DELINEATION OF HOMOGENEOUS ZONES BASED ON GEOSTATISTICAL MODELS ROBUST TO OUTLIERS[1]

DANILO PEREIRA BARBOSA[2]*, EDUARDO LEONEL BOTTEGA[3], DOMINGOS SÁRVIO MAGALHÃES VALENTE[4], NERILSON TERRA SANTOS[2], WELLINGTON DONIZETE GUIMARÃES[5]

ABSTRACT – Measures of the apparent electrical conductivity (ECa) of soil are used in many studies as indicators of spatial variability in physicochemical characteristics of production fields. Based on these measures, management zones (MZs) are delineated to improve agricultural management. However, these measures include outliers. The presence or incorrect identification and exclusion of outliers affect the variogram function and result in unreliable parameter estimates. Thus, the aim of this study was to model ECa data with outliers using methods based on robust approximation theory and model-based geostatistics to delineate MZs. Robust estimators developed by Cressie–Hawkins, Genton and MAD Dowd were tested. The Cressie–Hawkins semivariance estimator was selected, followed by the semivariogram cubic fit using Akaike information criterion (AIC). The robust kriging with an external drift plug-in was applied to fitted estimates, and the fuzzy k-means classifier was applied to the resulting ECa kriging map. Models with multiple MZs were evaluated using fuzzy k-means, and a map with two MZs was selected based on the fuzzy performance index (FPI), modified partition entropy (MPE) and Fukuyama–Sugeno and Xie–Beni indices. The defined MZs were validated based on differences between the ECa means using mixed linear models. The independent errors model was chosen for validation based on its AIC value. Thus, the results demonstrate that it is possible to delineate an MZ map without outlier exclusion, evidencing the efficacy of this methodology.

Keywords: Robust statistics. Precision agriculture. Apparent soil electrical conductivity. Spatial variability. fuzzy k-means.

DELINEAMENTO DE ZONAS HOMOGÊNEAS POR GEOESTATÍSTICA BASEADA EM MODELOS ROBUSTA À OUTLIERS

RESUMO – Diversas pesquisas utilizam medidas de condutividade elétrica aparente do solo (CEa) como indicador da variabilidade espacial de atributos físico-químicos existentes no campo de produção. Com base nestas medidas, zonas de manejo (ZM) são delineadas para aperfeiçoamento da gestão agrícola. Entretanto, estas amostras têm apresentado presença de outliers. Todavia, a presença ou incorreta detecção e exclusão de outliers altera o formato do variograma, exibindo estimativas não fidedignas para os seus parâmetros. Dessa forma, objetivou-se nesta pesquisa, tratar dados amostrais da CEa por meio de métodos robustos à presença de outliers, fundamentados na teoria de aproximações robusta e na geoestatística baseada em modelos, para o delineamento de ZM. Assim, estimadores robustos de Cressie Hawkins, Genton's e MAD Dowd foram avaliados. Nesta avaliação, selecionou-se o estimador de semivariância de Cressie Hawkins. E na sequência, optou-se pelo ajuste cúbico do semivariograma via Critério de Informação de Akaike (AIC). As estimativas obtidas com este ajuste foram aplicadas na plug-in robusto de krigagem. E coerentemente o mapa de krigagem da CEa obtido foi utilizado no classificador *fuzzy k-means*. Com uso do *fuzzy k-means,* diferentes ZM foram avaliadas, selecionando-se o mapa com duas ZM por meio dos índices FPI, MPE, Fukuyama-Sugeno e xie beni. As ZM estabelecidas foram validadas quanto as suas diferenças médias relativas à CEa por meio de modelos lineares mistos. Nesta validação optou-se pelo modelo de erros independentes, através do AIC. E dessa forma, diante a exposição dos resultados alcançados, foi possível delinear o mapa de ZM sem necessidade de recorrer à exclusão de outliers, evidenciando o mérito da metodologia empregada.

Palavras-chave: Estatística robusta. Agricultura de precisão. Condutividade elétrica aparente. Variabilidade espacial. *Fuzzy k-means*.

INTRODUCTION

With the development of precision agriculture, mapping of soil heterogeneity has become a relevant tool in agricultural management. Thus, precision agriculture may be defined as a systematic procedure to inspect and incorporate soil spatial variability in field management (HAGHVERDI et al., 2015). Such spatial variability may be caused by climatic, topographical and biological factors (CÓRDOBA et al., 2013).

Management of field spatial variability is performed through the delineation of management zones (MZs). MZs are field sub-regions that have similar needs based on soil physicochemical features. The delineation of these sub-regions allows specific input needs to be identified and thus decreases their usage, increasing agricultural sustainability (CÓRDOBA et al., 2013; BOTTEGA et al., 2017).

