REDUNDANT VARIABLES AND THE QUALITY OF MANAGEMENT ZONES

RICARDO SOBJAK[1], EDUARDO G. DE SOUZA[2], CLAUDIO L. BAZZI[3], MIGUEL A. URIBE-OPAZO[4], NELSON M. BETZEK[5]

ABSTRACT: Precision agriculture (PA) allows farmers to identify and address variations in an agriculture field. Management zones (MZs) make PA more feasible and economical. The most important method for defining MZs is a fuzzy C-means algorithm, but selecting the variable for use as the input layer in the fuzzy process is problematic. BAZZI et al. (2013) used Moran's bivariate spatial autocorrelation statistic to identify variables that are spatially correlated with yield while employing spatial autocorrelation. BAZZI et al. (2013) proposed that all redundant variables be eliminated and that the remaining variables would be considered appropriate on the MZ generation process. Thus, the objective of this work, a study case, was to test the hypothesis that redundant variables can harm the MZ delineation process. BAZZI This work was conducted in a 19.6-ha commercial field, and 15 MZ designs were generated by a fuzzy C-means algorithm and divided into two to five classes. Each design used a different composition of variables, including copper, silt, clay, and altitude. Some combinations of these variables produced superior MZs. None of the variable combinations produced statistically better performance that the MZ generated with no redundant variables. Thus, the other redundant variables can be discredited. The design with all variables did not provide a greater separation and organization of data among MZ classes and was not recommended.

KEYWORDS: precision agriculture, spatial correlation, relative efficiency.

VARIÁVEIS REDUNDANTES E A QUALIDADE DE ZONAS DE MANEJO

RESUMO: A agricultura de precisão proporciona aos agricultores identificar e tratar de forma adequada as variações encontradas na área agrícola. As zonas de manejo (ZMs) permitem a implantação da agricultura de precisão de forma viável e relativamente mais econômica. A forma mais importante para definir ZMs é usando o algoritmo fuzzy C-means. Um problema consiste em como selecionar a variável a ser usada como *layer* de entrada no processo fuzzy. Assim, o objetivo deste trabalho, foi testar a hipótese de que variáveis redundates podem prejudicar o processo de delineamento de ZMs. Este trabalho foi desenvolvido em uma área de 19,6 ha e 15 agrupamentos de ZMs foram gerados por meio do o algoritmo fuzzy C-means, dividindo-se em duas a cinco classes. Cada agrupamento usou uma composição diferente de variáveis, que são os atributos cobre, silte, argila, e altitude. Foi encontrado que algumas combinações dessas variáveis produziu melhores ZMs. Nenhuma combinação de variáveis produziu desempenho estatisticamente melhor que a ZM gerada apenas com as variáveis não redundantes. Assim, as variáveis redundantes podem ser descartadas. O agrupamento com todas as variáveis não forneceu maior separação e organização dos dados entre as classes de ZM, não sendo recomendado.

PALAVRAS-CHAVE: agricultura de precisão, correlação espacial, eficiência relativa.

[1] Technologist in Data Processing, MSc in Agricultural Engineering, Computer Science Departament, Technological Federal University of Paraná, Av. Brasil, 4232, 85884-000, Medianeira, Paraná, Brazil, ricardosobjak@utfpr.edu.br

[2] Mechanical Engineer, PhD in Mechanical Engineering, PGEAGRI, UNIOESTE - Western Paraná State University, Productivity researcher of the CNPq. Rua Universitária, 2069. 85819-110. Cascavel, Paraná. Brazil, Eduardo.souza@unioeste.br

[3] Technologist in Data Processing, PhD in Agricultural Engineering, Computer Science Departament, Technological Federal University of Paraná, Av. Brasil, 4232. 85884-000. Medianeira, Paraná. Brazil.

[4] Statistician, PhD in Statistics, Associated Professor, PGEAGRI, UNIOESTE - Western Paraná State University, Productivity researcher of the CNPq. Rua Universitária, 2069. 85819-110. Cascavel, Paraná. Brazil.

[5] Technologist in Data Processing, MSc in Agricultural Engineering, Computer Science Departament, Technological Federal University of Paraná, Av. Brasil, 4232. 85884-000. Medianeira, Paraná. Brazil.

INTRODUCTION

The spatial and temporal variability of soil attributes and weather conditions may affect soybean development (GUASTAFERRO et al., 2010). The spatial and temporal variations that affect production must be elucidated so that these variations can be considered in crop management. Precision agriculture (PA) is an agricultural management system based on spatial variations of soil and crops attributes in production fields. PA aims to optimize profitability, sustainability and environmental protection (MOLIN et al., 2010) by employing localized application of inputs to an agricultural field. Among studies seeking to evaluate the economic viability of PA, those that incorporate management zones (MZs) aim to analyze and provide recommendations on the division of production areas into smaller MZs that are treated differently. Such practices allow farmers to use the same systems used in conventional agriculture during crop management (RODRIGUES JUNIOR et al., 2011). According to ROUDIER et al. (2011), MZs simplify the representation of spatial variability in a cropping field. The management of such small and uniform regions (MZs) is considered by MOLIN & CASTRO (2008) one of the most challenging stages of PA.

Clustering methods are highly recommended for defining MZs (YAN et al., 2007; ILIADIS et al., 2010) and include the use of several attributes such as electrical conductivity, elevation, slope and soil texture, and nitrogen alone and in combination. Although any attribute may be related to crop yield, for DOERGE (2000), the ideal attribute is the correlation of predictable spatial information sources with yield. Clustering techniques for MZ generation include algorithms such as K-Means and Fuzzy C-Means (ILIADIS et al., 2010; VALENTE et al., 2012 and LI et al., 2013), which offer good results (VITHARANA et al., 2008; MORARI et al., 2009; MORAL et al., 2010; RODRIGUES JUNIOR et al., 2011; DAVATGAR et al., 2012; KWEON, 2012; BANSOD & PANDEY, 2013), which permit the automatic division of the studied field. In this approach, different data sources that are related to crop development factors can be used to generate MZs.

The most used unsupervised clustering algorithm is Fuzzy C-means (also known as Fuzzy K-means). This algorithm uses an iterative process to recalculate the cluster means and assign data points to clusters. Fuzzy C-means uses a weighting exponent to control the degree to which membership sharing occurs between classes (BEZDEK, 1981). This approach is important because it allows individuals to exhibit partial membership in each of a number of sets, thus enabling the study of continuous variability in natural phenomena (BURROUGH, 1989). Before a data cluster can be formed, an appropriate measure of similarity must be established, which is typically the normalized distance from an observation to the cluster mean in attribute space (TOU & GONZALEZ, 1974; JOHNSON, 1998). One of the more frequently used measures of similarity is the Euclidean distance, which gives equal weight to all measured variables and is sensitive to correlated variables (BEZDEK, 1981). Geometrically, the Euclidean distance generates clusters with a spherical shape, which rarely occurs in a real soil system (ODEH et al., 1992).

The influence of a variable (attribute) on yield must be measured before it can be selected and considered in the MZ determination process. The criteria to select a variable must consider its required spatial autocorrelation and the spatial correlation of the variable with yield. BAZZI et al. (2013) used Moran's bivariate spatial autocorrelation statistic to propose a procedure for selecting variables as input data for the Fuzzy C-Means algorithm.

Thus, this trial aims to test the hypothesis that redundant variables can harm the MZ delineation process.

MATERIAL AND METHODS

This work was conducted in a 19.6-ha commercial field from April 2009 to January 2011. The experimental field was cropped with soybean. The field is located in southern Brazil, western Paraná, municipality of Cascavel (Figure 1). Its geographical coordinates are 24°57'19" S and 53°33'59" W, with an average altitude above sea level of 706 m.

Soil samples were collected at 55 different sites in the field (2.8 points ha^{-1}, Figure 1) to obtain data on soil resistance to penetration (SRP) and soil chemical properties at a depth of 0 to 0.20 m. SRP was measured with a penetrometer Falker SoloTrack PLG1020. The sampling elements of soil texture were collected from 45 of the 55 sites. Soil sampling was conducted with the aid of an auger at a depth of 0-0.2 m, and for each of the sampling points, eight sub-samples were collected within a radius of 3 m from the point determined by the sampling grid (adapted from WOLLENHAUPT et al., 1994).

