ESTABLISHING MANAGEMENT ZONES USING SOIL ELECTRICAL CONDUCTIVITY AND OTHER SOIL PROPERTIES BY THE FUZZY CLUSTERING TECHNIQUE

José Paulo Molin[1]*; Cesar Nunes de Castro[2]

[1]USP/ESALQ - Depto. de Engenharia Rural, C.P. 09 - 13418-900 - Piracicaba, SP - Brasil.
[2]USP/ESALQ - Programa de Pós-Graduação em Máquinas Agrícolas.
*Corresponding author <jpmolin@esalq.usp.br>

ABSTRACT: The design of site-specific management zones that can successfully define uniform regions of soil fertility attributes that are of importance to crop growth is one of the most challenging steps in precision agriculture. One important method of so proceeding is based solely on crop yield stability using information from yield maps; however, it is possible to accomplish this using soil information. In this study the soil was sampled for electrical conductivity and eleven other soil properties, aiming to define uniform site-specific management zones in relation to these variables. Principal component analysis was used to group variables and fuzzy logic classification was used for clustering the transformed variables. The importance of electrical conductivity in this process was evaluated based on its correlation with soil fertility and physical attributes. The results confirmed the utility of electrical conductivity in the definition of management zones and the feasibility of the proposed method.
Key words: precision agriculture, soil physical properties, fuzziness

DELINEAMENTO DE UNIDADES DE GERENCIAMENTO DIFERENCIADO COM O USO DE CONDUTIVIDADE ELÉTRICA E ATRIBUTOS DO SOLO POR MEIO DE TÉCNICAS DE LÓGICA FUZZY

RESUMO: O delineamento de unidades de gerenciamento diferenciado, que pode determinar com sucesso regiões uniformes em termos de atributos da fertilidade dos solos que são de importância para o desenvolvimento das culturas, é uma das mais desafiadoras etapas da agricultura de precisão. Um método destacado consiste na utilização da informação associada à estabilidade da produtividade das culturas, contida nos mapas de produtividade; no entanto é possível se executar esse delineamento a partir de informações do solo. Nesse estudo o solo foi amostrado para condutividade elétrica e onze outras propriedades, buscando definir unidades de gerenciamento com uniformidade nessas propriedades. Análise de componentes principais foi utilizada para agrupar variáveis e lógica fuzzy foi utilizada para classificar as variáveis transformadas. A importância da condutividade elétrica nesse processo foi avaliada assim como as suas correlações com as propriedades físicas e químicas do solo. Confirmou-se a utilidade da condutividade elétrica na delimitação de unidades de gerenciamento diferenciado e a exeqüibilidade do método de delimitação proposto.
Palavras-chave: agricultura de precisão, propriedades físicas do solo, descontinuidade

INTRODUCTION

Precision agriculture presents promising perspectives in the development of new technologies and crop management propositions, optimizing inputs and allowing production cost reductions or increases in yield in addition to possible environmental benefits. Different strategies can be used in order to maximize the effectiveness of agricultural inputs applied on variable rates. One approach is based on management zones that represent a homogeneous combination of potential productivity-limiting factors, which are therefore permanent (Fridgen et al., 2001) and refer to geographic regions that present topography and soil attributes with minimal heterogeneity (Luchiari Jr. et al., 2000). The determination of homogeneous areas within a field is difficult to achieve due to the complex combination among factors which may influence yield.

Soil electrical conductivity (EC) has attracted attention as a mapping tool since it is a quick and economic method of indicating soil productivity (McBride et al., 1990). In turn, electrical conductivity depends on soil water content, chemical composition of the soil solution and soil exchangeable ions, soil clay content,

and the interaction between non-exchangeable and exchangeable ions (Nadler & Frenkel, 1980). EC has been used to monitor the spatial variability of several soil properties, such as soil water content (Sheets & Hendrickx, 1995), CEC and exchangeable Ca and Mg (McBride et al., 1990), and soil clay content (Machado et al., 2006). Minasny & McBratney (2000) presented a fuzzy clustering technique, known as fuzzy k-means, used by Fridgen et al. (2001) to identify natural clusters that occur among data.