However, field spatial variability must be determined to delineate MZs. Such knowledge is obtained by modelling the spatial characterization of factors that are relevant to field management. The apparent soil electrical conductivity (ECa) is among the relevant factors that must be considered (CÓRDOBA et al., 2016; PERALTA et al., 2015; HAGHVERDI et al., 2015; CÓRDOBA et al., 2013; SCUDIERO et al., 2013; PERALTA et al., 2013). The widespread use of such a feature is due to its ease of sampling, low cost and strong influence on plant growth and yield (PERALTA et al., 2013).

Depending upon climatic conditions and other factors, ECa distribution can either be symmetrical (CÓRDOBA et al., 2013) or asymmetrical (SCUDIERO et al., 2013; YAO et al., 2014; BOTTEGA, 2014; SHAHBAZI et al., 2013). The latter property has a significant effect on the fit of the variogram theoretical model (FU et al., 2016), on statistical kriging cross-validation (KERRY; OLIVER, 2007) and on the precision of the spatial variability map. Therefore, asymmetry compromises the delineation of MZs, negatively influencing field management.

Therefore, asymmetrical distributions of sampling data must be carefully analysed due to the irrefutable occurrence of outliers. Outliers are atypical values within a dataset (FU et al., 2016; CÓRDOBA et al., 2016). Several methods such as Q-Q plot, box plot, local indicators of spatial association (LISA) and cross-validation can be used to identify outliers (FU et al., 2016).

Once identified, outliers are excluded from the analysis (RAMOS et al., 2017; CÓRDOBA et al., 2016; FU et al., 2016; PICCINI; MARCHETTI; FRANCAVIGLIA, 2014). Outlier exclusion favours the normal distribution of the data without data transformation (FU et al., 2016) and allows the correct interpolation of the spatial variability map, and thus the correct management decision may be made based on the analysis (TAYLOR; MCBRATNEY; WHELAN, 2007).

However, the incorrect outlier exclusion can undermine statistical determinations (MARONNA; MARTIN; YORAI, 2006) and, consequently, compromise the estimates of variogram parameters, and, ultimately, field management decisions. In this context, the present study aims to delineate MZs without excluding outliers, using methods that are insensitive to outliers. Robust geostatistical methods will be used both to define variogram parameters and to interpolate the spatial variability map.

MATERIAL AND METHODS

ECa was used in this study due to the advantages of its use in delineating MZs, as described by Peralta et al. (2013). Georeferenced ECa data from Bottega (2014), collected from a rural property in the municipality of Ponta Porã in the state of Mato Grosso do Sul, Brazil (22°32'09"S, 55°43'33"W), were used. Samples were collected in February 2012 from 160 locations at 50-m regular intervals, using a receptor GPS (model GPSMAP 62, brand Garmin).

Georeferenced sample geographic coordinates were transformed from the World Geodetic System 84 (WGS84) to Universal Transverse Mercator (UTM) zone 21 South using the "rgdal" package in R (BIVAND et al., 2016), followed by descriptive analysis of samples focusing on data asymmetry and outlier detection, as outliers affect the variogram function (OLIVER; WEBSTER, 2014).

Values outside of the range defined as ± 3 times the standard deviation (sd) around the mean ($\bar{x}$) were defined as outliers, as described by Córdoba et al. (2016).

Once outliers were identified, data were analysed by robust geostatistics using the "georob" package (PAPRITZ, 2017) in R (TEAM, 2016). This methodology is unique in that it is based on the theoretical fundaments of the robust approach (KÜNSCH et al., 2013;) and on geostatistics based on models (DIGGLE; RIBEIRO, 2007).

Thus, bservations of ECa $y^T = (y(s_1), y(s_2), ..., y(s_n))$ sampled at positions s_i were used to obtain experimental variograms for the following semivariance estimators: methods-of-moments, proposed by Matheron (ISAAKS; SRIVASTAVA, 1989); the robust estimator, proposed by (CRESSIE; HAWKINS, 1980), which is insensitive to outliers, as opposed to Matheron's classic estimator; the robust estimator, proposed by Genton (GENTON, 1998); and the median absolute deviation (MAD) robust estimator (DOWD, 1984). These semivariance estimators were tested to verify which most satisfactorily modelled the spatial dependency pattern for the stationary process with

the least sensitivity to outliers.

The selected semivariance estimator provided the initial estimates of the semivariogram parameters. These initial estimates were used in the model proposed by Künsch et al. (2013) – $Y(s) = x(s)^T\beta + Z(s) + \varepsilon(s)$ – to predict $Z(s)$ at non-observed positions s_i. In this model, $x(s)^T\beta$ is defined as the trend, $Z(s)$ is the stationary Gaussian process of zero-mean and covariance matrix, R, and $\varepsilon(s)$ is the independent error with the squared scale parameter σ^2 (nugget effect) (KÜNSCH et al., 2013). R is defined as the covariance matrix as a function of distance h and $\theta^T = (\sigma_0^2, \alpha)$, where σ_0^2 is the plateau and σ^2 is the range (KÜNSCH et al., 2013).