The quantitative values obtained were potential hydrogen (pH), phosphorus (P), potassium (K), copper (Cu), zinc (Zn), iron (Fe), manganese (Mn), calcium (Ca), magnesium (Mg), carbon (C), aluminum (Al) and the contents of sand, clay and silt. The soybean yield data were collected from 130 sampling points manually collected and corrected at 12% water content. Each sample point was represented by the total mass collected on two lines in a path of one meter and, because the spacing was 0.45 m, each sample point was represented by an area of approximately 0.9 m^2.
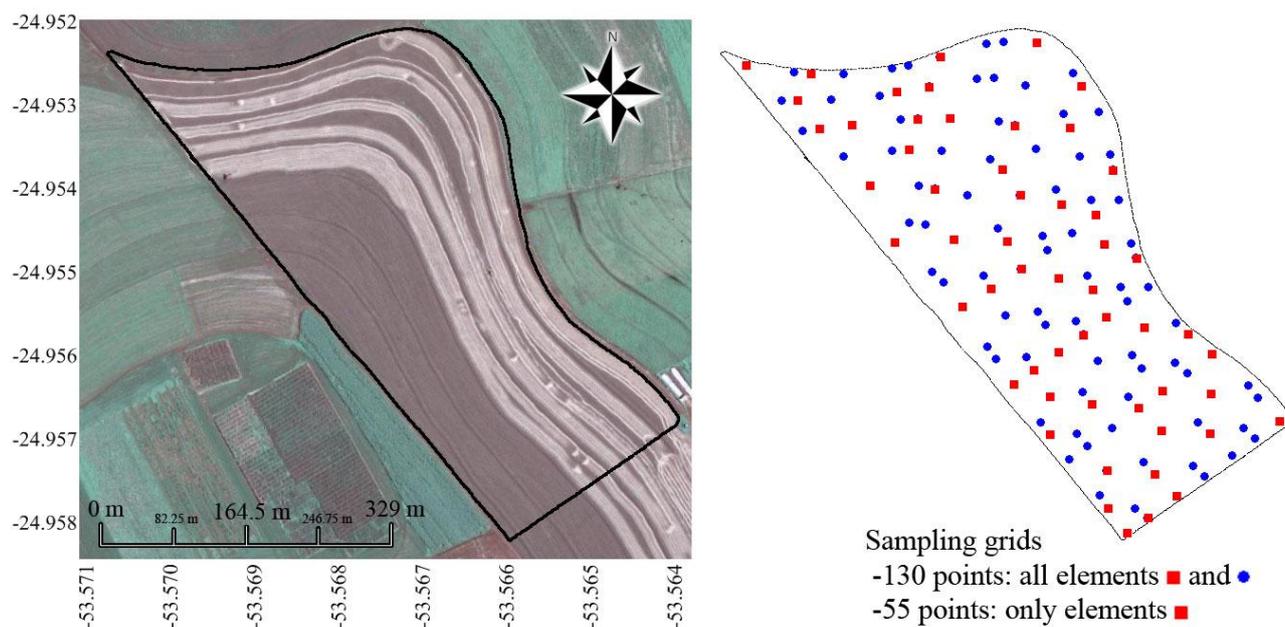
Source: Adapted from GOOGLE EARTH (2009).

FIGURE 1. Experimental field and its sampling grid. All red and blue elements represent the 130-point grid; red-only elements represent the 55-point grid. Municipality of Cascavel, Paraná, Brazil.

Based on statistical descriptive analysis of the data, measures of central tendency (mean and median) and measures of dispersion (variance, standard deviation and coefficient of variation) were determined. To evaluate the data normality at 5% probability, Anderson-Darling and Kolmogorov-Smirnov tests were performed, and those with normality for at least one of the tests were considered normal. To evaluate the correlation among chemical and physical properties of soil, topography and soybean yield, Moran's bivariate spatial autocorrelation statistic I_{YZ} (BONHAM et al., 1995), shown in [eq. (1), was applied. The hypothesis of no spatial autocorrelation was tested at the 5% significance level.

$$I_{YZ} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}.Y_i.Z_j}{S\sqrt{m_Y^2.m_Z^2}} \quad (1)$$

where,

w_{ij} is an element of matrix W of spatial association, which measures the association among elements at positions i and j (calculated by $w_{ij}=(1/1+D_{ij})$) so that D_{ij} is the distance between i and j points;

Y_i - is the value of variable Y transformed at point i ($i=1,…,n$) to obtain a zero mean, according to $Y_i = (Y_i - \overline{Y})$, where $\overline{Y}$ is the sampling mean of Y variable;

Z_j is the value of Z variable transformed at point j ($j=1,…,n$) to obtain a zero mean from the formula $Z_j = (Z_j - \overline{Z})$, where $\overline{Z}$ is the sample mean of Z variable;

S is the sum of all degrees of spatial association, obtained from elements w_{ij} for $i \neq j$;

m_Y^2 is the sample variance of variable Y_i, and

m_Z^2 is the sample variance of variable Z_j.

The spatial correlation matrix was generated for the analyzed variables; in the matrix, the autocorrelation values of each variable are the elements on the main diagonal, whereas the off-diagonal elements are the cross-correlations among variables. The coefficients were tested at the 1% and 5% significance levels. BAZZI et al. (2013) proposed the following procedure for selecting variables as input data for the Fuzzy C-Means algorithm. 1) Variables with spatial dependence and no significance at 95% significance are eliminated. 2) From variables with spatial dependence, those with no correlation with soybean yield are removed. 3) The descending order is calculated considering the degree of correlation with yield. 4) The redundant variables are eliminated (those that correlate with each other), giving preference to those variables with lower correlation with yield.

In this work, the redundant variables were not discarded, but different combinations of the variables were built, and each combination was defined as a MZ design. The combinations were constructed to assess whether redundant variables can harm the MZ delineation process.

According to the sample data for the selected variables, values were predicted for non-sampled regions in the studied field by the inverse of square distance. The data for the selected variables were normalized (Equation 2, MIELKE & BERRY, 2007) to ensure that no variable had greater weight in the process of dividing the field into MZs.

$$P_n = \frac{P_i - Median}{Amplitude} \tag{2}$$

where,

P_i is the pixel to be normalized and P_n is the normalized pixel.

For each design, the field was divided into MZs according to an unsupervised classification fuzzy C-means algorithm (CANNON et al., 1986) using the software FuzMe (MINASNY & MCBRATNEY, 2002). The software parameters included a minimum of two and a maximum of five MZ classes, Euclidean distances, a fuzzy exponent of 1.30, a maximum number of 300 interactions, and stopping criterion or error $\varepsilon = 0.0001$. The Euclidean distance was adopted because the data were already normalized. The others constants were used as suggested.

The yields of the MZ designs were evaluated to determine the best way to divide the field. The following rating methods were used:

1) Variance Reduction: (PING; DOBERMANN, 2003; XIANG et al., 2007), (Equation 3) for yield was carried out, with expectations that the sum of the variances of the data of the MZs is lower than the total variance.

$$VR = 1 - \frac{\sum_{i=1}^{c}\left(W_i * V_{ZM_i}\right)}{V_{Field}} * 100 \qquad (3)$$

where,

c is the number of management zones;

W_i is the proportion of area each management zone;

V_{ZM_i} is the variance of data each management zone, and

V_{Field} is the sample variance of data to field.

2) Analysis of variance: used to verify that there are significant mean differences in yield among MZ classes in each proposed design (XIANG et al., 2007; MOLIN & CASTRO, 2008; XIN-ZHONG et al., 2009). The HSD (Honest Significant Difference) Tukey test was applied at the 5% level of significance. In each internal MZ class, the yield was considered an independent variable and normally distributed.

3) Cluster validation indices: FPI (Fuzziness Performance Index) and MPE (Modified Partition Entropy) indices were used (McBRATNEY & MOORE, 1985, Equations 4 and 5) and were provided by FuzME software to assess whether the MZs with greater degrees of separation and organization among classes also corresponded to those with better RE indices. These indices have been applied in studies related to PA (MOLIN & CASTRO, 2008; FU et al., 2010; GUASTAFERRO et al., 2010; ARNO et al., 2011). The cluster class (MZ) with the greatest differentiation was the one in which these two indices reached approximately the minimum in each design.

$$FPI = 1 - \frac{c}{(c-1)}\left[1 - \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{ij})^2\right] \qquad (4)$$

$$MPE = -\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{c}\mu_{ij}\log(\mu_{ij}) \qquad (5)$$

where,

c is the number of clusters;

n is the number of observations, and

μ_{ij} is element ij of fuzzy matrix U.

