



## Data interpolation in the definition of management zones

Kelyn Schenatto<sup>1\*</sup>, Eduardo Godoy Souza<sup>2</sup>, Claudio Leones Bazzi<sup>3</sup>, Vanderlei Arthur Bier<sup>2</sup>, Nelson Miguel Betzek<sup>3</sup> and Alan Gavioli<sup>3</sup>

<sup>1</sup>Universidade Tecnológica Federal do Paraná, Prolongamento da Rua Cerejeira, s/n., 85892-000, Santa Helena, Paraná, Brazil. <sup>2</sup>Universidade Estadual do Oeste do Paraná, Cascavel, Paraná, Brazil. <sup>3</sup>Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brazil. \*Author for correspondence. E-mail: [kschenatto@utfpr.edu.br](mailto:kschenatto@utfpr.edu.br)

**ABSTRACT.** Precision agriculture (PA) comprises the use of management zones (MZs). Sample data are usually interpolated to define MZs. Current research checks whether there is a need for data interpolation by evaluating the quality of MZs by five indices – variance reduction (VR), fuzzy performance index (FPI), modified partition entropy index (MPE), Kappa index and the cluster validation index (CVI), of which the latter has been focused in current assay. Soil texture, soil resistance to penetration, elevation and slope in an experimental area of 15.5 ha were employed as attributes to the generation of MZ, correlating them with data of soybean yield from 2011-2012 and 2012-2013 harvests. Data interpolation prior to MZs generation is important to achieve MZs as a smoother contour and for a greater reduction in data variance. The Kriging interpolator had the best performance. CVI index proved to be efficient in choosing MZs, with a less subjective decision on the best interpolator or number of MZs.

**Keywords:** inverse of distance, inverse of square of distance, kriging.

## Interpolação de dados na definição de unidades de manejo

**RESUMO.** Uma forma de adoção da agricultura de precisão (AP) é a utilização de unidades de manejo (UMs), sendo que, para se definir UMs, normalmente os dados amostrais são interpolados. O presente trabalho objetivou verificar se há necessidade de interpolação dos dados, avaliando-se a qualidade das UMs por meio de cinco índices: redução da variância – RV, índice de desempenho fuzzy – FPI, índice da partição da entropia modificada – MPE, índice Kappa e o índice de validação de clusters – CVI, proposto neste trabalho. Como atributos candidatos à geração de UM, foram utilizadas a textura do solo, a resistência do solo à penetração, a elevação e a declividade, em uma área experimental de 15,5 ha, correlacionando-as com dados de produtividade da soja coletadas nas safras 2011/2012 e 2012/2013. Pode-se concluir que é importante a interpolação dos dados antes da geração das UMs para se obter UMs com contorno mais suave e maior redução da variância dos dados, tendo obtido melhor desempenho o interpolador krigagem. O índice CVI mostrou-se eficiente na escolha das UMs, possibilitando uma decisão menos subjetiva do melhor interpolador ou número de UMs.

**Palavras-chave:** inverso da distância, inverso do quadrado da distância, krigagem.

## Introduction

The modernization of agriculture through the technology of precision agriculture (PA) has resulted in the emergence of machines equipped with sensors and technological equipments that make agricultural activity increasingly competitive. Despite the advantages of PA, the cost of obtaining soil samples to characterize field variability restricts this technology to big producers (Yan, Zhou, Feng, & Hong-Yi, 2007).

So that the relationship between economic aspects and the benefits obtained with PA could be adjusted, modifications on PA concepts are conducted to identify regions within the field with similar characteristics. The regions are called management zones (MZs).

The definition of MZ makes PA techniques easier to apply since the same systems used in conventional agriculture may be employed in crop management. According to Yan et al., (2007), information on MZ reduces the number of soil analyses necessary to create application maps in farming operations and makes PA technology more attractive.

Working on soil attributes, several researchers, such as Peralta, Costa, Franco, and Balzarini (2013), Mzuku et al. (2005), Vrindts et al. (2005) and Delalibera, Weirich Neto, and Nagata (2012), have detected positive results working with MZ definition with different crops and with different parameters. Similarly, Suszek, Souza, Uribe-Opazo, and Nobrega (2011) and Bazzi, Souza, Uribe-

Opazo, Nobrega, and Rocha (2013) in their assays with soybean; Moral, Terrón, and Marques da Silva (2010) with apparent soil electrical conductivity; and Tagarakis, Liakos, Fountas, Koundouras, and Gemtos (2013), in the cultivation of grapes and soil attributes.

Among the techniques in MZ generation, one may mention clustering algorithms, such as K-Means and Fuzzy C-Means (Iliadis, Vangeloudh, & Spartalis, 2010; Arno Martinez-Casasnovas, Ribes-Dasi, & Rosell, 2011; Valente, Queiroz, Pinto, Santos, & Santos, 2012; Li, Shi, Wu, Li, & Li, 2013), which provide good results (Mingoti & Lima, 2006; Jipkate & Gohokar, 2012).

In the case of the software used to run these algorithms, the highlights are FuzMe (Minasny & Mcbratney, 2002), used by researchers Molin and Castro (2008); the Management Zone Analyst - MENZA (Fridgen et al., 2004); and Software for Definition of Management Zones - SDMZ, developed and used by Bazzi et al. (2013). The clustering algorithms have as input only data rates of the attributes to be clustered, which need datasets with the same number of sample elements. There is often the need to perform data interpolation. Coupled to this demand, data interpolation also becomes necessary when there is a small number of sample points and it is necessary to generate MZs, obtaining as a result maps with precise definitions of sub-regions in the area (Bazzi et al., 2013). If data interpolation must be used, the most used interpolators are inverse distance (ID), inverse square distance (ISD) or kriging (KRI). Kriging is a robust interpolator and has usually provided better results with regard to ID and ISD, though it requires in-depth knowledge in geo-statistics (Mazzini & Schettini, 2009; Guastaferro et al., 2010).