The unique characteristic of this method is the use of robust restricted maximum likelihood (RREML), proposed by KÜNSCH et al. (2013), which maximizes the restricted Gaussian log likelihood, in addition to the parameters described in Diggle and Ribeiro (2007) $(\sigma^2, \sigma_0^2, \alpha)$ and the latent variable z.

Using the experimental semivariogram's initial entry values in the model, the semivariogram models that exhibited convergence using this robust methodology were evaluated. Among the evaluated models, the one with lowest Akaike information criterium (AIC) value was selected, providing the parameter estimates used in the kriging plug-in (KÜNSCH et al., 2013):

$$\hat{Z}(s_0) = x(s_0)^T\hat{\beta} + \gamma_{\hat{\theta}}^T(s_0)\Gamma_{\hat{\theta}}^{-1}\hat{Z}$$

where $\Gamma_{\hat{\theta}}$ is the estimated covariance matrix of Z and $\gamma_{\hat{\theta}}$ is the estimated covariance vector between Z and $Z(s_0)$. After kriging was performed, the following descriptive statistics for kriging validation were obtained: mean prediction error (MPE) and mean square deviation ratio (MSDR) (OLIVER; WEBSTER, 2014).

The resulting modelled map of the ECa spatial structure was processed with the fuzzy k-means algorithm using the "e1071" package in R (DIMITRIADOU et al., 2008). This algorithm ranks the interpolated values into classes to minimize the sum of the squared distances within the domain of a defined cluster (SCUDIERO et al., 2013). This methodology defines a diffused element in which a given observed point may belong to more than one class, depending upon a pre-defined weighting exponent (SCUDIERO et al., 2013).

Then, the weighting exponent was pre-defined at 1.35, as described by Odeh; Chittleborough; McBratney (1992), and was used in the cluster definition, with the resulting maps featuring 2 to 5 clusters. The following indices were used to define the optimal number of clusters: fuzziness performance index (FPI), modified partition entropy (MPE), Fukuyama–Sugeno (FS) and Xie-–Beni (Xb) (RAMOS et al., 2017; CÓRDOBA et al., 2016; YAO et al., 2014). The lowest index values were used to define the number of classes. The use of lowest values resulted in better separation among clusters and a higher similarity of values within each cluster (RAMOS et al., 2017; SONG et al., 2009).

Once the number of clusters was identified, a spatial filter to smooth the continuity of zones and reduce the fragmentation between classes was applied to the respective map (LARK, 1998). Following the protocol of Córdoba et al. (2016), a median spatial filter with 9 x 9 pixels was applied to improve smoothing, without fragmentation. Thus, one pixel is defined as the median of its neighbours (GONZALEZ; WOODS, 2008).

After applying the filter for the smoothing of the classes, MZs were validated by verifying any significant differences between MZ means. For that purpose, random samples from each MZ were selected and mixed linear models (MLMs) were defined considering the MZ as a fixed effect and random errors as spatially correlated.

This configuration, also used by Córdoba et al. (2016), was defined using the following MLMs: spherical and exponential models, with and without the nugget effect, and the independent errors model. These models were also applied in the validation of MZs in the present study.

Each defined model was evaluated based on the AIC, with the lowest value used to select the model. Next, the means of the MZs and their respective confidence intervals were reported. In the MZ validation process, the packages "nlme" (PINHEIRO et al., 2016) and "lsmeans" (RUSSELL, 2016) in R were used do define the MLMs and mean tests, respectively.

RESULTS AND DISCUSSION

Following the transformation of coordinates from WGS84 to UTM zone 21 South, descriptive statistics for the ECa dataset were obtained (Table 1. Descriptive statistics of apparent electrical conductivity (ECa) sampled from 0- to 20-cm depth in a rural property in the municipality of Ponta Porã, Mato Grosso do Sul.).

Table 1. Descriptive statistics of apparent electrical conductivity (ECa) sampled from 0- to 20-cm depth in a rural property in the municipality of Ponta Porã, Mato Grosso do Sul, Brazil.

Feature	N[1]	Minimum	Mean	Median	Maximum	Sd[2]	CV (%)[3]	Cs[4]	Ck[5]
ECa[6]	160	2.74	6.19	5.94	19.31	2.13	34.44	2.24	9.29

[1]Sample size; [2]Standard deviation; [3]Coefficient of variation; [4]Asymmetry coefficient; [5]Coefficient of kurtosis; [6]Soil apparent electrical conductivity in 2012 (mS m^{-1}).

The ECa dataset was comprised of 160 samples and had an amplitude of 16.57 mS m^{-1}, a 34% variation around a mean of 6.19 mS m^{-1} and displayed positive asymmetry (Table 1. Descriptive statistics of apparent electrical conductivity (ECa) sampled from 0- to 20-cm depth in a rural property in the municipality of Ponta Porã, Mato Grosso do Sul.).