4) The Kappa index has been used to measure the accuracy of thematic classifications (BAZZI et al., 2008; BASTIANI et al., 2012; DALPOSSO et al., 2012) and is recommended as an appropriate measure of accuracy for all elements from the error matrix. It was used to evaluate the spatial agreement among maps of more efficient MZs (Equation 6).

$$K = \frac{n\sum_{i=1}^{r}x_{ii} - \sum_{i=1}^{r}(x_{i+}x_{+i})}{n^2 - \sum_{i-1}^{r}(x_{i+}x_{+i})} \qquad (6)$$

where,

K is the Kappa agreement index;

n is the number of observations (sample points);

r is the number of classes in the error matrix;

x_{ii} is the number of combinations of row i and column i;

x_{i+} is the total observations of row i; x_{+i} is the total observations of column i.

The classification degree for Kappa index was defined according to LANDIS & KOCH (1977) as follows: very strong, values between 0.81 to 1.00; strong, values from 0.61 to 0.80; moderate, values between 0.41 to 0.60; weak, values between 0.21 to 0.40; and no agreement, values between 0.00 and 0.20.

RESULTS AND DISCUSSION

The contents of C, pH, clay content and altitude (Table 1) were highly homogeneous (PIMENTEL GOMES & GARCIA, 2002), whereas soybean yield and the content attributes of Fe, Mn, Ca, silt, and soil resistance to penetration into layers at depths of 0 - 0.20 m (SRP_0_20) and 0.10 - 0.20 m (SRP_10_20) exhibited average homogeneity. The sample data for Mg content exhibited low homogeneity. Finally, Cu, Zn, P, Al, K, sand and soil resistance to penetration in the 0 - 0.10 m layer (SRP_0_10) exhibited heterogeneity.

The soil texture (Table 1) was at least 60% clay, indicating a loamy soil. The field also exhibited significant compaction in the 0.10 - 0.20 m layer and greater resistance to penetration, as indicated by an average value of 1.94 MPa and a maximum value of 2.40 MPa. The lowest soil compaction in the field was 0.31 MPa in the layer from 0 to 0.10 m.

TABLE 1. Descriptive statistics of the data.

Attributes	N	Minimum	Mean	Median	Maximum	Standard deviation	CV (%)	Normal*
Yield (t ha^{-1})	130	2.09	3.29	3.29	4.64	0.52	16 (m)	Yes
Cu content (mg dm^{-3})	55	1.00	2.24	2.11	4.21	0.93	41 (vh)	Yes
Zn content (mg dm^{-3})	55	1.00	1.61	1.32	4.04	0.79	49 (vh)	No
Fe content (mg dm^{-3})	55	14.00	18.40	18.00	30.00	2.85	16 (m)	Yes
Mn content (mg dm^{-3})	55	37.00	49.65	49.00	68.00	6.69	13 (m)	Yes
P content (mg dm^{-3})	55	6.40	15.12	13.80	39.60	7.05	47 (vh)	No
C content (g dm^{-3})	55	28.05	33.08	33.12	37.79	2.39	7 (l)	Yes
pH	55	4.50	5.24	5.20	6.00	0.35	7 (l)	Yes
Ca content (cmolc dm^{-3})	55	3.79	6.74	6.75	9.90	1.34	20 (m)	Yes
Mg content (cmolc dm^{-3})	55	1.44	2.22	2.12	3.82	0.54	24 (h)	Yes
Al content (cmolc dm^{-3})	55	0.00	0.04	0.00	0.61	0.11	260 (vh)	Yes
K content (cmolc dm^{-3})	55	0.09	0.28	0.28	0.60	0.11	38 (vh)	Yes
SRP_0_10 (MPa)	55	0.31	0.62	0.63	1.09	0.20	31 (vh)	Yes
SRP_0_20 (MPa)	55	0.89	1.28	1.29	1.65	0.18	14 (m)	Yes
SRP_10_20 (MPa)	55	1.43	1.94	1.96	2.40	0.20	11 (m)	Yes
Altitude (m)	55	701.0	706.3	702.0	712.0	3.33	0.5 (l)	Yes
Clay content (%)	45	60.00	70.00	72.00	79.00	5.00	7 (l)	No
Sand content (%)	45	6.00	10.00	9.00	19.00	4.00	39 (vh)	No
Silt content (%)	45	15.00	20.00	19.00	25.00	2.00	11 (m)	Yes

N - Number of sampling units; *: Not Normal at 5% significance; coefficient of variation (CV): low (l), medium (m), high (h), very high (vh); SRP_0_10: soil resistance to penetration at a depth of 0 - 0.10 m; SRP_0_20: soil resistance to penetration in the layer at a depth of 0 - 0.20 m; SRP_10_20: soil resistance to penetration in the 0.10 - 0.20 m layer.

Most of the studied variables exhibited a normal distribution at 5% significance, including yield, Cu, Fe, Mn, C, pH, Ca, Mg, Al, K, SRP_0_10, SRP_0_20, SRP_10_20, altitude and silt (Table 1). The exceptions were the contents of Zn, P, clay and sand. Previous studies have also observed non-normality for clay (CORÁ et al., 2014), silt (ZUCOLOTO et al., 2011), P and Zn (LEÃO et al., 2010). No transformation was necessary for the variables that were not normal because these variables were not correlated with yield.

The average yield obtained in the study area was 3.29 t ha^{-1}. The yield map (Figure 2) was generated by ordinary kriging interpolation, based on the best semivariogram, obtained with the Gaussian model (nugget effect in 0215 and reach 397 m).
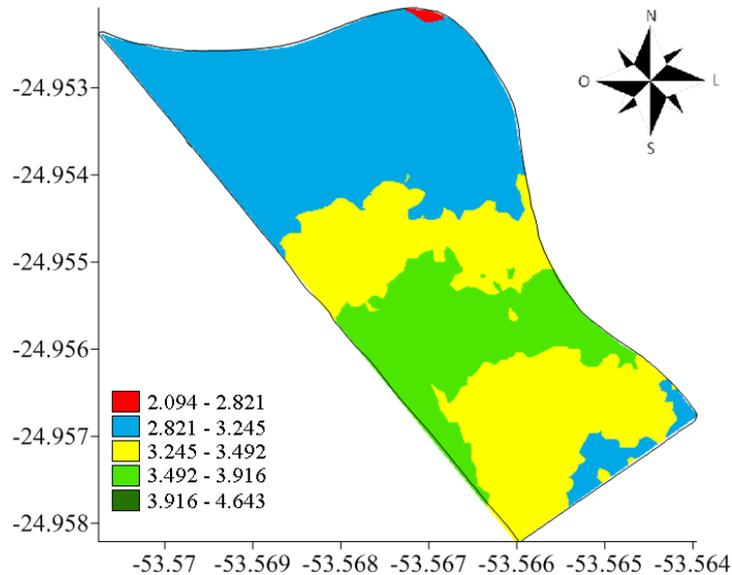

FIGURE 2. Soybean yield map (t ha^{-1}) in a 19.6-ha commercial field, Municipality of Cascavel, Paraná, Brazil.

Based on the estimates of Moran's bivariate spatial autocorrelation statistic (I_{YZ}, Figures 3), variables such as Cu, clay, silt and altitude have significant spatial autocorrelations (diagonal matrix) and significant spatial correlations with soybean yield.