The objectives of this study were to monitor soil EC in a field under no-till, compare it in relation to the spatial variability of soil physicochemical characteristics, analyze the correlation among variables, run a multivariate analysis (principal components analysis) among all variables and delineate soil management zones using EC and other physicochemical variables via the fuzzy k-means clustering technique.

MATERIAL AND METHODS

The study was conducted in a 35.8 ha area, located in Ponta Grossa, state of Paraná, Brazil (50º12' W, 25º9' S), cultivated under no-tillage system since the 1980s during the summer cropping seasons on a soybean-corn rotation system and during the winter season with wheat (*Triticum* spp.) or black oat (*Avena strigosa Schreb*) as a cover crop. The soil type in the area was classified on two major orders: Oxisol and Inceptsol. Soil EC was measured in October 2002 using a Veris 3100® Soil EC Mapping System (Veris Technologies, Inc., Salina, KS, USA). It collected EC information simultaneously at two depths, providing both shallow (0–0.3 m) and deep readings (0–0.9 m). The equipment was run at 6 m wide passes and it uses six electrodes as sensors, connected to smooth disk coulters which penetrate the soil. As the assembly moves across the field, a pair of these electrodes transmits an electric current to the soil, while the other two pairs measure the potential difference verified in the electromagnetic field generated in the soil due to the electric current applied.

The data logger software makes the conversion of the voltage drop verified in the soil into EC, recorded as mS m^{-1}. The EC data acquisition system is connected to a GPS receiver with differential correction provided by geostationary satellite (DGPS). Readings emitted by the sensor are recorded once per second, totalizing 12,709 points for the experimental field, and linked to geographic position.

Georeferenced soil samples were collected in the winter of 2001, at a 0–0.1 m depth, totaling 71 samples for a sample density of 1.9 samples ha^{-1}. Determinations were made for: soil pH, determined in a 0.01 M CaCl$_2$ solution; organic matter (OM) (g dm^{-3}), determined by the Walkley-Black method; phosphorus (P) (mg dm^{-3}), extracted by resin; K (mmol$_c$ dm^{-3}), Ca (mmol$_c$ dm^{-3}) and Mg (mmol$_c$ dm^{-3}), extracted by resin, and clay (g kg^{-1}) and sand (g kg^{-1}), determined by the total dispersion method. In addition to these attributes, calculations were made for the sum of bases (SB) (mmol$_c$ dm^{-3}), cation exchange capacity (CEC) (mmol$_c$ dm^{-3}), and base saturation (V) (%).

In order to analyze the data, experimental semivariograms were initially built for the distributions of EC and of the other soil variables and the semivariogram classic estimator, or moments method (Isaaks & Srivastava, 1989) was used. After analyzing the spatial dependence of these variables, estimates were obtained for unsampled locations, in order to generate surface maps. Ordinary kriging was used to digitize and visualize the information.

To verify possible relations between variables, the Pearson's coefficient of linear correlation was calculated between the soil physicochemical variables and soil EC at both reading depths, using the values of cells interpolated by the ordinary kriging technique as data for each of them.

A principal components analysis (PC) was conducted, which is a multivariate statistical technique that linearly transforms a set of data of several variables. In practice, this technique transforms interdependent variables into independent and significant ones. This linear transformation allowed the original data set to be compressed into a substantially smaller set of non-correlated variables, the PCs, which represent most of the information contained in the original data set (Afifi & Clark, 1996). This technique was used to determine which soil variables (including EC) were the most important for soil variability characterization and, by using the new variables (PCs) obtained from it, to implement the classification process, in order to define soil management zones. PCs were selected after an evaluation of the various criteria presented by Afifi & Clark (1996); PCs whose correlation matrix cumulative values explained 80% of the total variance of data were retained. This analysis was made using a commercial statistical software.

The next step consisted in classifying the selected PCs aimed at identifying the natural clusters. This classification process was used as a method for clustering the principal components, where each natural cluster comprises a distinct soil management zone. The fuzzy k-means algorithm was used - FuzME program (Minasny & McBratney, 2000). Initially, it was necessary to select a distance criterion (Euclidean, mahalanobis, or diagonal) and the fuzzy exponents that

measure the degree of overlapping between groups. Information on this technique can be found in Burrough et al. (1997) and Minasny & McBratney (2000).