Similar CV values for ECa, ranging from 17.61% to 44.49%, and a higher mean ECa, ranging from 12.79 to 27.42, were obtained by Peralta et al. (2013). These differences are due to different sample depths (0 to 90 cm), a higher sand content and lower clay content, which lead to higher water accumulation and, thus, lower ECa values (PERALTA et al., 2013).

Scudiero et al. (2013) reported a similar ECa mean (1.14 dS m^{-1}) and also detected positive asymmetry (1.05). ECa distribution asymmetry also has been reported in several other studies (YAO et al., 2014; BOTTEGA et al., 2017; SHAHBAZI et al., 2013; AGGELOPOOULOU et al.; 2013).

Positive asymmetry (Table 1. Descriptive statistics of apparent electrical conductivity (ECa) sampled from 0- to 20-cm depth in a rural property in the municipality of Ponta Porã, Mato Grosso do Sul.) had atypical values in the overall dataset (Figure 1).

Such atypical values were evaluated for potential outliers, and three values were identified outside of the range $[\bar{x} \pm sd: 6.19 \pm 2.13]$ using the boxplot (Figure 1). The outlier positions were determined (Figure 2).
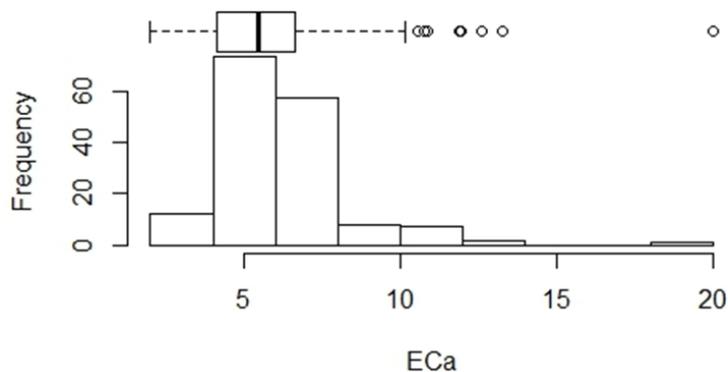
Figure 1. Histogram and boxplot of apparent electrical conductivity (ECa) distribution.
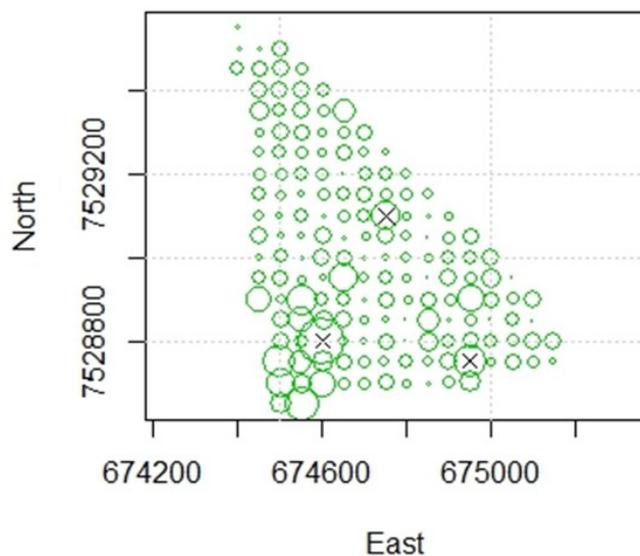
Figure 2. Sample mesh where X denotes the positions of the outliers.

Due to the depth of samples analysed for ECa and the specific physicochemical properties of the soil, Scudiero et al. (2013) and Córdoba et al. (2016) reported different thresholds for the detection of outliers, $[\bar{x} \pm sd: 1.14 \pm 0.72]$ and $[\bar{x} \pm sd: 3.8 \pm 7]$, respectively.

