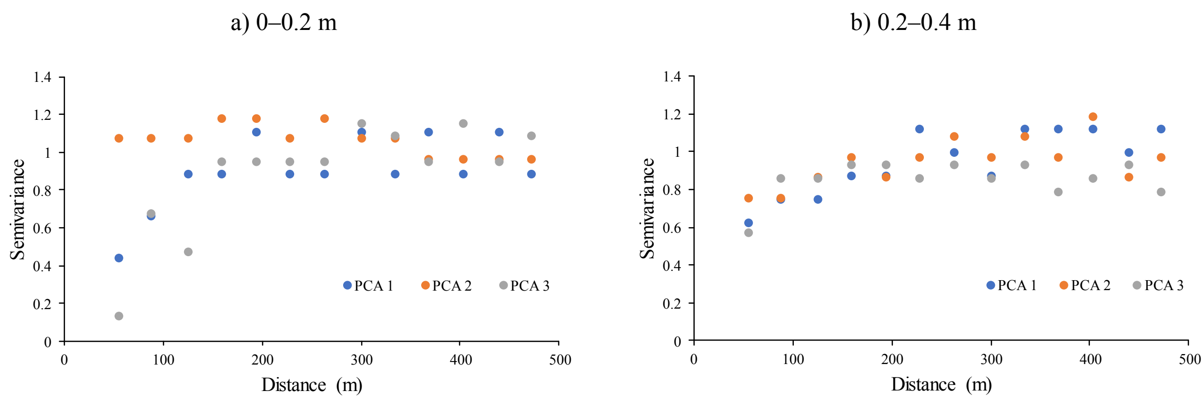

Figure 3. Scaled semivariogram of principal components for soybean yield and soil chemical properties at 0–0.2 and 0.2–0.4 m soil depth under a no-tillage system.

The use of a scaled semivariogram demonstrated that PCA is promising for defining management zones mainly because the possibility of grouping several soil and plant variables in one component allows the identification of common spatial patterns (VIEIRA et al., 1997; SIQUEIRA et al., 2015b; BUSS et al., 2019). However, when PCA results from many components are grouped into different variables, some data information can be lost (JEFFERS, 1978; SILVA et al., 2010). Our PCA considers only those components that explain more than 60% of the data variability, which indicates the efficiency of describing the spatial variability of the eigenvalue scores of the components.

The map of spatial variability of PCA 1 at the 0.2–0.4 m soil depth (Figure 4c) did not present similar distribution of contour lines with the maps of the other components. The pH was the soil property that more contributed to explaining the PCA 1 variability at the 0.2–0.4 m soil depth (Table 4). However, its spatial variability map (Figure 2e) did not have similarity with the PCA 1 map (Figure 4c). The new variable described by PCA 1 included and explained the variability that was not considered when only geostatistical analysis was performed. According to Silva et al. (2010), in principal component interpretation, individualized analysis must be performed only if the variables are independent, and priority must be given to a group of variables that explain a component. Therefore, we must analyze the components as a whole and identify a common variability pattern in all spatial variability maps.

The maps of spatial variability for PCA (Figure 4) showed two management zones: one in the upper half and the other in the lower half. The analysis of the spatial variability maps indicated that the new variables that were identified by the correlation matrix demonstrated an inverse behavior, which means that the upper area presented higher values for the components at the 0–0.2 m soil depth (Figures 4a and 4b), while the lower side area presented high values at the 0.2–0.4 m soil depth (Figure 4c, 4d, and 4e). From the analysis of soybean yield and principal components, it was observed that soybean yield was more influenced by the layers in the 0–0.2 m soil depth. Therefore, this soil layer must be considered when establishing a management zone in an area.