Last, an analysis of variance (ANOVA) was run for the EC variables and other soil attributes between the different management zones, to verify whether there were significant differences between variable means in each zone. The zone to which each individual (cell) belonged was used as a dependent variable in the ANOVA. The Tukey HSD test ("Honest Significant Difference") was used to make comparisons between management zones for samples of different sizes; this analysis was performed using a commercial statistical software.

RESULTS AND DISCUSSION

The chemical and physical properties from soil samples are presented on Table 1, and the pH the coefficient of variation shows exceptionally a large variability within samples. Semivariograms were calculated for the soil EC variables at both depths, as well as for the other soil properties. The following parameters are presented in Table 2: nugget effect, sill, range, and model for the calculated semivariograms, in addition to the error sum of squares values, which was the criterion adopted to select the best fit for each model and structural component ($C1/C0+C1$), representing the amount of data variance which can be explained by spatial dependence (Isaaks & Srivastava, 1989).

Table 1 - Statistics of chemical and physical properties from soil samples (OM = organic matter; SB = sum of bases; CEC = cation exchange capacity; V = base saturation).

Variable	Mean	S.D.	Skewness	CV
P (mg dm^{-3})	50.20	30.56	1.31	60.88
OM (g dm^{-3})	38.17	11.00	-0.57	28.81
pH	4.89	0.15	0.73	3.11
K (mmol$_c$ dm^{-3})	2.55	1.02	0.28	40.11
Ca(mmol$_c$ dm^{-3})	29.47	8.99	0.42	30.51
Mg(mmol$_c$ dm^{-3})	15.36	6.04	0.92	39.32
SB(mmol$_c$ dm^{-3})	47.38	15.57	0.59	32.87
CEC(mmol$_c$ dm^{-3})	99.49	19.99	-0.34	20.09
V%	46.80	8.68	0.14	18.55
Clay (g kg^{-1})	284.60	10.63	-0.40	37.35
Sand (g kg^{-1})	596.10	15.39	0.36	25.82

Table 2 - Semivariogram parameters for shallow (0 - 0.3 m) and deep (0 - 0.9 m) soil EC readings and for other soil physicochemical variables (OM = organic matter; SB = sum of bases; CEC = cation exchange capacity; V = base saturation).

Variable	Nugget effect (C0)	Sill (C0+C1)	Range (a)	SSE[1]	C1/C0+C1[2]	Model[3]
EC (0-0.3 m)	0.30	1.09	15.85	0.05	0.72	Exp.
EC (0-0.9 m)	0.08	0.31	11.65	$3*10^{-3}$	0.74	Exp.
P	223.2	542.7	253.5	9,045.6	0.59	Sph.
OM	4.9	63.0	375.8	123.3	0.92	L/w sill
pH	0.01	0.03	579.6	$1*10^{-5}$	0.67	Gaus.
K	0.41	0.68	352.2	$8*10^{-3}$	0.40	Sph.
Ca	16.13	83.32	340.6	57.40	0.81	Exp.
Mg	3.51	30.51	198.2	8.00	0.88	Exp.
SB	35.42	173.52	393.8	363.8	0.80	Sph.
CEC	1.73	149.83	253.6	849.5	0.99	Sph.
V%	40.59	67.53	375.8	87.42	0.40	Sph.
Clay	15.46	-	-	$7*10^{-3}$	-5	Power
Sand	18.55	-	-	774.9	-	Power

[1]Error sum of squares; [2]Structural component; [3]Exp.= Exponential; Sph.= Spherical; L/w sill = Linear with sill; Gaus.= Gaussian; Power = Power function.