In this context, current assay evaluates whether data interpolation used in the generation of MZs is justified. The quality of MZs was also evaluated by five indices, of which one is the cluster validation focused upon in current analysis.

## Material and methods

Sampling points were determined and an experimental area of approximately 15.5 ha in Céu Azul, Paraná State, Brazil, was delineated by a topographic GPS Trimble Geo Explorer XT 2005 and software PathFinder. The central geographical location lies at 25°06'32"S and 53°49'55"W, at an average altitude of 752 m. The area has been cultivated under a tilling system for

more than 10 years, with a crop sequence of soybean, wheat, oats and corn, for commercial purposes.

Stable attributes were used in the definition of MZ, with the exclusion of chemical soil attributes and thus satisfying the general recommendations by the literature (Doerge, 2000). A sampling grid was used, with irregular sampling points, located on an imaginary line among the intermediate contour lines. Forty sampling points (2.58 points ha<sup>-1</sup>) were defined in which data on altitude, slope, texture (clay, silt and sand), soil penetration resistance (SPR), density, macro- and micro-porosity and total porosity were collected.

Soil sampling was carried out with an auger, at a depth of 0-0.2 m; eight sub-samples were collected for each sampling point within a radius of 3 m from the point determined by the sampling grid (adapted from Wollenhaupt, Wolkonski, & Clayton, 1994).

SPR was determined by using a Falker PenetroLOG electronic penetration meter, with which four measurements around the point defined in the sampling grid (a maximum distance of 3 m) were taken and, subsequently, the average measurements for the representation of the point, at depths 0-0.1, 0.1-0.2 and 0.3-0.2 m. Non-deformed soil samples were collected by a volumetric ring to determine water content, soil density, macro-porosity, micro-porosity and total porosity, following methodology by Embrapa (1997). Altitude was obtained by electronic total station Topcon GPT-7505.

Soybean yield during agricultural years 2011-2012 and 2012-2013 was measured by harvest monitor AFS PRO 600 locked to combined CASE IH 2388 and filtered data, suggested by Michelin, Souza, and Uribe-Opazo (2007).

Moran's bivariate spatial autocorrelation statistics (Czaplewski & Reich, 1993) were used to assess the spatial correlation between the analyzed attributes and to establish the spatial correlation matrix, which checks which attributes influence the yield positively or negatively, and whether a sample is correlated spatially (spatial autocorrelation). The attributes used in the generation of MZ were selected by the variable selection method proposed by Bazzi et al. (2013).

For a greater detailing of the attributes under analysis, data were interpolated to a 5 x 5 m grid by using the following methods: nearest neighbor (NN), inverse distance (ID), inverse square distance (ISD) and ordinary kriging (KRI). Whereas ID, ISD and KRI have been chosen because they are the most used methods in PA (Robinson & Metternicht, 2006; Ortega & Santibáñez, 2007; Lopez-Lozano, Casterad,

& Herreno, 2010), the NN method is a deterministic interpolation method in which the estimated value is always equal to its nearest sample without considering any other. In fact, NN is exact and preserves the sampled rates. ArcGIS Software was used for map plotting and geostatistical analysis, with the classic estimator of Matheron. Theoretical models (spherical, exponential, Gaussian) have been adjusted to the semivariogram, and the best model was selected, taking into account the cross-validation statistics.

Through interpolated values with regard to selected soil attributes, MZs were generated by Fuzzy C-Means clustering method. During MZ evaluation, the variance reduction - VR (Ping & Dobermann, 2003; Xiang, Yu-Chun, Zhong-Qiang, & Chun-Jiang, 2007) was calculated for soybean yield (Equation 1), expecting that the sum of the variances of MZs data would be lower than the total variance. The average comparison test (ANOVA) among MZs was also undertaken to identify whether they had significant average difference for soybean yield, assuming that there is no spatial dependence inside each MZ.

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

where,  $c$  is the number of management zones;  $W_i$  is the proportion of area in each management zone;  $V_{ZM_i}$  is the variance of data in each management zone;  $V_{Field}$  is the sample variance of data to field.

Further, the fuzziness performance index (FPI; Equation 2) was employed for the evaluation of MZs, which determined the separation degree (or confusion) among the fuzzy  $c$ -clusters of a set of data  $X$  and the modified partition entropy index (MPE; Equation 3) which estimated the disorganization established by a specified number of clusters.

$$FPI = 1 - \frac{c}{(c-1)} \left[ 1 - \sum_{j=1}^n \sum_{i=1}^c (u_{ij})^2 / n \right] \quad (2)$$

where,  $c$  = number of clusters;  $n$  = number of observations;  $u_{ij}$  = is the element  $ij$  of the fuzzy membership matrix  $U$ .

$$MPE = \frac{- \sum_{j=1}^n \sum_{i=1}^c u_{ij} \log(u_{ij}) / n}{\log c} \quad (3)$$

where,  $c$  = number of clusters;  $n$  = number of observations;  $u_{ij}$  = is the element  $ij$  of the fuzzy membership matrix  $U$ .