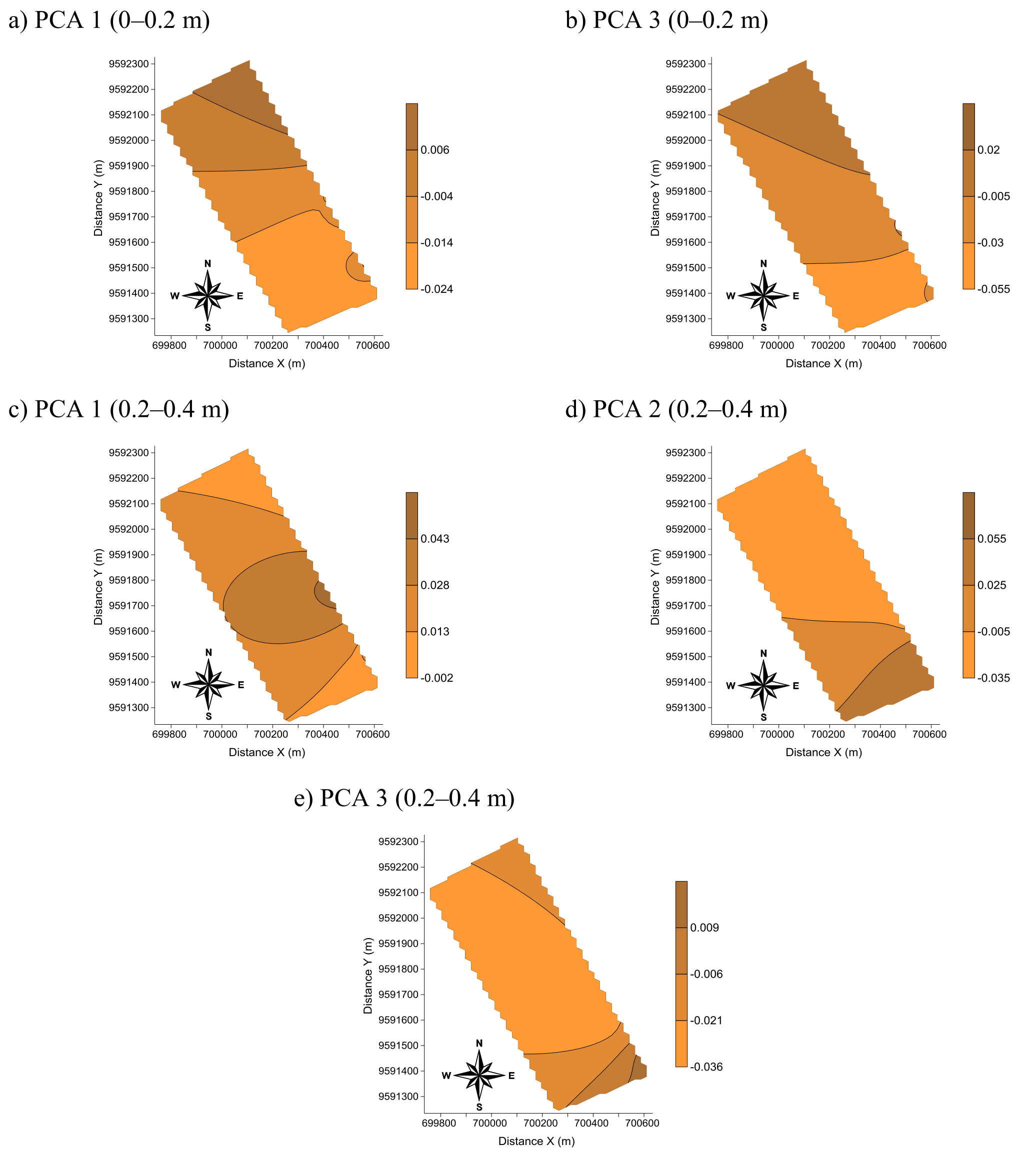

Figure 4. Spatial distribution maps of principal components at 0–0.2 m soil depth [PCA 1 (a) and PCA 3 (b)] and 0.2–0.4 m soil depth [PCA 1 (c), PCA 2 (d) and PCA 3 (e)].