In general, the spatial dependence of these variables explained a large part of their variation, as can be observed from the structural component values (Table 2). In addition, a difference can be noted between the low sum of square errors (SSE) for the EC variable at both depths, as opposed to the higher values of this index, on average, for the other soil variables analyzed in the laboratory. This fact is explained by the sample density difference between the first variables (355 samples ha^{-1}) in relation to the latter (1.9 samples ha^{-1}), causing a more erratic behavior of the semivariance function for the variables in the second group, consequently with poorer model fitting. One of the advantages of soil EC monitoring relies on this fact: it allows many samples to be collected quickly and at low cost, and generates maps that model soil spatial variability in a more precise manner.

With the semivariograms, surface maps could be generated for each variable using kriging interpolation. Figure 1 presents some of these maps, specifically for the shallow (0–0.3 m) and deep EC values (0–0.9 m), and for clay and sand contents. To the observer, the similarity between these maps should be clear. The "softer" appearance of the distribution of soil texture in relation to soil EC is caused by the sample density difference. A visual inspection of these maps indicates a good agreement between both. This relationship between EC and soil texture was previously reported by Williams & Hoey (1987) and, recently, in a field near by this one, by Machado et al. (2006).

The correlation between soil EC for the shallow and deep readings and soil physicochemical attributes was evaluated based on the generated surface maps (Table 3). Regarding the relation between EC in the shallow reading and the soil attributes analyzed in the laboratory, a number of medium and strong correlations could be identified between EC and P, OM, pH, K, Ca, Mg, SB, CEC, V%, clay, and sand. Several medium and high-intensity correlations were also verified between deep-reading EC and the following soil attributes: P, OM, pH, K, Ca, Mg, SB, CEC, V%, clay, and sand. Specifically, the strong relation between EC and soil texture is a promising indication toward its use to define soil management zones in the study area.

From the correlations between soil variables and soil EC at both depths, a number of correlations with intensities from moderate to strong can be identified between them. These interactions are limiting factors for the use of simple correlation analysis in the interpretation of these data. Principal component analysis was then used to reduce the number of variables, which are to be used later to generate management zones, without loss of relevant information. Trials were also done to evaluate whether EC, at both depths, could somehow help to define soil management zones (Table 4).

Two PCs were selected among the total set of 13 original variables, which together explained 84.5% of the total variance of those data. The first PC (responsible for 76.7% of variability) is strongly influ-
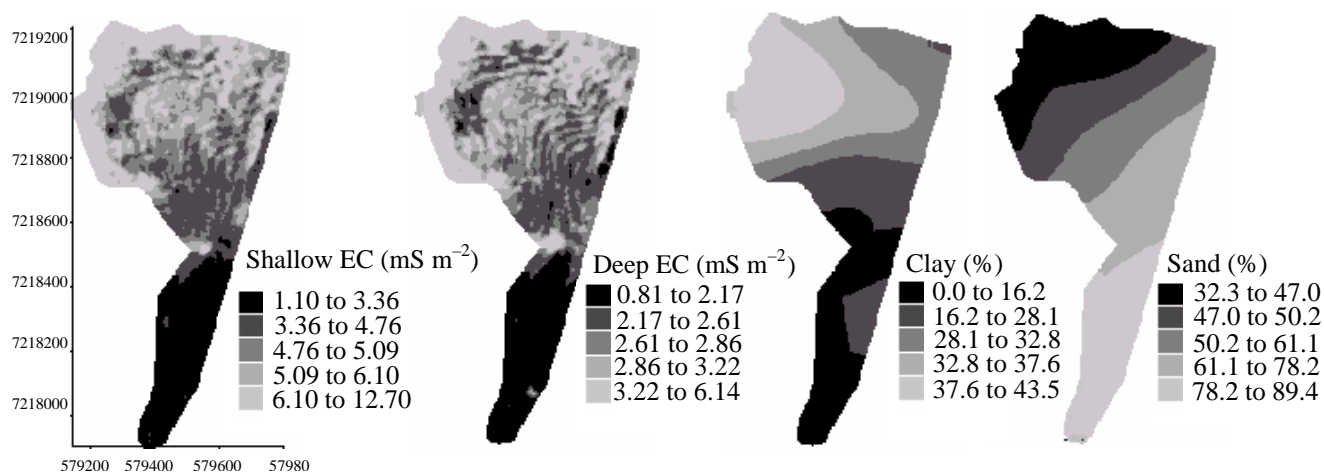

Figure 1 - Shallow (0 – 0.3 m) and deep-reading (0 – 0.9 m) soil EC maps, and soil clay and sand content maps.