	Yield	P	Ca	C	Cu	Fe	K	Mg	Mn	pH	Zn	Clay	Sand	Silt	SRP_0_10	SRP_0_20	SRP_10_20	Elevation
Yield	0.01073 NS																	
P	0.00111 NS	0.00718 NS																
Ca	-0.01344 NS	0.00559 NS	-0.00180 NS															
C	0.00926 NS	-0.02345 NS	0.02197 NS	0.05272 **														
Cu	0.09284 **	-0.04511 **	0.03882 *	0.13867 **	0.32329 **													
Fe	-0.02350 NS	0.01471 NS	0.01403 NS	-0.01779 NS	-0.11762 **	0.00034 NS												
K	-0.0477 **	-0.02259 NS	-0.01080 NS	0.01079 NS	0.00704 NS	0.02019 NS	0.01134 NS											
Mg	0.00716 NS	0.04114 **	-0.00307 NS	0.01637 NS	-0.00344 NS	0.02835 NS	0.00747 NS	0.02235 NS										
Mn	-0.00982 NS	-0.02602 NS	0.00902 NS	0.06041 **	0.13363 **	-0.02841 *	0.02327 NS	0.02086 NS	0.04074 *									
pH	0.01258 NS	0.01632 NS	-0.0076 NS	0.01628 NS	0.02835 NS	0.01909 NS	-0.00698 NS	0.00471 NS	0.00869 NS	-0.0138 NS								
Zn	-0.02200 NS	0.01076 NS	0.00147 NS	0.00621 NS	-0.01190 NS	0.01919 NS	0.00061 NS	0.02871 *	-0.0093 NS	0.00926 NS	-0.0114 NS							
Clay	0.04330 **	0.06212 NS	-0.03984 **	0.11502 **	0.26229 **	0.10471 **	0.01598 NS	0.03840 **	-0.1041 **	-0.0155 NS	0.03276 *	0.28478 **						
Sand	0.02561 NS	-0.06567 **	0.03305 *	0.10584 **	0.21812 **	-0.09276 **	0.02958 *	-0.03500 *	0.08969 **	0.01052 NS	-0.02583 NS	-0.25138 **	0.22187 **					
Silt	0.05521 **	-0.03216 *	-0.03537 *	0.08518 **	0.23203 **	-0.08338 **	-0.01232 NS	-0.02899 NS	0.08694 **	0.01760 NS	-0.03127 *	-0.22827 **	0.19785 **	0.18888 **				
SRP_0_10	0.01059 NS	0.03990 **	-0.00808 NS	-0.01250 NS	-0.04246 *	0.02687 NS	-0.01127 NS	0.03252 *	-0.00763 NS	0.00350 NS	0.02135 NS	0.09391 **	-0.08729 **	-0.06812 **	0.01062 NS			
SRP_0_20	-0.00492 NS	0.04279 **	-0.00608 NS	0.02671 NS	-0.09166 **	0.07407 *	-0.00759 NS	0.03581 *	-0.01860 NS	0.00897 NS	0.01972 NS	0.12771 **	-0.11280 **	0.10227 **	0.02115 NS	0.02860 NS		
SRP_10_20	-0.01958 NS	0.03664 **	-0.00293 NS	-0.03557 *	-0.12204 **	0.03775 **	-0.00243 NS	0.03123 *	-0.02593 NS	0.01208 NS	0.01453 NS	0.13457 **	-0.11441 **	0.11497 **	0.02664 NS	0.02953 *	0.02685 NS	
Elevation	0.05351 **	0.05718 **	-0.00492 NS	0.01013 NS	0.11911 **	-0.06030 NS	0.03314 *	0.07079 **	0.00071 NS	0.01741 NS	0.05345 **	0.16686 **	0.14225 **	0.14194 **	-0.07764 **	-0.09709 **	-0.09562 **	0.18560 **

* p-value significant at 5% probability, ** p-value significant at 1% probability, NS = not significant (p-value > 0.05)

FIGURE 3. Bivariate spatial autocorrelation statistic of the analyzed chemical variables and yield in the studied field.

For a soil with the same characteristics as the soil studied here, BAZZI et al. (2013) also observed autocorrelation among these attributes (Cu, clay and silt), but only clay and Cu were correlated spatially with soybean yield.

Based on the variables that exhibited both spatial autocorrelation spatial correlation with yield (copper, silt, altitude and clay, in order of decreasing importance), fifteen MZ designs (all possible combinations) were produced (Table 2). This table shows the calculated Moran's statistic. Because the values are not normalized, even small values of this statistic can be statistically significant (non-zero). The important outcome is whether the statistic is significant at 0.01 or 0.05.

TABLE 2. Combination of variables to generate management zones.

Design	Acronym for design	Variables of design
1	01-CuSiAlCl	Copper, Silt, Altitude and Clay
2	02-CuSiAl	Copper, Silt and Altitude
3	03-CuSi	Copper and Silt
4	04-Cu	Copper
5	05-CuSiCl	Copper, Silt and Clay
6	06-CuAlCl	Copper, Altitude and Clay
7	07-SiAlCl	Silt, Altitude and Clay
8	08-CuAl	Copper and Altitude
9	09-CuCl	Copper and Clay
10	10-SiAl	Silt and Altitude
11	11-SiCl	Silt and Clay
12	12-ClAl	Clay and Altitude
13	13-Si	Silt
14	14-Al	Altitude
15	15-Cl	Clay

Cu: copper; Si: silt; Al: altitude; Cl: clay.

The generated maps represent the division of the field in two, three, four and five classes (Figures 4-7) after the design of the division into MZs with FuzME. Thus, the maps present the MZs based on the previous interpolated variables, and thus no new interpolation was necessary to generate the maps. Good correlation can be seen between the yield map (Figure 2) and the MZs.
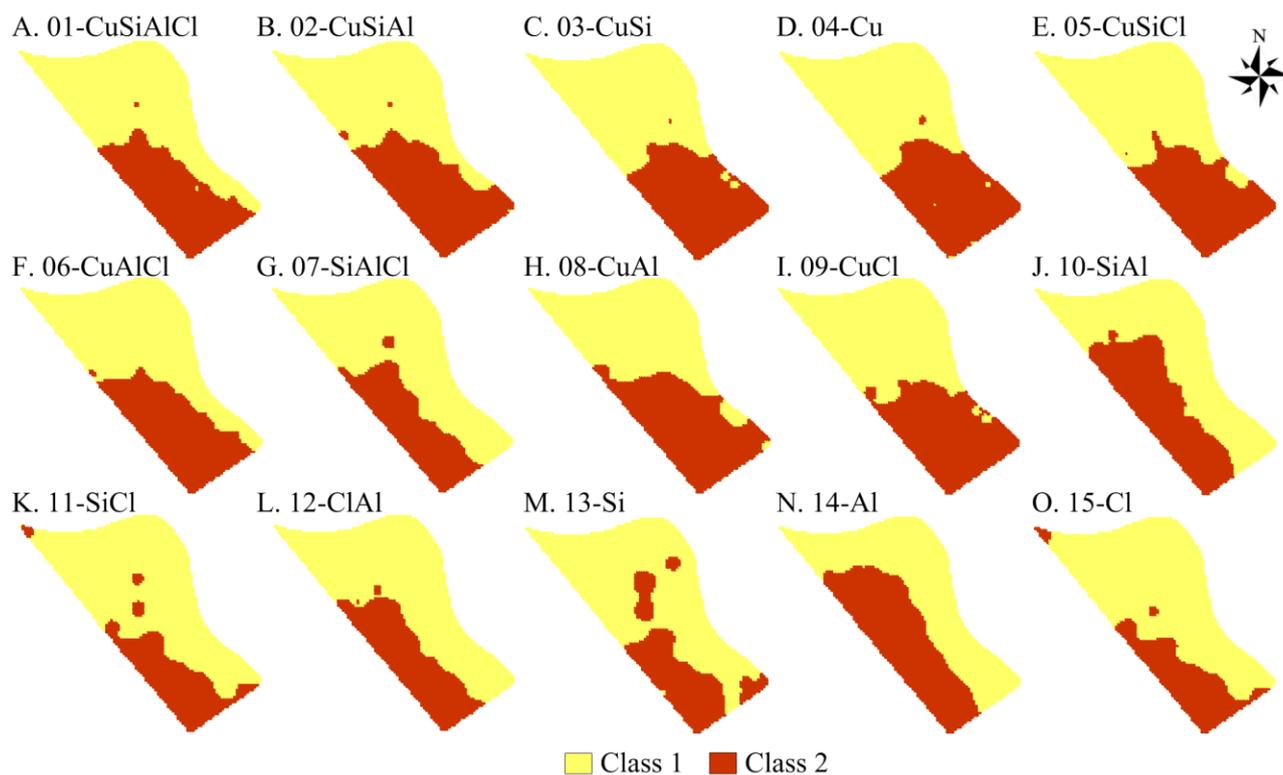