In the case of FPI, the rates close to 0 indicate distinct classes, with a small sharing degree of members (data), whereas rates close to 1 indicate no distinct classes, with high degree of sharing of members between the classes (Fridgen et al., 2004). In the case of MPE, rates close to 1 indicate the predomination of disorganization, whereas rates close to 0 show a better organization (Boydell & Mcbratney, 2002). The best grouping (cluster) number of a set of data set is based on the minimum rate of FPI and MPE. So that the situation in which estimates point to different models would be avoided, the cluster validation index (CVI, Equation 4) has been proposed in current assay which, in the selection of  $j$  clusters, provides a rate closer to zero as FPI and MPE. Thus, in choosing between groupings, the best clustering is that with the lowest CVI.

$$CVI = \frac{FPI_i}{\text{Max}_{i=\{1,...,j\}} FPI_i} + \frac{MPE_i}{\text{Max}_{i=\{1,...,j\}} MPE_i} \quad (4)$$

The MZs generated by the original data without interpolation and the ones generated by ID, ISD and kriging were also compared by Kappa (K) and Tau indexes (Dalposso, Uribe-Opazo, Mercante, Johann, & Borssoi, 2012). The degree of agreement was classified according to Landis and Koch (1977) for K, namely,  $0 < K \leq 0.2$ : no agreement;  $0.2 < K \leq 0.4$ : weak;  $0.4 < K \leq 0.6$ : moderate;  $0.6 < K \leq 0.8$ : strong;  $0.8 < K \leq 1$ : very strong.

## Results and discussion

The descriptive analysis of data (Table 1) revealed that the highest coefficients of variation (CV) occurred with the 2012 (39%) and 2013 (20%) yields and slope (147%). Although soil compaction rates increased between 2011 and 2012, rates for SPR were between low and some limitations, following Canarache (1990). Total porosity averaged  $0.45 \text{ cm}^3$ , of which  $0.11 \text{ cm}^3$  corresponded to macro-porosity and  $0.34 \text{ cm}^3$  to micro-porosity. Bulk density rates ranged between  $1.24$  and  $1.47 \text{ g cm}^{-3}$ , with mean  $1.35 \text{ g cm}^{-3}$ . Cavalcante, Alves, Souza, and Perreira (2011) reported very similar results.

The attributes for MZs definition were selected as from the spatial correlation matrix (Moran index) and by eliminating the redundant variables (Bazzi et al., 2013). This resulted in the attribute altitude for the 2012-2013 harvest, and in altitude

and SPR 0-0.1 m for the 2012-2013 harvest, for which the geo-statistical analyses were prepared (Table 2).

Two, three and four MZs were generated by the Fuzzy C-Means clustering method, employing data interpolated by ID, ISD and KRI methods for a 5 x 5 m grid. In the case of 2011-2012 soybean harvest, MZs with the attribute altitude were generated

(Figure 1). MZs with attributes altitude and RSP 0-0.10 m were generated within the same area for the 2012-2013 soybean harvest (Figure 2). The importance of interpolating the sampling data (Figures 1a and 2a) in MZs generation may be understood by interpolation viewing with NN (Figures 1b and 2b), which provides more deformed MZs than the other selected interpolation methods.

**Table 1.** Descriptive statistics of the soil's physical attributes in 2011 and 2012.

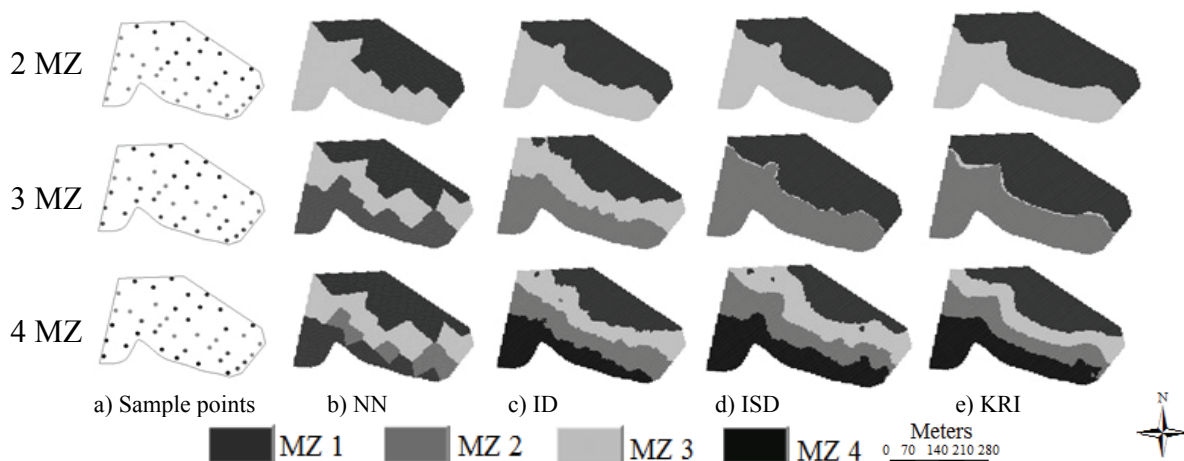
Physical attributes of the soil – 2011								
Attribute/unit	Minimum	Mean	Median	Maximum	SD	CV(%)	Asymmetry	Kurtosis
SPR 0-0.1 (MPa)	1.01	1.58	1.52	2.79	0.43	27	1.36	1.86
SPR 0.1-0.2 (MPa)	1.42	2.00	1.98	2.69	0.31	15	0.38	-0.58
SPR 0.2-0.3 (MPa)	1.49	1.91	1.93	2.31	0.19	10	-0.09	2.51
Physical attributes of the soil – 2012								
Total porosity (cm <sup>3</sup> )	0.42	0.45	0.44	0.49	0.02	4	0.32	-0.42
Macro-porosity (cm <sup>3</sup> )	0.06	0.11	0.10	0.16	0.02	21	0.49	-0.03
Micro-porosity (cm <sup>3</sup> )	0.30	0.34	0.34	0.37	0.02	4	-0.28	-0.67
Density (g cm <sup>-3</sup> )	1.24	1.35	1.34	1.47	0.06	4	0.05	-1.03
Humidity (%)	16.6	23.3	24.2	29.3	2.83	12	-0.33	-0.30
SPR 0-0.1 (MPa)	1.59	2.24	2.21	3.32	0.39	17	0.64	0.70
SPR 0.1-0.2 (MPa)	1.69	2.27	2.25	2.72	0.23	9	-0.14	0.17
SPR 0.2-0.3 (MPa)	1.34	1.92	1.90	2.29	0.23	11	-0.21	-0.28
Soil texture – 2011								
Clay (%)	60.0	67.9	68.0	74.0	3.56	5	-0.30	-0.48
Silt (%)	14.0	20.6	20.0	26.0	2.65	12	-0.13	0.30
Sand (%)	6.0	11.5	11.0	17.0	2.88	25	0.19	-0.84
Topography – 2012								
Altitude (m)	651.3	662.7	662.1	676.4	6.86	1	0.19	-1.02
Slope (°) *	0.00	2.80	1.28	21.65	4.10	147	3.03	11.20
Soybean yield								
Yield 2011/2012*	0.51	2.44	2.39	4.40	0.96	39	0.00	-1.43
Yield 2012/2013*	1.30	3.93	4.00	11.73	0.80	20	2.18	20.14