Once outliers were identified, their influence on the variogram function was evaluated. Thus, the semivariance estimators developed by Matheron, Genton, Cressie and Hawkins, and MAD Dowd were compared regarding the expected typical behaviour of an experimental semivariogram, considering ECa data. These semivariance estimators, except that of Matheron, must be used in situations related to the pertinence of the outliers to the dataset (OLIVER; WEBSTER, 2014).
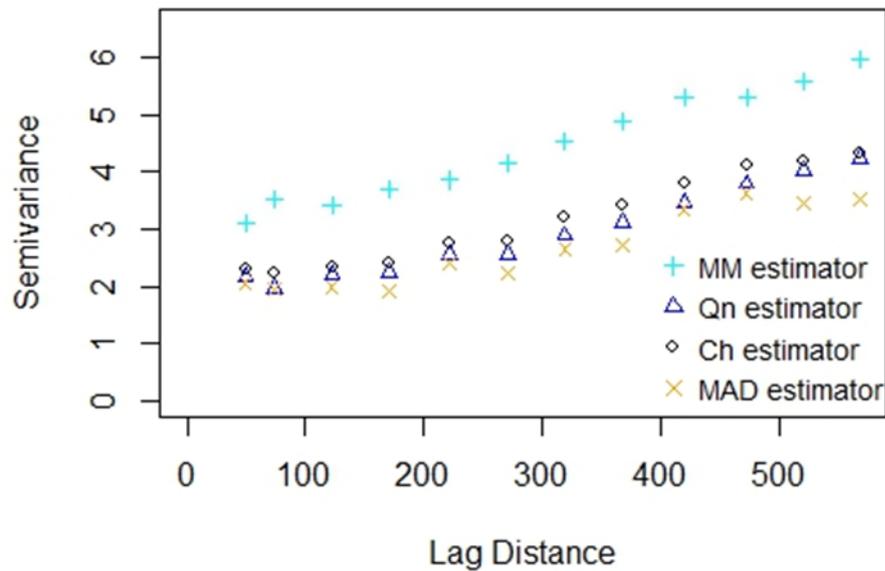
Figure 3. Experimental semivariogram for the semivariance estimators of Matheron (MM), Genton (Qn), Cressie–Hawkins (Ch) and Dowd (MAD) applied to the ECa sampling data.

It is noteworthy that the outliers affect the semivariance function for the Matheron estimator, which does not display a constant plateau (Figure 3). However, the other semivariance estimators (Genton, Cressie and Hawkins, and Dowd), which are robust to outliers, exhibit similar behaviour regarding both trend and plateau stability (Figure 3). The Cressie–Hawkins estimator was selected for the following analysis, as it presented the smallest kriging variance, mean prediction error and mean squared error (Table 2).

Based on the Cressie and Hawkins semivariogram, the cubic model was fitted as it displayed the lowest AIC value (577) compared to the other models: Gaussian (620), geniting (690) and penta (705). The cubic model fitted to the semivariances (Figure 4) displayed a nugget effect of 1.96, a plateau at 3.9 and a range of 630.

Table 2. Statistical cross-validation for the semivariance estimators: Matheron (MM), Genton (Qn), Cressie–Hawkins (Ch) and Dowd (MAD).

Estimators	Var[1]	ME[2]	MSDR[3]
MM	0.0	0.0	0.91
Qn	0.1	0.0	0.87
Ch	0.0	0.0	1.01
MAD	0.1	0.0	0.88

[1]Kriging variance; [2]Mean prediction error; [3]Mean squared deviation ratio.
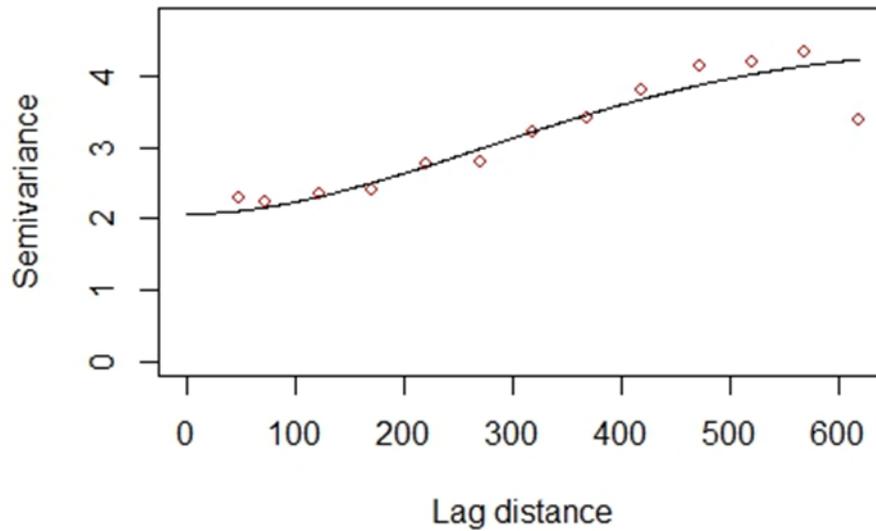
Figure 4. Cubic model fitted to the semivariances of the apparent electrical conductivity (ECa) data.