FIGURE 4. Designs of MZs divided into two classes.
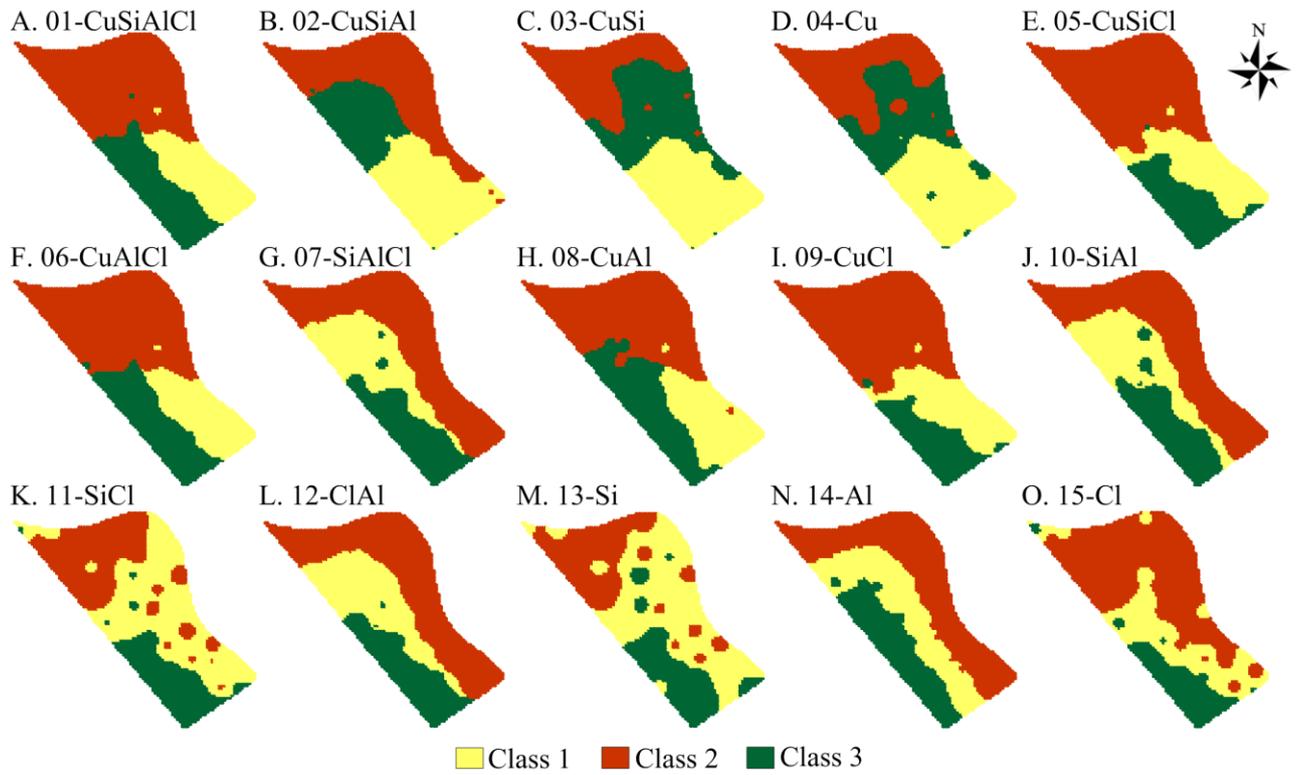
FIGURE 5. Designs of MZs divided into three classes.
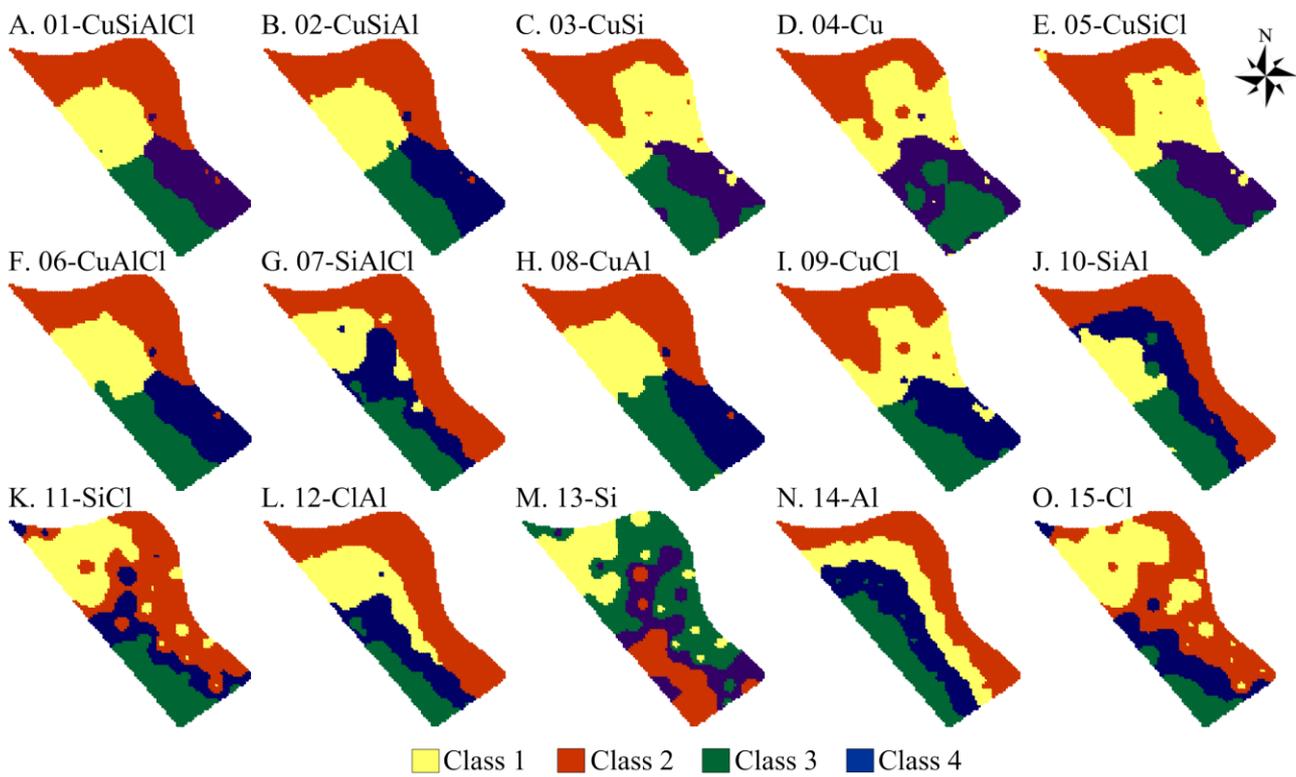
FIGURE 6. Designs of MZs divided into four classes.
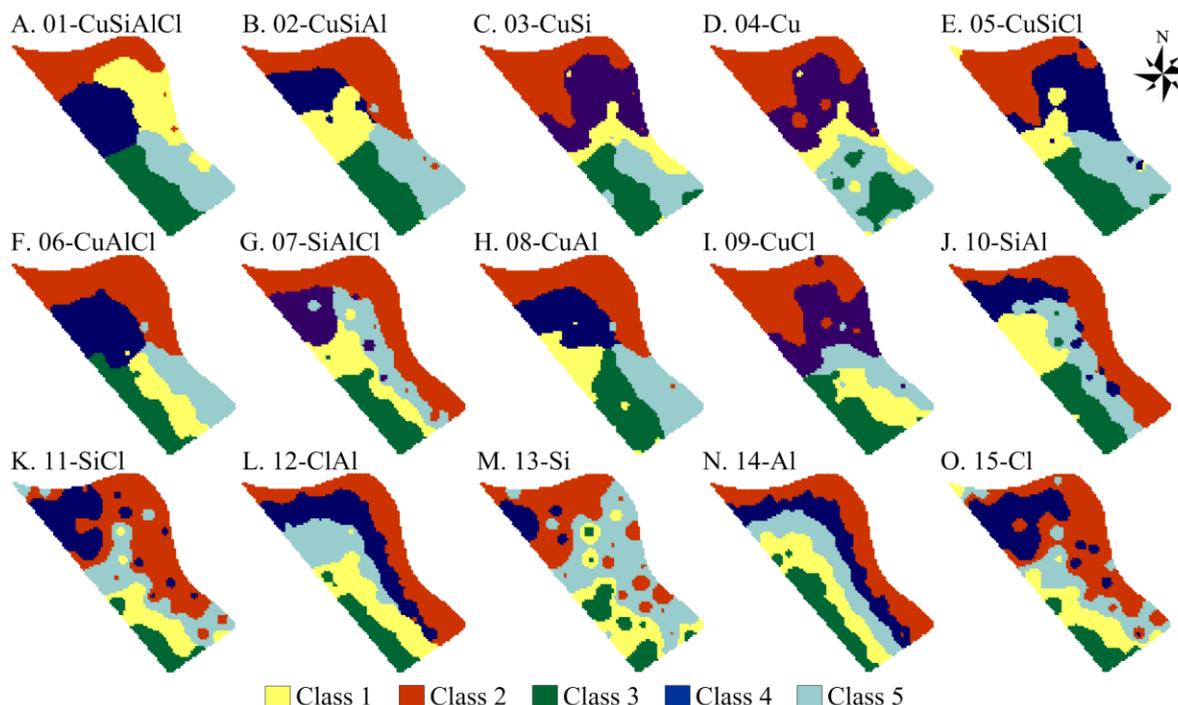
FIGURE 7. Designs of MZs divided into five classes.