Overall, the individualized maps of spatial variability (Figure 2) and variable maps grouped into components (Figure 4) showed similarity in the spatial pattern of the contour lines, particularly with the yield map (Figure 2a). This validated the separation of the areas in the two management zones, corroborating the results of Silva et al. (2010), Bitencourt et al. (2016), Lima et al. (2013), Tripathi et al. (2015), Buttafuoco et al. (2017), and Jianshu (2019).

The spatial analysis of principal components allows the construction of maps that are more homogeneous when compared to the original data, as stated by Silva et al. (2010), Lima et al. (2013), Tripathi et al. (2015), and Buss et al. (2019). Maps of spatial variability of data analyzed by multivariate techniques reduce the dimensionality of the original data, which usually have low spatial and temporal stability (GAVIOLI et al., 2016).

Our results demonstrated that PCA, together with geostatistical analysis, improved the evaluation of the spatial variability of soil chemical properties because it reduced the number of maps to be analyzed. The combination of these techniques improves decision-making related to soil fertility management and is an innovation that can be used in precision agriculture.

CONCLUSIONS

The multivariate analysis grouped the OM, pH, P, K, Ca, Mg, and Na with the soybean yield, and the three components explained 65.34% (0–0.2 m) and 70.5% (0.2–0.4 m) of data variability.

The maps of spatial variability of principal components were similar to the soybean yield map, showing the efficiency of the integration of geostatistical and multivariate techniques to define management zones.

The management zones defined by PCA 1, PCA 2, and PCA 3 confirmed that different management strategies were necessary for each soil depth in the study area.

ACKNOWLEDGEMENTS

Thanks to Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA – process UNIVERSAL-001160/17, BD-02105/17, COOP-04938/18, BESTEXT-00361/19, BINST-00362/19, UNIVERSAL-00976/19, POS-GRAD-02423/21 and RESOLUÇÃO-FAPEMA-N07-03/05/2022). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Finance Code 001). Thanks also to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Process 429354/2016-9, 307619/2016-8, 103961/2018-6 and 312515/2020-0).

REFERENCES

ALARCÓN-JIMÉNEZ, M. F. et al. Management zones based on corn yield and soil physical attributes. **Agronomia Colombiana**, 33: 373-382, 2015.

BITENCOURT, D. G. B. et al. Multivariate and geostatistical analyses to evaluate lowland soil levelling effects on physics-chemical properties. **Soil & Tillage Research**, 156: 63-73, 2016.

BUSS, R. N. et al. Spatial and multivariate analysis of soybean productivity and soil physical-chemical attributes. **Revista Brasileira de Engenharia Agricola e Ambiental**, 23: 446-453, 2019.

BUTTAFUOCO, G. et al. Geostatistical modelling of within-field soil and yield variability for management zones delineation: a case study in a durum wheat field. **Precision Agriculture**, 18: 37-58, 2017.

CAMBARDELLA, C. A. et al. Field scale variability of soil properties in central Iowa soil. **Soil Science of America Journal**, 58: 1501-1511, 1994.

CONAB – Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira: grãos, safra 2021/22 – 7º levantamento**. Brasília, DF: CONAB, 2022. 94 p.

CÓRDOBA, M. A. et al. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, 146: 95-107, 2016.

FREDDI, O. D. S. et al. Physical-chemical quality of a latossol under direct seeding and soybean-corn succession in the cerrado-amazonian ecotone. **Revista Caatinga**, 30: 991-1000, 2017.

GAVIOLI, A. et al. Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, 127: 302–310, 2016.

GOLDEN SOFTWARE. **Surfer 12: powerful contouring, gridding, and surface mapping**. Colorado: Golden Software. 2014, 1056 p.

GUEDES FILHO, O. et al. Geostatistical analysis of crop yield maps in a long term no tillage system. **Bragantia**, 69: 9-18, 2010.