Values obtained for the variogram parameter estimates (Figure 4) were used as entry data in the kriging plug-in using the package "*georob*" (PAPRITZ, 2017). With this function, the ECa kriging was obtained (Figure 5) as well as the descriptive statistics for kriging cross-validation (ME = 0.1 and MSDR = 1.66).
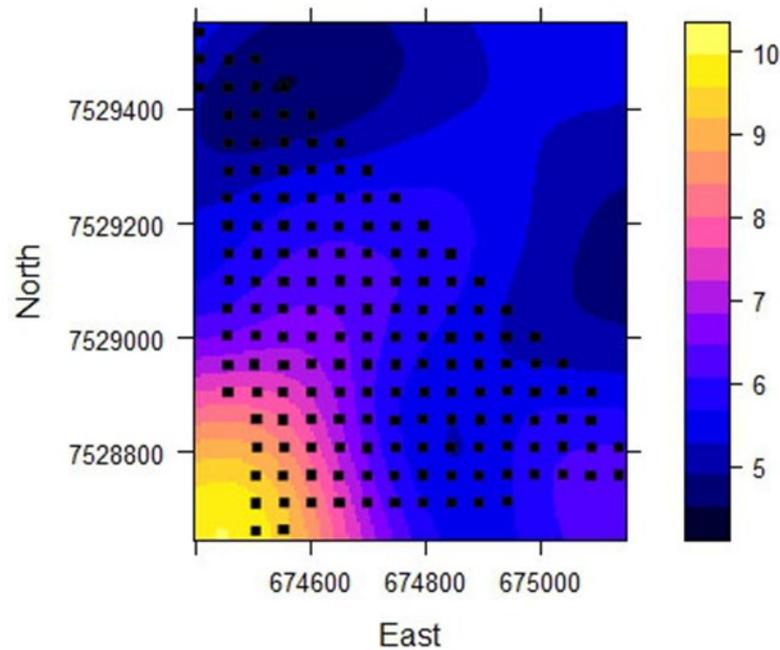
Figure 5. Kriging plug-in data interpolation (PAPRITZ, 2017) defining the spatial variability structure of the apparent electrical conductivity (ECa).

The summarized values for kriging cross-validation support the analysis, given that the mean error approaches zero and the mean square error ratio is approximately one, which indicates unbiased kriging (OLIVER; WEBSTER, 2014).

The interpolation displaying the spatial variability structure of ECa (Figure 5) was used as complementary information in the fuzzy k-means algorithm. Such an algorithm was used to group the interpolated values into zones (RODRIGUES JUNIOR et al., 2011).

The number of zones to be evaluated varies among studies; in general, two to six zones are used (RAMOS et al., 2017; CÓRDOBA et al., 2016; HAGHVERDI et al., 2015; YAO et al., 2014). Thus, two to five zones were evaluated using the "*cmeans*" function and weighting exponent of 1.35 (ODEH; CHITTLEBOROUGH; MCBRATNEY, 1992) (Figure 6).
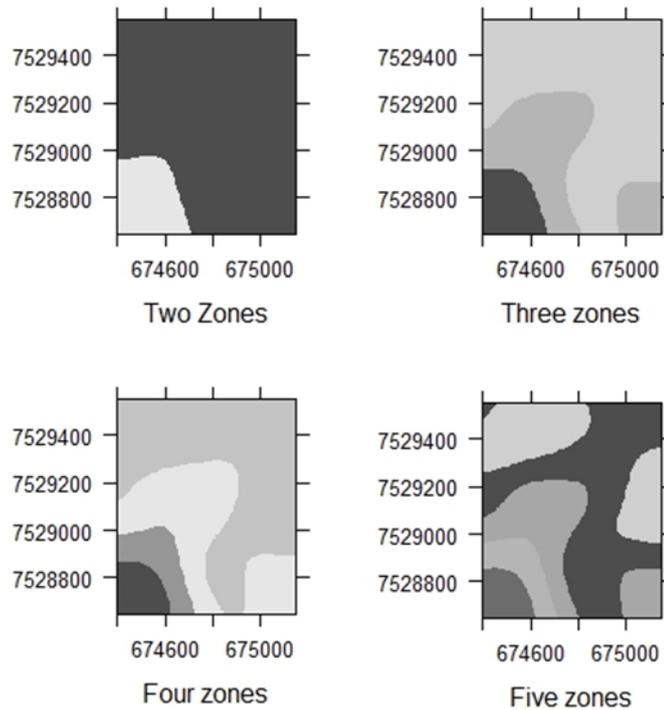
Figure 6. Spatial distribution with two, three, four and five management zones (MZs) within the study area, determined based on the fuzzy k-means algorithm.

The optimal number of MZs is achieved when the FPI, MPE, FS and Xb indices (RAMOS et al., 2017; CÓRDOBA et al., 2016; YAO et al., 2014) are lowest (RAMOS et al., 2017; SONG et al., 2009). In this study, based on these indices, the optimal number of MZs was two.

Similarly, Córdoba et al. (2013), Ramos et al. (2017) and Yao et al. (2014) reported two MZs was optimal. Once the optimal number of MZs is defined, a validation is performed to verify the significance of the difference between the MZ mean values.

MLMs were fitted to 1,000 samples randomly selected from each MZ and evaluated regarding their AIC values. In these models, MZs were considered as a fixed effect and errors were spatially correlated with exponential and spherical spatial structures, with and without nugget effect, and an independent errors model.