Each MZ class (Figures 4-7) represents a group of data based on the variables used in each design. Because the studied variables exhibited a significant spatial correlation with soybean yield, the formed sub-regions (classes) tend to represent different potentials of soybean yield within the total field. The use of different combinations of the selected variables to generate MZs influences how the field is divided.

The soybean yield data were analyzed using Tukey's test to confirm that the MZs classes had different production potentials. The MZs were divided into two classes with significantly different averages for some designs (Table 3). The MZs that belonged to different classes were 01-CuSiAlCl, 02-CuSiAl, 03-CuSi, 04-Cu, 05-CuSiCl, 06-CuAlCl, 07-SiAlCl, 08-CuAl, 09-CuCl and 12-ClAl. ARNO et al. (2011) also determined by ANOVA that differences in yield were evident only when the field was subdivided into two MZs.

TABLE 3. Soybean yield averages in the MZ classes and comparison using Tukey's test at the 5% level of significance.

Division	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D14	D15
2 classes	3.18 a	3.14 a	3.15 a	3.14 a	3.17 a	3.16 a	3.21 a	3.14 a	3.15 a	3.22 a	3.25 a	3.21 a	3.26 a	3.21 a	3.25 a
	3.44 b	3.45 b	3.45 b	3.46 b	3.45 b	3.46 b	3.45 b	3.45 b	3.45 b	3.39 a	3.37 a	3.43 b	3.35 a	3.38 a	3.37 a
	3.10 a	3.10 a	3.03 a	3.04 a	3.12 a	3.10 a	3.19 a	3.10 a	3.13 a	3.19 a	3.12 a	3.20 a	3.18 a	3.18 a	3.15 a
3 classes	3.45 b	3.33 ab	3.34 b	3.34 b	3.41 b	3.39 b	3.30 a	3.34 ab	3.42 b	3.26 ab	3.34 ab	3.28 a	3.31 a	3.28 ab	3.45 b
	3.46 b	3.46 b	3.45 b	3.45 b	3.52 b	3.51 b	3.45 a	3.50 b	3.49 b	3.51 b	3.42 b	3.46 a	3.35 a	3.48 b	3.47 b
	3.06 a	3.07 a	3.04 a	3.04 a	3.03 a	3.07 a	3.20 a	3.07 a	3.01 a	3.12 a	3.09 a	3.20 a	3.17 a	3.14 a	3.09 a
	3.34 ab	3.29 ab	3.29 ab	3.26 a	3.30 ab	3.31 ab	3.21 a	3.28 ab	3.36 b	3.27 ab	3.17 ab	3.25 a	3.29 a	3.27 a	3.23 ab
4 classes	3.39 b	3.42 b	3.35 ab	3.33 ab	3.41 b	3.35 ab	3.30 a	3.36 ab	3.43 b	3.37 ab	3.33 ab	3.39 a	3.33 a	3.34 a	3.39 ab
	3.48 b	3.48 b	3.55 b	3.58 b	3.49 b	3.51 b	3.52 a	3.52 b	3.46 b	3.50 b	3.50 b	3.51 a	3.37 a	3.47 a	3.53 b
	3.00 a	3.09 a	3.01 a	3.05 a	3.02 a	3.04 a	3.14 a	3.07 a	3.02 a	3.17 a	3.07 a	3.14 a	2.92 a	3.15 a	3.08 a
	3.19 ab	3.16 a	3.23 ab	3.21 ab	3.24 ab	3.28 ab	3.19 a	3.27 a	3.26 ab	3.18 a	3.23 ab	3.22 a	3.20 ab	3.21 a	3.18 ab
5 classes	3.35 ab	3.40 a	3.39 ab	3.22 ab	3.30 ab	3.31 ab	3.34 a	3.41 a	3.27 ab	3.31 a	3.27 ab	3.27 a	3.27 ab	3.30 a	3.25 ab
	3.43 b	3.42 a	3.48 b	3.50 b	3.41 ab	3.45 b	3.40 a	3.41 a	3.52 b	3.32 a	3.33 ab	3.39 a	3.38 ab	3.39 a	3.53 b
	3.48 b	3.44 a	3.48 b	3.52 b	3.49 b	3.55 b	3.51 a	3.44 a	3.56 b	3.52 a	3.58 b	3.53 a	3.42 b	3.50 a	3.55 b

The designs refer to Table 2. D01: 01-CuSiAlCl; D02: 02-CuSiAl; D03: 03-CuSi; D04: 04-Cu; D05: 05-CuSiCl; D06: 06-CuAlCl; D07: 07-SiAlCl; D08: 08-CuAl; D09: 09-CuCl; D10: 10-SiAl; D11: 11-SiCl; D12: 12-ClAl; D13: 13-Si; D14: 14-Al; D15: 15-Cl.

As shown in Figure 8, none of the MZs exhibited RE values less than one. The following MZs had relative efficiencies of 1: 11-SiCl, which contains two classes; 13-Si, which contains 2, 3 and 4 classes; and 15-Cl, which contains two classes. Thus, these forms of division into MZs will

not reduce the variance of the yield, in agreement with Tukey's test (Table 3), which revealed no significant differences among the yields. The other MZs with RE > 1 exhibited a variance reduction, suggesting that these divisions can be used as a source of recommendation and analysis.
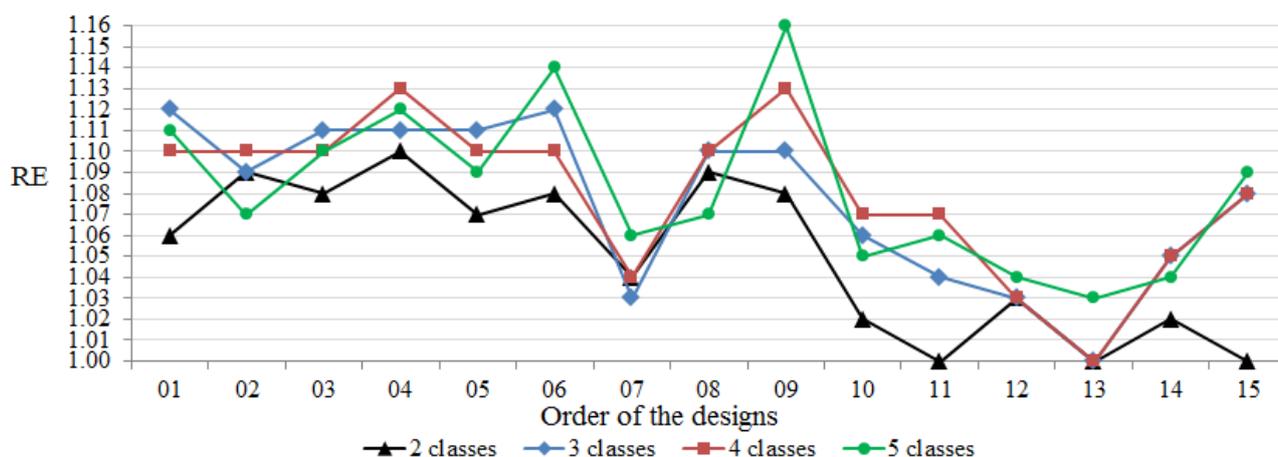

FIGURE 8. Performance of divisions in the management zone according to Relative Efficiency (RE). The order of the designs is 01-CuSiAlCl; 02-CuSiAl; 03-CuSi; 04-Cu; 05-CuSiCl; 06-CuAlCl; 07-SiAlCl; 08-CuAl; 09-CuCl; 10-SiAl; 11-SiCl; 12-ClAl; 13-Si; 14-Al; and 15-Cl.