Table 3 - Coefficients of correlation between shallow (0 - 0.3 m) and deep-reading (0 - 0.9 m) soil EC and other soil physicochemical attributes (OM = organic matter; SB = sum of bases; CEC = cation exchange capacity; V = base saturation).

Variables	P	OM	pH	K	Ca	Mg	SB	CEC	V%	Clay	Sand
EC (0-0.3 m)	-0.68	0.86	-0.41	0.87	0.84	0.63	0.84	0.77	0.76	0.75	-0.86
EC (0-0.9 m)	-0.64	0.82	-0.42	0.81	0.81	0.60	0.81	0.74	0.73	0.66	-0.79

enced by all original variables, with the exception of pH. This can be viewed as the inherent fertility potential of the soil. Due to the direct relation between the original variables and this PC, regions with higher values for this PC are the most fertile. Furthermore, it is important to highlight the great intensity of the influence of both shallow and deep soil EC readings on this PC (having correlations with the first principal component equal to 0.91 and 0.87, respectively), indicating the importance of this information in helping to explain the soil general physicochemical spatial variability of the area. The second PC showed a more intense relation with pH only, approximately the contrary of the first PC, for which the relative influence of pH was low.

In the following step, the fuzzy k-mean continuous classification technique was used to define soil management zones based on the two PCs presented in Table 4. The Euclidean distance criterion and a fuzzy exponent of 1.2 were used. The Fuzzy Performance Index (FPI) and the Modified Partition Entropy (MPE) were used to determine the optimal number of management zones in each area. These two indices are provided by the FuzMe program itself (Minasny & McBratney, 2000). The optimal number of cluster classes (management zones) is then defined as the number at which these two indices reach their minimum value (Fridgen et al., 2001). FPI and MPE indices reached a minimum value for a di-

vision of the area into three classes (three management zones), indicating that this is the optimal number of soil management zones based on the sampled variables (Figure 2).

The two previously selected PCs were then submitted to fuzzy classification, where an individual could have total, partial, or null participation in each of the different classes. The maps for the three clusters with their respective management zones resulting from this classification are presented in Figure 3.

The final fuzzy classification result provides participation function values for each individual (datum) in the original set of data for each of the generated classes. This function may assume values between 0.0 (no possibility of participation in the class) and 1.0 (total participation in the class). In this study, the cluster of individuals whose participation function in the
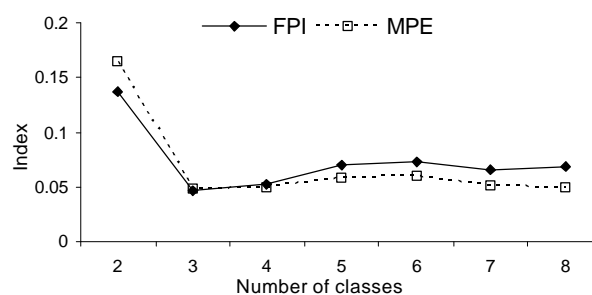

Figure 2 - Cluster performance graph as a function of number of classes.

Table 4 - Principal components analysis for the soil physicochemical variables collected (OM = organic matter; SB = sum of bases; CEC = cation exchange capacity; V = base saturation).

Variance components	Principal components	
	1	2
Variance	9.98	1.01
Proportion (%)	76.75	7.78
Cumulative proportion (%)	76.75	84.53
Variables	Correlation with principal components	
EC (0-0.3 m)	0.91	-0.06
EC (0-0.9 m)	0.87	-0.03
P	-0.83	-0.38
OM	0.96	0.14
pH	-0.51	-0.79
K	0.98	0.03
Ca	0.95	-0.17
Mg	0.76	-0.19
SB	0.95	-0.21
CEC	0.90	0.05
V%	0.87	-0.32
Clay	0.85	-0.03
Sand	-0.95	0.04

same given class was higher than 0.50 were considered belonging to a distinct management zone, as can be seen in maps from A through C in Figure 3.