SD – Standard Deviation; CV (%) – Coefficient of Variation; \* Not Normal at 5% probability.

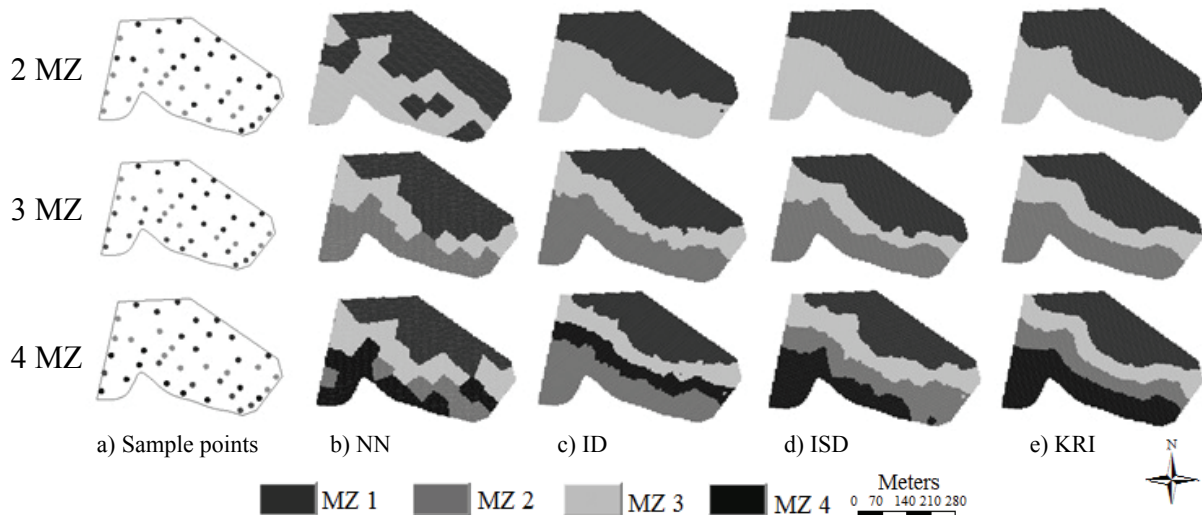
**Table 2.** Semivariogram parameters for the choice of the best model of selected attributes to generate MZs.

Attribute	Model	Nugget	Partial Sill	Sill	Range	ICE
Altitude	Spherical	0	60.048	60.048	60.048	1.68
	<u>Exponential</u>	0	<u>69.495</u>	<u>69.495</u>	<u>69.495</u>	<u>1.60</u>
	Gaussian	6.850	55.485	62.335	62.337	1.71
SPR 0-0.1 m	Spherical	9.873	9.577	19.450	1,945	1.81
	<u>Exponential</u>	<u>8.538</u>	<u>1.018</u>	<u>9.556</u>	<u>1.872</u>	<u>1.57</u>
	Gaussian	1.142	9.497	10.639	2,092	1.95

C<sub>0</sub> – Nugget Effect; C<sub>1</sub> – Contribution; IDE% – Space Dependency Index; VR – reducing measurement error; SER – standard deviation of reducing measurement error; ICE – Error comparison index; Underlined items indicate the best model.



**Figure 1.** Management Zones (MZs) generated for the 2011/2012 harvest with two, three and four MZs with Fuzzy C-Means, using sampling points (a), and the interpolators: NN – nearest neighbor (b); ID – inverse of distance (c); ISD – inverse square of distance (d); KRI – ordinary kriging (e).



**Figure 2.** Management Zones (MZs) generated for the 2012-2013 harvest with two, three and four MZs with Fuzzy C-Means, using sampling points (a) and the interpolators: NN – nearest neighbor (b); ID – inverse distance (c); ISD – inverse square distance (d); KRI – ordinary kriging (e).

In the evaluation of MZs for the two harvests (2011-2012 and 2012-2013), ANOVA registered a significant difference in soybean yield for all zones (2, 3 and 4), regardless of the interpolation method used (Tables 3 and 4). However, relative efficiency (RE) was greater for the first harvest, perhaps due to the fact that, in the first year, the bottom section of the area (areas corresponding to yield, ranging between 0.5 and 4.1 t ha<sup>-1</sup>, Figure 3a,) had to be replanted and thus provided lower yields than the top area. Figures 3 and 4 demonstrate that there was a natural separation of data according to the replanting process. Consequently, VR (which is as high as the reduction of data variability or variance) ranged between 59 and 74% for the 2011-2012 harvest and from zero to 2% for the 2012-2013 one (Figures 5c and 5d).