JEFFERS, J. N. R. **An introduction to system**

analysis: with ecological applications. Edward Arnold Publishers: London. 1978. 198 p.

JIANSHU, L. V. Multivariate receptor models and robust geostatistics to estimate source apportionment of heavy metals in soils. **Environmental Pollution**, 244: 72-83, 2019.

LIMA, J. S. S. et al. Variabilidade espacial de atributos químicos de um latossolo vermelhor-amarelo cultivado em plantio direto. **Revista Ciência Agronômica**, 44: 16-23, 2013.

MASOUD, A. A. et al. Assessment of groundwater and soil quality degradation using multivariate and geostatistical analyses, Dakhla Oasis, Egypt. **Journal of African Earth Sciences**, 142: 64-81, 2018.

OUMENSKOU, H. et al. Multivariate statistical analysis for spatial evaluation of physicochemical properties of agricultural soils from Beni-Amir irrigated perimeter, Tadla plain, Morocco. **Geology, Ecology, and Landscapes**, 3: 1-12, 2018.

R CORE TEAM. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.r-project.org/>. Acesso em: 16 out. 2018.

RAIJ, B. V. et al. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas,SP: Instituto Agronômico. 2001. 285 p.

SANTOS, H. G. et al. **Sistema brasileiro de classificação de solos**. 5. ed. ver. e ampl. Brasília, DF: Embrapa, 2018.

SILVA, A. S. et al. Variabilidade espacial de atributos químicos de um latossolo vermelho-amarelo húmico cultivado com café. **Revista Brasileira de Ciência do Solo**, 34: 15-22, 2010.

SILVA, D. A. S.; LIMA, J. S. S. Avaliação da variabilidade do estado nutricional e produtividade de café por meio da análise de componentes principais e geoestatística. **Revista Ceres**, 59: 271-277, 2012.

SILVA, R. A.; SIQUEIRA, G. M. Multifractal analysis of soil fauna diversity indexes. **Bragantia**, 79: 120-133, 2020.

SIQUEIRA, G. M. et al. Estacionariedade do conteúdo de água de um Espodossolo Humilúvico. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 19: 439-448, 2015b.

SIQUEIRA, G. M. et al. Spatial distribution of soil apparent electrical conductivity measured by electromagnetic induction and sugarcane yield. **Bragantia**, 74: 215-223, 2015a.

SIQUEIRA, G. M. et al. O. Multifractal and joint multifractal analysis of soil invertebrate fauna, altitude, and organic carbon. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 26: 248-257, 2022.

SOBRAL, L. F. et. al. **Guia prático para interpretação de resultados de análises de solos**. Aracaju, SE: Embrapa Tabuleiros Costeiros, 2015. 13 p.

STATSOFT. **Statistica 7.0**. Tulsa: StatSoft. 2015, 250 p.

TRIPATHI, R. et al. Delineation of soil management zones for a rice cultivated area in eastern India using fuzzy clustering. **Catena**, 133: 128-136, 2015.

URIBEETXEBARRIA, A. et al. Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling. **Geoderma**, 319: 185-193, 2018.

VIEIRA, S. R. et al. Scaling of semivariograms and the kriging estimation of field-measured properties. **Revista Brasileira de Ciência do Solo**, 21: 525-533, 1997.

VIEIRA, S. R. et al. Variabilidade espacial dos teores foliares de nutrientes e da produtividade da soja em dois anos de cultivo em um latossolo vermelho. **Revista Brasileira de Ciência do Solo.** 34: 1503-1514, 2010.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.; ALVAREZ V. H.; SCHAEFER, G. R. (Eds.). **Tópicos em Ciência do Solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000. v. 2, cap. 1, p. 1-54.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of soil physical properties in the field. In: HILLEL, D. (Eds.). **Applications of Soil Physics.** New York, NY: Academic Press, 1980. v. 1, cap. 13, p. 319-344.