The independent errors model was selected as it had the lowest AIC value (3,072) and it detected significant differences between the means from the two defined MZs (Figure 7).
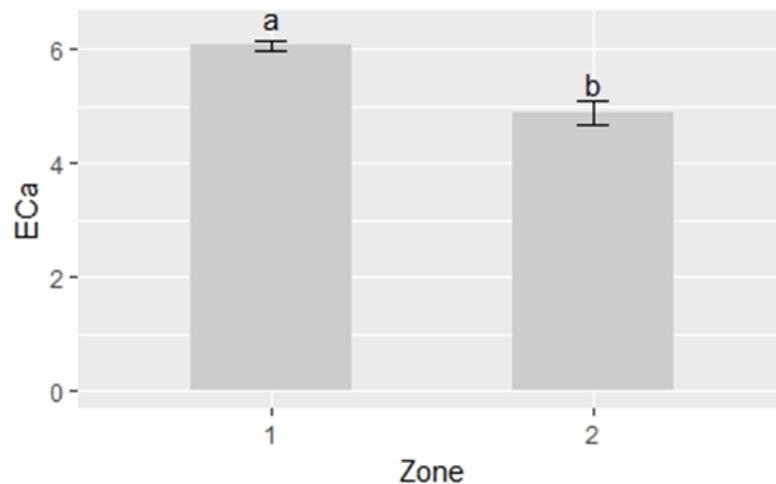
Figure 7. Mean apparent electrical conductivity (ECa) in the defined management zones (MZs) and confidence intervals. Letters 'a' and 'b' indicate statistically significant differences (P<0.001) between MZ means.

With the selection of the independent errors model, it is possible to also verify the non-overlap of the confidence interval for the means obtained for each of the defined MZs (Figure 7), which highlights the heterogeneity among the MZs. Finally, after validating the MZs and identifying their optimal number, it was possible to confirm the use of two MZs (Figure 8).
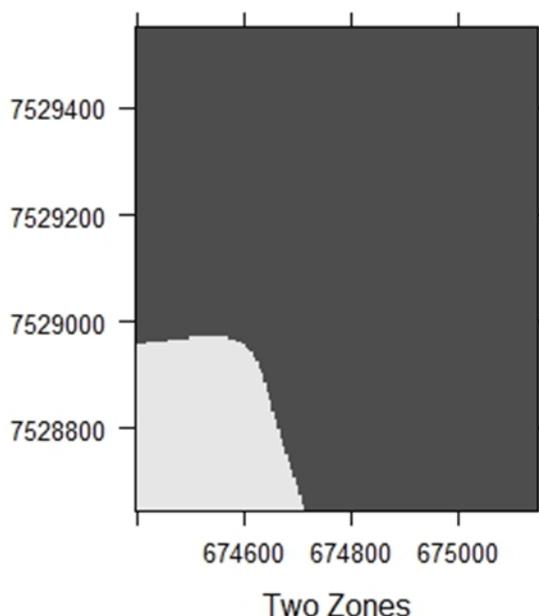
Figure 8. Spatial characterization of the apparent electrical conductivity (ECa) of the management zones (MZs) determined based on robust geostatistics.

The spatial distribution displayed in Figure 8 exhibits a degree of continuity in the limits of the obtained MZ, which was due to the application of a median smoothing filter (CÓRDOBA et al., 2016), comprised of a 9 x 9-pixel window. In this filter, each pixel is defined as the median of its internal neighbours within each window.

CONCLUSION

The use of robust estimators of semivariance was suitable to the dataset studied, as it defined the semivariogram function in spite of asymmetry and outliers. Thus, the application of a robust Gaussian restricted log-likelihood methodology included in the kriging plug-in resulted in MZs that can easily be managed, justifying its application. The procedure enabled the understanding of the spatial variability of soil physicochemical features that correlate with ECa. Thus, a probable soil correction is easily identified, which consequently reduces agricultural management costs.

ACKNOWLEDGMENTS

REFERENCES

AGGELOPOOULOU, K. et al. Delineation of management zones in an apple orchard in Greece using a multivariate approach. **Computers and Electronics in Agriculture**, v. 90, s/n., p. 119-130, 2013.

BIVAND, R. et al. **Bindings for the geospatial Data Abstraction Library**. R package version 1. 1–10. 2016.

BOTTEGA, E. L. **Utilization of management zones in soy production in the Brazilian savanna**. 2014. 78 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2014.

BOTTEGA, E. L. et al. Precision agriculture applied to soybean: Part I - Delineation of management zones. **Australian Journal of Crop Science**, v. 11, n. 5, p. 573-579, 2017.

CÓRDOBA, M. A. et al. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, v. 143, s/n., p. 95-107, 2016.

CÓRDOBA, M. et al. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. **Computers and Electronics in Agriculture**, v. 97, s/n., p. 6-14,

2013.

CRESSIE, N.; HAWKINS, D. M. Robust estimation of the variogram: I. **Journal of the International Association for Mathematical Geology**, v. 12, n. 2, p. 115-125, 1980.