FPI, MPE and CVI indexes (Figure 5c-h) showed a similar behavior, or rather, the decision for the best interpolation method was the same, regardless of the index used. Results show that interpolation by ID was the worst in the 2011-2012 harvest. Data on the two harvests did not reveal any gain by implementing the CVI index, due to the similar performances of FPI and MPE. However, specialized literature (Li, Shi, Wu, Li, & Li, 2008; Valente et al., 2012) reports instances in which the above failed to occur and that CVI would provide a less subjective decision on the best interpolation or on the number of management areas.

Analyses of CVI and VR for the two harvests demonstrated that kriging provided the best results. The best zones for 2011-2012 and 2012-2013 harvests were respectively two and four. It should be

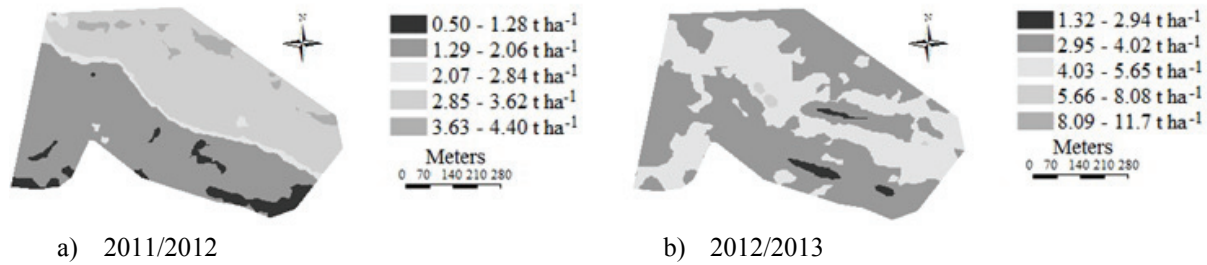
underscored that the methodology used to generate MZ for each harvest separately is not recommended and was only used to show the efficiency of MZ selection indexes. However, the MZs should be ideally generated by stable attributes selected according to average normalized yield during several years.

The influence of the type of interpolation in the area defined by each interpolation method for each MZ (Table 5) ranged between 0 and 30%. Greatest differences were reported between ISD and ID and ID and KRI for the division in 3 MZ. A maximum difference of 3% occurred in the case of the division into 2 MZ. For the division in 4 MZ, the greatest differences were found among MZs generated by ID and ISD (15%).

In an overall evaluation, ANOVA showed that kriging was the method that obtained more significant differences among MZs generated (Tables 3 and 4), whilst ID provided less significant differences among the divisions.

Comparisons ID x ISD, ID x KRI and ISD x KRI (Table 6) for 2, 3 and 4 MZs were employed to evaluate map concordance according to the type of interpolator.

The 2011-2012 soybean harvest demonstrates that the concordance between the maps decreased when there was an increase in MZs. Concordance was classified very strong only for 2 MZs; concordance ranged between average and very strong for division in 3 and 4 MZs. In the 2012-2013 harvest, there was little difference among the maps for the division into two MZs, whereas the maps showed very strong concordance for all comparisons.

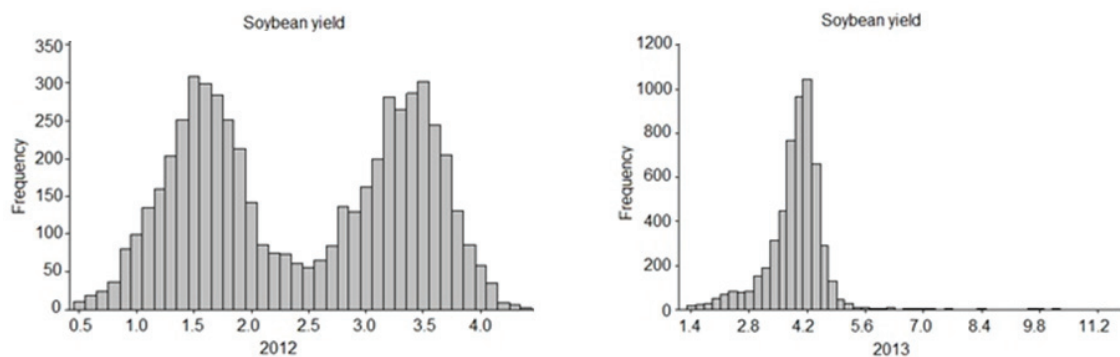


**Figure 3.** Soybean yield map in 2011-2012 (a) and 2012-2013 (b) harvests interpolated by kriging.

**Table 3.** Descriptive statistics on soybean yield data in the 2011-2012 harvest, separated by MZ and cluster validation indices.