The spatial continuity of the management zones presented in Figure 3 is quite large, and may be considered an indication that the classification was successful, since it will later facilitate management operations in the area. It is interesting to compare the management zone maps for this area against the spatial distribution maps for clay and sand (Figure 1). Based on the comparison between these information layers it becomes easy to notice the influence of soil texture on the soil's behavior. Due to the relation between EC measurements at both depths with texture (r = 0.75 and 0.66 for the ratios between shallow and deep EC readings with soil clay content, respectively), the legitimacy of this information in defining management zones for this area can be observed.

The last step of the study consisted in running an analysis of variance (ANOVA) in order to verify whether there were differences between the EC variables at both depths and soil physicochemical characteristics among the different manage-ment zones created by the fuzzy k-means algorithm. The greater the differences the more the validity of the performed division is confirmed; the differences also confirm whether the process of reduction of the number of variables (principal components analysis) was able to suitably express their spatial variability model. The ANOVA results for soil variables among management zones are presented in Table 5.

There is a considerable relation between shallow and deep soil EC and soil clay content, with the means for these variables among zones arranged in ascending order. In addition, most soil attributes followed this tendency, with zone 3 showing the highest soil fertility indices. The differences found between zones for all these variables were significant, demonstrating the viability of the proposed classification, and indicating that a differentiated management is possible between zones for the sampled soil attributes. However, more researches showing the utility of EC with the same objective for other locations with soil characteristics diverse from those found in this study are needed.
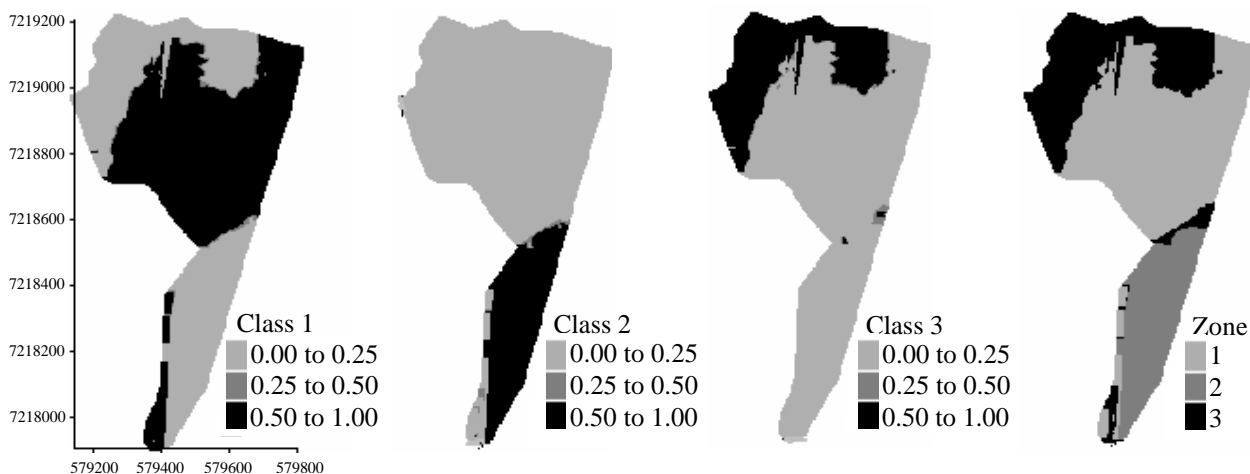

Figure 3 - Maps showing the spatial distribution of participation function values for each individual in the three classes generated after classification by the fuzzy-k-means algorithm of two principal components selected and corresponding map showing the resulting management zones.

Table 5 - Mean soil physicochemical variables per management zone (OM = organic matter; SB = sum of bases; CEC = cation exchange capacity; V = base saturation).