The division of the 09-CuCl MZs (Figure 9) into 5 classes resulted in the highest RE (1.16) among all the MZs. For division into four MZs, 04-Cu and 09-CuCl designs reduced the yield variance most effectively (RE = 1.13). Combinations 01-CuSiAlCl and 06-CuAlCl exhibited higher RE (1.12) compared to the other divisions in three classes. By contrast, for divisions into two classes, the 04-Cu design had the highest RE (1.10). In general, the RE values were highest in divisions with higher numbers of classes, although for 3, 4 and 5 classes, not all means differed from each other within each design (Table 3).

According to the calculated values of FPI and MPE (Figure 9), the designs in which fewer variables were used exhibited better separation and organization of data among the MZ classes, as indicated by the low values for these indices. Thus, designs such as 04-Cu, 13-Si, 14-Al and 15-Cl, in which only one variable was used to divide MZs, exhibited the lowest levels of FPI and MPE. However, only the 04-Cu design had a high RE, whereas the 13-Si, 14-Al and 15-Cl designs had, in general, the lowest RE values.

When the variable Cu was not included in the evaluation, the MZs were less effective (low RE, Figure 9) and did not all exhibit differences in means in each design (Tukey test), such as the following cases: 10-SiAl 11-SiCl, 12-ClAl, 13-Si, 14-Al and 15-Cl. An exception was the 07-SiAlCl design (Table 3).
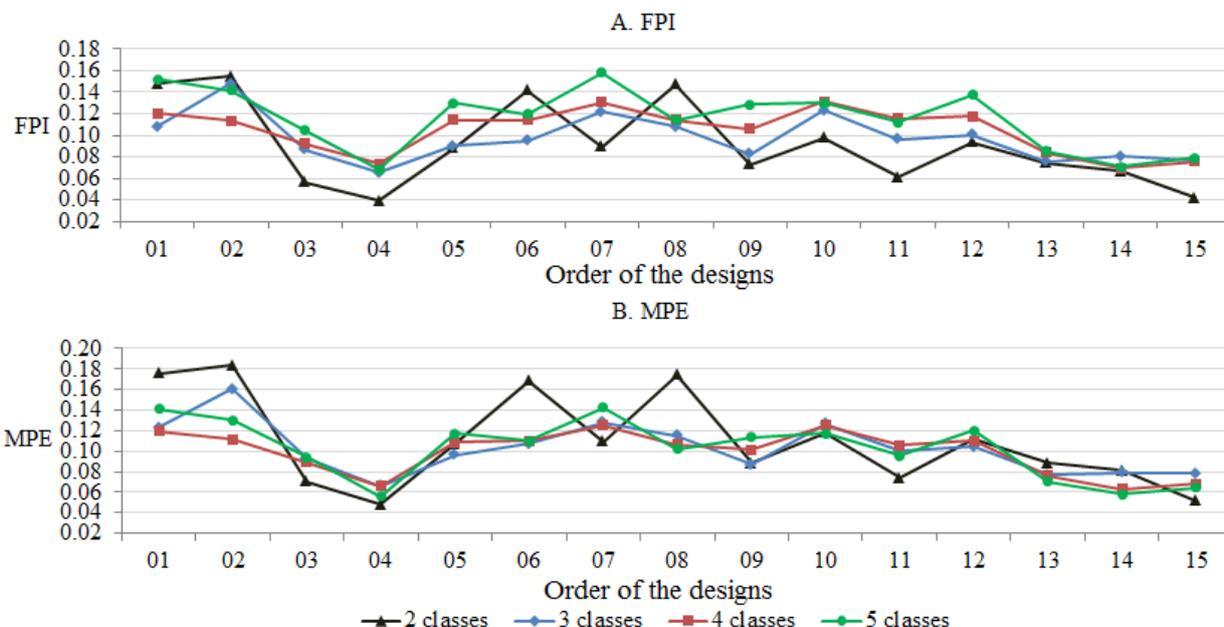
FIGURE 9. Performance of divisions in the management zone according to FPI (Fuzziness Performance Index) and MPE (Modified Partition Entropy). The order of the designs is 01-CuSiAlCl; 02-CuSiAl; 03-CuSi; 04-Cu; 05-CuSiCl; 06-CuAlCl; 07-SiAlCl; 08-CuAl; 09-CuCl; 10-SiAl; 11-SiCl; 12-ClAl; 13-Si; 14-Al; and 15-Cl.

Six designs of MZs were compared using the Kappa index (Table 4). These MZs used the variable Cu and in general had higher RE values in divisions of 2 to 5 classes (01-CuSiAlCl, 02-CuSiAl, 03-CuSi, 04-Cu, 06-CuAlCl and 09-CuCl). The objective was to evaluate the degree of visual agreement among the maps of 2, 3, 4 and 5 classes.

For the MZs that were divided into two classes, there was greater visual agreement among the compared maps. Eight comparisons were in very strong agreement, whereas seven comparisons yielded strong agreement. There were lower levels of agreement among the visual maps for MZs divided into 3, 4 and 5 classes. However, the difference was greatest for MZs divided into 5 classes; two comparisons exhibited strong agreement, seven moderate and six weak.

TABLE 4. Kappa index among maps divided into management zones with 2, 3, 4 and 5 classes for the experimental designs 01-CuSiAlCl, 02-CuSiAl, 03-CuSi, 04-Cu, 06-CuAlCl and 09-CuCl.

Compared experimental designs	2 classes	3 classes	4 classes	5 classes
01-CuSiAlCl x 02-CuSiAl	0.87 (VS)	0.31 (W)	0.96 (VS)	0.59 (MO)
01-CuSiAlCl x 03-CuSi	0.72 (S)	0.30 (W)	0.61 (S)	0.61 (S)
01-CuSiAlCl x 04-Cu	0.69 (S)	0.30 (W)	0.42 (MO)	0.44 (MO)
01-CuSiAlCl x 06-CuAlCl	0.95 (VS)	0.97 (VS)	0.95 (VS)	0.35 (W)
01-CuSiAlCl x 09-CuCl	0.76 (S)	0.85 (VS)	0.64 (S)	0.48 (MO)
02-CuSiAl x 03-CuSi	0.83 (VS)	0.56 (MO)	0.60 (MO)	0.38 (W)
02-CuSiAl x 04-Cu	0.80 (S)	0.56 (MO)	0.40 (W)	0.25 (W)
02-CuSiAl x 06-CuAlCl	0.89 (VS)	0.31 (W)	0.93 (VS)	0.59 (MO)
02-CuSiAl x 09-CuCl	0.86 (VS)	0.33 (W)	0.61 (S)	0.29 (W)
03-CuSi x 04-Cu	0.97 (VS)	0.90 (VS)	0.66 (S)	0.71 (S)
03-CuSi x 06-CuAlCl	0.74 (S)	0.29 (W)	0.57 (MO)	0.38 (W)
03-CuSi x 09-CuCl	0.96 (VS)	0.33 (W)	0.86 (VS)	0.59 (MO)
04-Cu x 06-CuAlCl	0.71 (S)	0.29 (W)	0.38 (W)	0.29 (W)
04-Cu x 09-CuCl	0.93 (VS)	0.34 (W)	0.67 (S)	0.48 (MO)
06-CuAlCl x 09-CuCl	0.78 (S)	0.83 (VS)	0.60 (MO)	0.52 (MO)

VS: Very strong; S: strong; MO: moderate; W: weak.

The 04-Cu experimental design displayed good results for RE, FPI and MPE, very strong agreement with the 03-CuSi and 09-CuCl designs and strong agreement with the 01-CuSiAlCl, 02-CuSiAl and 06-CuAlCl experimental designs for the division of MZs into 2 classes. For MZs divided into three classes, the Cu-04 design exhibited very strong agreement with the 03-CuSi design. For MZs divided into 4 and 5 classes, the 04-Cu design exhibited strong agreement with the 03-CuSi design.

CONCLUSIONS

Some combinations of these variables that were cross-correlated produced better MZs. None of the variable combinations produced a statistically better performance than the MZ generated using only copper, i.e., the design with only no redundant variables. Thus, the other redundant variables can be discredited. The design with all variables did not provide a greater separation and organization of data among MZ classes.

ACKNOWLEDGEMENTS

REFERENCES

ARNO, J.; MARTINEZ-CASASNOVAS, J. A.; RIBES-DASI, M.; ROSSEL, J. R. Clustering of grape yield maps to delineate site-specific. **Spanish Journal of Agricultural Research**, Washington, v. 9, n. 3, p. 721-729, 2011.