Nº	Interp.	MZ	% area	N	Min	Median	Max	D. P.	C. V. (%)	VR%	FPI	MPE	CVI
2	Not interpolated	1	47	2,635	0.53	3.22b	4.40	0.56	31	59.2	0.09	0.11	1.50
		2	53	2,939	0.51	1.74a	4.27	0.65	43				
	ID	1	49	2,783	0.64	3.23a	4.4	0.51	16	69.4	0.17	0.19	1.40
		2	51	2,791	0.51	1.64b	4.1	0.54	33				
	ISD	1	49	2,783	0.64	3.23a	4.4	0.51	16	69.4	0.09	0.10	1.45
		2	51	2,791	0.51	1.64b	4.1	0.54	33				
	KRI	1	52	3,009	0.64	3.18a	4.4	0.57	18	71.2	0.08	0.10	1.40
		2	48	2,565	0.51	1.56b	4.1	0.44	28				
	Not interpolated	1	33	1,789	1.50	3.33a	4.4	0.40	12	59.8	0.09	0.10	1.41
		2	32	1,844	0.53	2.52b	4.33	0.91	36				
		3	35	1,941	0.51	1.54c	3.54	0.39	25				
		4	33	1,789	1.50	3.33a	4.4	0.40	12				
3	ID	1	31	1,676	2.16	3.33a	4.4	0.40	12	63.4	0.18	0.20	1.47
		2	32	1,842	0.54	2.65b	4.3	0.87	33				
		3	37	2,056	0.51	1.51c	2.6	0.35	23				
		4	33	1,789	1.50	3.33a	4.4	0.40	12				
	ISD	1	48	2,744	0.64	3.24a	4.4	0.51	16	70.6	0.09	0.10	1.45
		2	2	82	1.21	2.76b	4.1	0.71	26				
		3	50	2,748	0.51	1.62c	4.1	0.52	32				
		4	23	1,272	0.51	1.56d	2.64	0.36	23				
	KRI	1	50	2,878	0.64	3.22a	4.4	0.52	27	73.6	0.08	0.10	1.40
		2	2	151	0.88	2.25b	3.9	0.86	38				
		3	48	2,545	0.51	1.56c	4.1	0.43	28				
		4	23	1,272	0.51	1.56d	2.64	0.36	23				
4	Not interpolated	1	33	1,789	1.50	3.33a	4.40	0.40	12	63.0	0.08	0.09	1.26
		2	14	846	0.53	2.97b	4.33	0.74	25				
		3	30	1,667	0.52	1.88c	4.26	0.78	42				
		4	23	1,272	0.51	1.56d	2.64	0.36	23				
	ID	1	26	1,361	2.17	3.35a	4.3	0.38	0.11	74.4	0.19	0.20	1.50
		2	25	1,544	0.64	3.09b	4.4	0.62	0.20				
		3	25	1,386	0.53	1.63c	3.9	0.53	0.33				
		4	24	1,283	0.51	1.54d	2.6	0.35	0.23				
	ISD	1	18	964	2.16	3.34a	4.3	0.40	0.12	68.4	0.10	0.10	1.50
		2	27	1,583	0.64	3.24b	4.4	0.48	0.15				
		3	25	1,376	0.53	1.97c	4.2	0.80	0.40				
		4	30	1,651	0.51	1.52d	2.6	0.36	0.24				
	KRI	1	28	1,555	2.16	3.36a	4.4	0.39	0.12	73.0	0.10	0.10	1.50
		2	21	1,269	0.64	3.08b	4.3	0.58	0.19				
		3	22	1,170	0.53	1.74c	4.1	0.66	0.38				
		4	29	1,580	0.51	1.53d	2.6	0.35	0.23				

\* ANOVA with 95% significance; N - Number of samples; Min - Minimum; Max - Maximum; D.P. - Standard Deviation; C.V. - Coefficient of Variation; VR - Variance reduction; FPI - Fuzziness Performance Index; MPE - Modified Partition Entropy; CVI - Cluster Validation Index.



**Figure 4.** Soybean yield histograms for 2012 and 2013.

**Table 4.** Descriptive statistics of soybean yield data in the 2012-2013 harvest, separated by MZ and cluster validation indices.

N <sup>o</sup>	Interp.	MZ	% area	N	Min	Man	Max	D. P.	C. V.(%)	VR%	FPI	MPE	CVI
2	Not interpolated	1	41	2,168	1.30	4.00	11.73	0.81	20	0.0	0.16	0.19	1.5
		2	59	3,393	1.34	3.89	11.67	0.80	21				
	ID	1	51	2,606	1.30	4.05a	11.7	0.92	23	0.0	0.11	0.13	1.23
		2	49	2,955	1.34	3.82b	11.6	0.67	18				
	ISD	1	51	2,633	1.30	4.05a	11.7	0.91	22	0.0	0.11	0.14	1.22
		2	49	2,928	1.34	3.82b	11.6	0.68	18				
	KRI	1	53	2,719	1.30	4.04a	11.7	0.90	22	0.0	0.11	0.13	1.26
		2	47	2,842	1.34	3.82b	11.6	0.69	18				
3	Not interpolated	1	38	1,859	1.43	4.08a	11.76	0.92	23	0.0	0.14	0.15	1.23
		2	12	785	1.30	3.91b	11.33	0.93	24				
		3	50	2,917	1.34	3.84c	9.67	0.66	17				
	ID	1	43	2,255	1.30	3.99a	11.7	0.81	20	1.0	0.13	0.15	1.43
		2	21	1,115	1.36	4.09b	11.6	0.98	24				
		3	36	2,191	1.34	3.79c	11.6	0.67	17				
	ISD	1	42	2,212	1.30	3.98a	11.7	0.82	20	1.0	0.15	0.16	1.47
		2	22	1,135	1.36	4.11b	11.6	0.96	23				
		3	36	2,214	1.34	3.78c	11.6	0.68	18				
	KRI	1	38	2,008	1.30	3.96a	11.7	0.82	21	0.0	0.12	0.14	1.36
		2	26	1,401	1.34	4.08b	11.6	0.95	23				
		3	36	2,152	1.37	3.80c	11.6	0.65	17				
4	Not interpolated	1	36	1,955	1.36	4.00a	11.68	0.90	23	1.0	0.15	0.16	1.31
		2	21	1,014	1.45	4.01b	11.73	0.75	19				
		3	12	785	1.30	3.91ac	11.33	0.93	24				
		4	31	1,807	1.34	3.83d	7.59	0.64	17				
	ID	1	36	1,920	1.30	3.96a	11.7	0.83	21	2.0	0.15	0.15	1.50
		2	21	1,066	1.42	4.23b	11.6	0.93	22				
		3	10	592	1.36	3.92c	9.5	0.74	19				
		4	33	1,983	1.34	3.77d	11.7	0.67	18				
	ISD	1	29	1,672	1.30	3.97a	11.3	0.81	20	0.0	0.16	0.16	1.50
		2	24	1,112	1.43	4.17b	11.7	1.00	24				
		3	25	1,564	1.34	3.78c	11.6	0.74	20				
		4	22	1,213	1.37	3.86d	7.5	0.60	15				
	KRI	1	31	1,724	1.30	3.94a	11.7	0.86	22	1.0	0.14	0.15	1.50
		2	20	875	1.43	4.25b	11.6	0.97	23				
		3	19	1,267	1.34	3.90c	11.6	0.75	19				
		4	30	1,695	1.37	3.78d	7.5	0.63	22				