Zone	Variables*†												
	ECs	ECd	P	OM	pH	K	Ca	Mg	SB	CEC	V%	Clay	Sand
1	5.2[b]	2.8[b]	40.2[b]	40.5[a]	4.8a	2.0c	2.7[b]	29.6[b]	15.0[b]	47.0[b]	46.4[b]	29.7[b]	57.9[b]
2	2.4[a]	1.7[a]	93.1[c]	20.1[b]	5.0[c]	1.0[a]	1.3[a]	19.4[a]	9.6[a]	29.5[a]	41.5[a]	13.1[a]	83.9[c]
3	6.6[c]	3.4[c]	38.9[a]	46.7c	4.9[b]	1.6[b]	3.2[c]	37.0[c]	20.2[c]	61.7[c]	51.7[c]	37.5[c]	43.7[a]

*ECs = shallow-reading electrical conductivity (0 - 0.3 m); ECd = deep-reading electrical conductivity (0 - 0.9 m); Clay = clay content; Sand = Sand content. †Variable means in the same column followed by different letters are different between management zones by Tukey HSD test for unequal samples (P = 0.05).

CONCLUSIONS

A manner to define differentiated soil management zones was achieved in this work using EC and other soil properties, in addition to geostatistics, principal component analysis, and fuzzy logic to handle the data. It was proved to be a viable procedure for the area, allowing the delimitation of homogeneous and distinct regions among them with reference to the soil attributes and the resulting three management zones represent a reasonable number for practical use. The relations between shallow and deep soil EC and clay content indicated the viability of using this information to delineate soil management zones, since it is the main factor controlling the spatial variability of other soil fertility indicators.

ACKNOWLEDGMENTS

REFERENCES

AFIFI, A.A.; CLARK, V. **Computer-aided multivariate analysis**. New York: Chapman & Hall, 1996. 455 p.

BURROUGH, P.A.; GAANS, P.F.M. van; HOOTSMANS, R. Continuous classification in soil survey: spatial correlation, confusion and boundaries. **Geoderma**, v.77, p.115-135, 1997.

FRIDGEN, J.J.; KITCHEN, N.R.; SUDDUTH, K.A. Variability of soil and landscape attributes within sub-field management zones. In: INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE, 5., Bloomington, 2000. **Proceedings.** Madison: ASA-CSSA-SSSA, 2001. CD-ROM.

ISAAKS, E.H; SRIVASTAVA, R.M. **Applied geostatistics**. New York: Oxford University Press, 1989. 561p.

LUCHIARI JR., A.; SHANAHAN, J.; FRANCIS, D.; SCHLEMMER, M.; SCHEPERS, J; LIEBIG, M.; SCHEPERS A.; PAYTON, S. Strategies for establishing management zones for site specific nutrient management. In: INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE, 5., Bloomington, 2000. **Proceedings.**, Madison: ASA-CSSA-SSSA, 2001. CD-ROM.

MACHADO, P.L.O.A.; BERNARDI, A.C.C.; VALENCIA, L.I.O.; MOLIN, J.P.; GIMENEZ, L.M.; SILVA, C.A.; ANDRADE, A.G.A.; MADARI, B.E.; MEIRELLES, M.S.P.M. Mapeamento da condutividade elétrica e relação com a argila de Latossolo sob plantio direto. **Pesquisa Agropecuária Brasileira**, v.41, p.1023-1031, 2006.

McBRIDE, R.A; GORDON, A.M; SHRIVE, S.C. Estimating forest soil quality from terrain measurements of apparent electrical conductivity. **Soil Science Society of American Journal**, v.54, p.255-260, 1990.

MINASNY, B.; McBRATNEY, A.B. FuzME version 2.1. 2000. Available at: http://www.usyd.edu.au/su/agri/acpa. Accessed 23 Feb. 2002.

NADLER, A; FRENKEL, H. Determination of soil solution electrical conductivity from bulk soil electrical conductivity measurements by the four electrode method. **Soil Science Society of American Journal**, v.44, p.1216-1221, 1980.

SHEETS, K.R; HENDRICKX, J.M.H. Noninvasive soil water content measurement using electromagnetic induction. **Water Resources Research**, v.31, p.2401-2409, 1995.

WILLIAMS, B.G.; HOEY, D.; The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. **Australian Journal of Soil Research**, v.25, p.21-27, 1987.