BANSOD, B. S.; PANDEY, O. P. An application of PCA and fuzzy C-means to delineate management zones and variability analysis of soil. **Eurasian Soil Science**, Silver Spring, v. 46, n. 5, p. 556-564, 2013.

BASTIANI, F. de; URIBE-OPAZO, M. A.; DALPOSSO, G. H. Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 2, p. 393-404, 2012.

BAZZI, C. L.; SOUSA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; PINHEIRO R. Influência da distância entre passadas de colhedora equipada com monitor de colheita na precisão dos mapas de produtividade na cultura do milho. **Engenharia Agrícola**, Jaboticabal, v. 28, n. 2, p. 355-363, 2008.

BAZZI, C. L. ; SOUZA, E. G. ; Uribe-Opazo, M. A. ; Nóbrega, L. H. P. ; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, Jaboticabal, v. 34, n.5, p. 952-964, 2013.

BEZDEK, C. J. **Pattern recognition with fuzzy objective function algorithms**. New York: Plenum Press, 1981. 256 p.

BONHAM C. D.; REICH R. M.; LEADER K. K. Spatial cross-correlation of Bouteloua gracilis with site factors. **Grassland Science**, Oxford, v. 41, n. 3, p. 196-201, 1995.

BURROUGH, P.A. Fuzzy mathematical methods for soil survey and land evaluation. **Journal of Soil Science**, Oxford, n.40, p.477-492, 1989.

CANNON, R. L.; DAVE, J. V.; BEZDEK, J. C. Efficient implementation of the Fuzzy C-means clustering algorithms. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v.8, n.2, p.248-255, 1986.

CORÁ, J. E.; ARAUJO, A. V.; PEREIRA, G. T.; BERALDO, J. M. G. Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-açúcar. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 28, n. 6, p. 1013-1021, 2014.

DAVATGAR, N.; NEISHABOURI, M. R.; SEPASKHAH, A. R. Delineation of site specific nutrient management zones for a paddy cultivated area based on soil fertility using fuzzy clustering. **Geoderma**, Amsterdam, v. 173-174, n. 1, p. 111-118, 2012.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; MERCANTE, E.; JOHANN, J. A.; BORSSOI, J. A. Comparison measures of maps generated by geostatistical methods. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 1, p. 174-183, 2012.

DOERGE, T. A. **Site-specific management guidelines**. Norcross: Potash & Phosphate Institute, 2000. 135 p.

FU, Q.; WANG, Z.; JIANG, Q. Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO. **Mathematical and Computer Modelling**, Oxford, v. 51, n. 11–12, p. 1299-1305, 2010.

GOOGLE EARTH. **Versão 5.0. Software de imagens**. 2009. Ano da imagem: 6 mar. 2010. Disponível em <http://earth.google.com>. Acesso em: 13 abr. 2011.

GUASTAFERRO, F.; CASTRIGNANO, A.; BENEDETTO, A. de; SOLLITTO, D.; TROCCOLI, A.; CAFARELLI, B. A comparison of different algorithms for the delineation of management zones. **Precision Agriculture**, Dordrecht, v. 11, n. 6, p. 600–620, 2010.

ILIADIS, L.S.; VANGELOUDH, M.; SPARTALIS,S. An inteligent system employing an enhanced fuzzy C-Means clustering model: Application in the case of forest fires. **Computers and Eletronics in Agriculture**, Netherlands, v.70, n.1, p.276-284, 2010.

JOHNSON, D.E. **Applied multivariate methods for data analysis**. Pacific Grove: Brooks/Cole Publ., 1998. 567 p.

KWEON, G. Delineation of site-specific productivity zones using soil properties and topographic attributes with a fuzzy logic system. **Biosystems Engineering**, London, v. 112, n. 4, p. 261-277, 2012.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, Arlington, v. 33, n. 1, p. 159-174, mar. 1977.

LEÃO, M. G. A. de; MARQUES JÚNIOR, J.; SOUZA, Z. M. de; SIQUEIRA, D. S.; PEREIRA, G. O relevo na interpretação da variabilidade espacial dos teores de nutrientes em folha de citros. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 14, n. 11, p. 1152-1159, 2010.

LI, Y.; SHI, Z.; WU, H.; LI, F.; LI, H. Definition of management zones for enhancing cultivated land conservation using combined spatial data. **Environmental Management**, New York, v. 52, p. 792-806, 2013.

MCBRATNEY, A. B.; MOORE, A. W. Application of fuzzy sets to climatic classification. **Agricultural and Forest Meteorology**, Amsterdam, v. 35, n. 1-4, p. 165-185, 1985.

MIELKE, P. W. J; BERRY, K. J. **Permutation methods**: a distance function approach. New York: Springer, 2007. 439 p.

MINASNY, B.; MCBRATNEY, A. B. **FuzME version 3**. Sidney: Australian Centre for Precision Agriculture, The University of Sydney, 2002. Disponível em <http://sydney.edu.au/agriculture/acpa/software/fuzme.shtml>. Acesso em 7 mar. 2011.

MOLIN, J. P.; CASTRO, C. N. de C. Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. **Scientia Agricola**, Piracicaba, v. 65, n. 6, p. 567-573, 2008.

MOLIN, J. P.; MOTOMIYA, A. V. de A.; FRASSON, F. R.; FAULIN, G. D. C.; TOSTA, W. Test procedure for variable rate fertilizer on coffee. **Acta Scientiarum Agronomy**, Maringá, v. 32, n. 4, p. 569-575, 2010.

MORAL, F. J.; TERRÓN, J. M.; MARQUES DA SILVA, J. R. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil and Tillage Research**, Amsterdam, v. 106, n. 2, p. 335-343, 2010.

MORARI, F.; CASTRIGNANÒ, A.; PAGLIARIN, C. Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. **Computers and Electronics in Agriculture**, New York, v. 68, n. 1, p. 97-107, 2009.

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

PIMENTEL GOMES, F. P.; GARCIA, C. H. **Estatística aplicada a experimentos agronômicos e florestais**. Piracicaba: FEALQ, 2002. 305 p.

PING, J. L.; DOBERMANN, A. Creating spatially contiguous yield classes for site-specific management. **Agronomy Journal**, v.95, p.121-113, 2003.

RODRIGUES JUNIOR, F. A.; VIEIRA, L. B.; QUEIROZ, D. M. de; SANTOS, N. T. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 15, n. 8, p. 778-787, 2011.

ROUDIER, P.; TISSEYRE, B.; POILVE, H.; ROGER, J. M. A technical opportunity index adapted to zone-specific management. **Precision Agriculture**, Dordrecht, v. 12, n. 1, p. 130-145, 2011.

TOU, J.; GONZALEZ, R. **Pattern recognition principles**. Reading: Addison-Wesley, 1974. 377 p.

VALENTE, D.S.M.; QUEIROZ, D.M.; PINTO, F.A.C.; SANTOS, N.T.; SANTOS, F.L. Definition of management zones in coffee production fields based on apparent soil. **Scientia Agricola**, Piracicaba, v. 69, n. 3, p. 173-179, 2012.

VITHARANA, U. W. A.; MEIRVENNE, M. V.; SIMPSON, D.; COCKX, L.; BAERDEMAEKER, J. de. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. **Geoderma**, Amsterdam, v. 143, n. 1-2, p. 206-215, 2008.

WOLLENHAUPT, N. C.; WOLKONSKI, R. P.; CLAYTON M. K. Mapping soil test phosphorus and potassium for variable-rate fertilizer applications. **Journal of Production Agriculture**, Madison, v. 7, n. 4, p. 441-448, 1994.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHUN-JIANG, Z. Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences In China**, Maryland, v. 6, n. 2, p. 180-188, fev. 2007.

XIN-ZHONG, W.; GUO-SHUN, L.; HONG-CHAO, H.; ZHEN-HAI, W.; QING-HUA, L.; XU-FENG, L.; WEI-HONG, H.; YAN-TAO, L. Determination of management zones for a tobacco field based on soil fertility. **Computers and Electronics in Agriculture**, New York, v. 65, n. 2, p. 168-175, 2009.

ZUCOLOTO, M.; LIMA, J. S. S; COELHO, R. I.; QUARTEZANI, W. Z. Variabilidade espacial das frações granulométricas e da produção de bananeira 'Prata Anã'. **Idesia**, Arica, v. 29, n. 2, p. 47-52, 2011.