\* ANOVA with 95% significance; N- Number of samples; Min – Minimum; Max – Maximum; D.P. – Standard Deviation; C.V. – Coefficient of Variation; VR – Variance reduction; FPI – Fuzziness Performance Index; MPE – Modified Partition Entropy; CVI – Cluster Validation Index.

**Table 5.** Higher percentage of difference in area among ID, ISD and KRI interpolation methods as a function of the number of management zones (MZs).

Crop	No. of MZ	#ID-ISD	#ID-KRI	#ISD-KRI
Soybean 2011/2012	2	0%	0%	3%
	3	30%	30%	2%
	4	8%	4%	10%
Soybean 2012/2013	2	0%	2%	2%
	3	1%	5%	4%
	4	15%	9%	8%

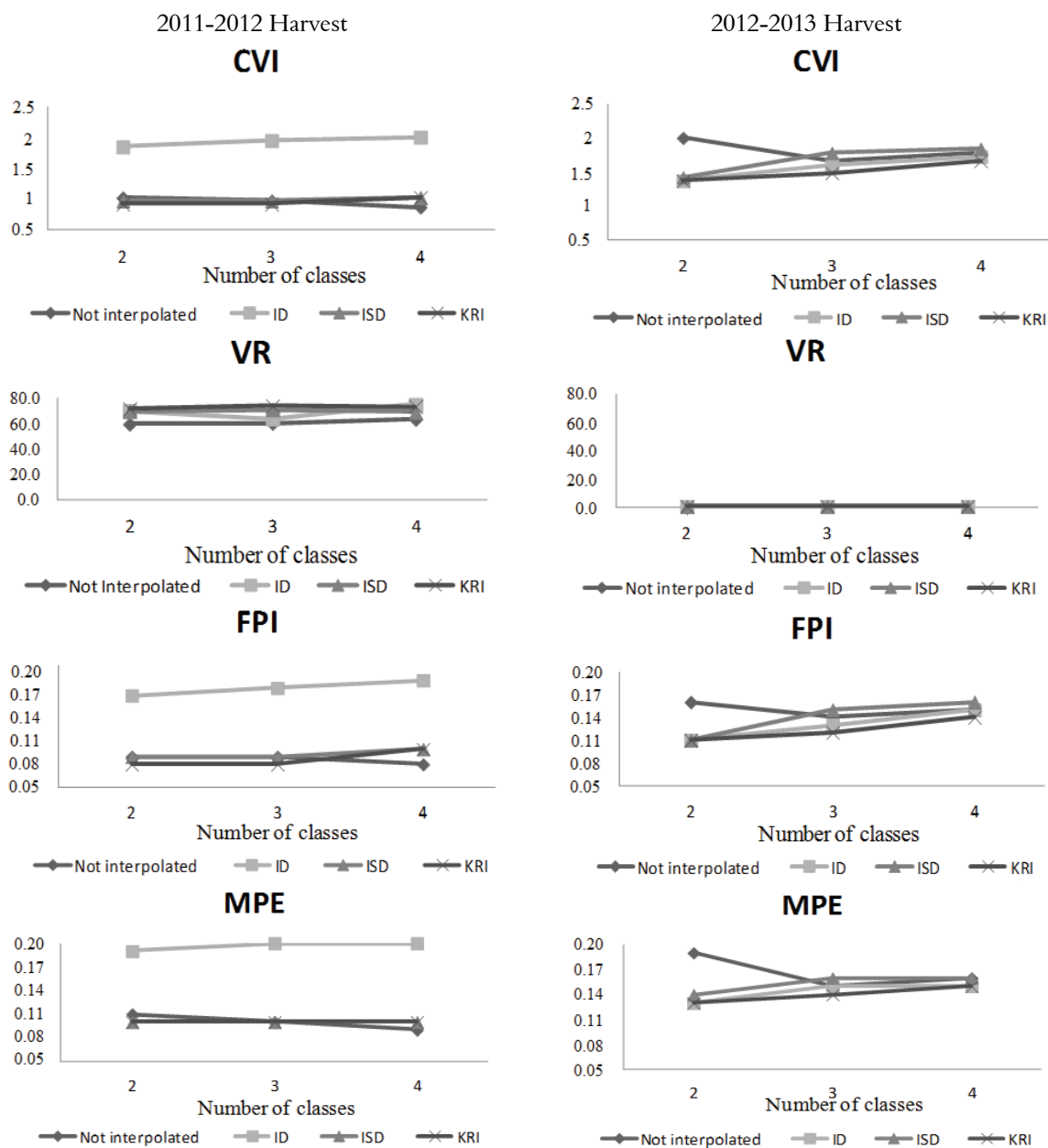
# - Difference in area between the interpolation methods.