DIGGLE, P. J.; RIBEIRO, P. J. **GEOSTATISTICS, Model-based**. Springer series in statistics. Springer, New York, 2007. 228 p.

DIMITRIADOU, E. et al. **Misc functions of the Department of Statistics (e1071), TU Wien**. R package version 1. 2008. 5-24 p.

DOWD, P. A. **The variogram and kriging: robust and resistant estimators. Geostatistics for natural resources characterization**. Springer: Dordrecht, 1984. p. 91-106.

FU, W. et al. Outlier identification of soil phosphorus and its implication for spatial structure modeling. **Precision Agriculture**, v. 17, n. 2, p. 121-135, 2016.

GENTON, M. G. Highly robust variogram estimation. **Mathematical Geology**, v. 30, n. 2, p. 213-221, 1998.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Upper Saddle River, New Jersey 2008. 977 p.

HAGHVERDI, A. et al. Perspectives on delineating management zones for variable rate irrigation. **Computers and Electronics in Agriculture**, v. 117, s/n., p. 154-167, 2015.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. Oxford University Press, New York, 1989. 561 p.

KERRY, R.; OLIVER, M. A. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. **Geoderma**, v. 140, n. 4, p. 383-396, 2007.

KÜNSCH, H. R. et al. **Robust estimation of the external drift and the variogram of spatial data**. In: ISI 58th World Statistics Congress of the International Statistical Institute. Eidgenössische Technische Hochschule Zürich, 2013.

LARK, R. M. Forming spatially coherent regions by classification of multi-variate data: an example from the analysis of maps of crop yield. **International Journal of Geographical Information Science**, v. 12, n. 1, p. 83-98, 1998.

MARONNA, R. A. R. D.; MARTIN, D.; YOHAY, V. Robust statistics. John Wiley & Sons, Chichester, 2006. 57 p.

OLIVER, M. A.; WEBSTER, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. **Catena**, v. 113, s/n., p. 56-69, 2014.

ODEH, I. O. A.; CHITTLEBOROUGH, D. J.; MCBRATNEY, A. B. Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. **Soil Science Society of America Journal**, v. 56, n. 2, p. 505-516, 1992.

PAPRITZ, A. **georob: Robust Geostatistical Analysis of Spatial Data. R package version 0.3-4, 2017**. Disponível em: <67 https://CRAN.R-project.org/package=georob> Acesso em 20 set. 2017.

PERALTA, N. R. et al. Delineation of management zones to improve nitrogen management of wheat. **Computers and Electronics in Agriculture**, v. 110, s/n., p. 103-113, 2015.

PERALTA, N. R. et al. Delineation of management zones with measurements of soil apparent electrical conductivity in the southeastern pampas. **Canadian Journal of Soil Science**, v. 93, n. 2, p. 205-218, 2013.

PICCINI, C.; MARCHETTI, A.; FRANCAVIGLIA, R. Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment. **Ecological Indicators**, v. 36, s/n., p. 301-314, 2014.

PINHEIRO, J. et al. **nlme: Linear and nonlinear mixed effects models. R package version, v. 3.1-128, 2016**. Disponível em: <http://CRAN.R-project.org/package=nlme>. Acesso em: 10 set. 2017.

RAMOS, F. T. et al. Defining management zones based on soil attributes and soybean productivity. **Revista Caatinga**, v. 30, n. 2, p. 427-436, 2017.

RODRIGUES JUNIOR, F. A. et al. Generation of coffee crops management zones using the SPAD sensor and leaf analysis. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 8, p. 778-787, 2011.

RUSSELL V. L. Least-Squares Means: The R Package lsmeans. **Journal of Statistical Software**, v. 69. n. 1, p. 1-33, 2016.

SCUDIERO, E. et al. Delineation of site-specific management units in a saline region at the Venice Lagoon margin, Italy, using soil reflectance and

apparent electrical conductivity. **Computers and Electronics in Agriculture**, v. 99, s/n., p. 54-64, 2013.

SHAHBAZI, F. et al. Geostatistical analysis for predicting soil biological maps under different scenarios of land use. **European Journal of Soil Biology**, v. 55, s/n., p. 20-27, 2013.

SONG, X. et al. The delineation of agricultural management zones with high resolution remotely sensed data. **Precision Agriculture**, v. 10, n. 6, p. 471-487, 2009.

TEAM, R. Core. **R: A language and environment for statistical computing [Computer software]**. R Foundation for Statistical Computing, Vienna, Austria. 2016.

TAYLOR, J. A.; MCBRATNEY, A. B.; WHELAN, B. M. Establishing management classes for broadacre agricultural production. **Agronomy Journal**, v. 99, n. 5, p. 1366-1376, 2007.

YAO, R. J. et al. Determination of site-specific management zones using soil physico-chemical properties and crop yields in coastal reclaimed farmland. **Geoderma**, v. 232, s/n., p. 381-393, 2014.