**Table 6.** Comparison between management zone (MZ) maps generated by different interpolators using the Kappa Index.

Crop	MZ	Interpolators	Kappa
Soybean 2011-2012	2	ID x ISD	1.00(VS)
		ID x KRI	0.94(VS)
		ISD x KRI	0.94(VS)
	3	ID x ISD	0.54(M)
		ID x KRI	0.56(M)
		ISD x KRI	0.92(VS)
	4	ID x ISD	0.73(S)
		ID x KRI	0.84(VS)
		ISD x KRI	0.77(S)
Soybean 2012-2013	2	ID x ISD	0.96(VS)
		ID x KRI	0.93(VS)
		ISD x KRI	0.94(VS)
	3	ID x ISD	0.89(VS)
		ID x KRI	0.91(VS)
		ISD x KRI	0.91(VS)
	4	ID x ISD	0.78(S)
		ID x KRI	0.89(VS)
		ISD x KRI	0.80(S)

Concordance: weak (W); moderate (M); strong (S); very strong (VS).





**Figure 5.** Graphs for cluster validation index (CVI), variance reduction (VR), fuzziness performance index (FPI) and modified partition entropy index (MPE), according to the number of zones (2, 3 and 4) and the type of interpolation (ID – inverse distance, ISD – inverse square distance and KRI – ordinary kriging) for the 2011-2012 and 2012-2013 harvests.

## Conclusion

Data interpolation prior to the generation of management zones (MZ) is highly important to obtain MZs with smooth contours and to have a greater reduction of data variance;

The cluster validation index (CVI) proposed in current research was efficient in the choice of MZs and provided a less subjective decision on the best interpolator or on the number of MZs;

The best interpolation method was kriging and justifies the selection of the most robust interpolator in MZs generation.

## Acknowledgements

The authors would like to thank the State University of Western Paraná (UNIOESTE), the Federal Technological University of Paraná (UTFPR), the Araucária Foundation (Fundação Araucária), the Coordination for the Upgrading of Higher Education Personnel (CAPES), and the National Council for Scientific and Technological Development (CNPq) for funding. Thanks are also due to agronomist-engineer Aldo Tasca for the assignment of the research area.



## References

- Arno, J., Martinez-Casasnovas, J. A., Ribes-Dasi, M., & Rosell, J. R. (2011). Clustering of grape yield maps to delineate site-specific management zones. *Spanish Journal of Agricultural Research*, 9(3), 721-729.
- Bazzi, C. L., Souza, E. G., Uribe-Opazo, M. O., Nobrega, L. H., & Rocha, D. M. (2013). Management Zones definition using soil chemical and physical attributes in a soybean area. *Engenharia Agrícola*, 33(5), 1-14.
- Boydell, B., & Mcbratney, A. B. (2002). Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture*, 3(1), 9-23.
- Canarache, A. (1990). PENETR – A generalized semi-empirical model estimating soil resistance to penetration. *Soil & Tillage Research*, 16(1), 51-70.
- Cavalcante, E. G. S., Alves, M. C., Souza, Z. M., & Pereira, G. T. (2011). Variabilidade espacial de atributos físicos do solo sob diferentes usos e manejos. *Engenharia Agrícola e Ambiental*, 15(3), 237-243.
- Czaplewski, R. L., & Reich, R. M. (1993). *Expected value and variance of Moran's bivariate spatial autocorrelation statistic under permutation*. Fort Collins, CO: U. S. Department of Agriculture, Rocky Mountain Forest and Range Experiment Station. Research Paper. Vol. 309.
- Dalposso, G. H., Uribe-Opazo, M. A., Mercante, E., Johann, J. A., & Borssoi, J. A. (2012). Comparison measures of maps generated by geostatistical methods. *Engenharia Agrícola*, 32(1), 174-183.
- Delalibera, H. C., Weirich Neto, P. H., & Nagata, N. (2012). Management zones in agriculture according to the soil and landscape variables. *Engenharia Agrícola*, 32(6), 1197-1204.
- Doerge, T. A. (2000) *Site-specific management guidelines*. Norcross: Potash & Phosphate Institute.
- Empresa Brasileira de Pesquisa Agropecuária (Embrapa). (1997). *Manual de métodos de análise de solo*. Rio de Janeiro: Embrapa-CNPQ.
- Fridgen, J. J., Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Wiebold, W. J., & Fraisse, C. W. (2004). Management zone analyst (MZA): software for subfield management zone delineation. *Agronomy Journal*, 96(1), 100-108.
- Guastaferro, F., Castrignanó, A., De Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precision Agriculture*, 11(1), 600-620.
- Iliadis, L. S., Vangeloudh, M., & Spartalis, S. (2010). An intelligent system employing an enhanced fuzzy C-Means clustering model: application in the case of forest fires. *Computers and Electronics in Agriculture*, 70(1), 276-284.
- Jipkate, B. R., & Gohokar, V. V. (2012). A comparative analysis of Fuzzy C-Means clustering and K-Means clustering algorithms. *International Journal of Computational Engineering*, 2(3), 737-739.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Li, Y., Shi, Z., Wu, C., Li, H., & Li, F. (2008). Determination of potential management zones from soil electrical conductivity, yield and crop data. *Journal of Zhejiang University Science*, 9(1), 68-76.
- Li, Y., Shi, Z., Wu, H., Li, F., & Li, H. (2013). Definition of management zones for enhancing cultivated land conservation using combined spatial data. *Environmental Management*, 52(1), 792-806.
- López-Lozano, R., Casterad, M. A., & Herrero, J. (2010). Site-specific management zones in a commercial maize plot delineated using very high resolution remote sensing and soil properties mapping. *Computers and Electronics in Agriculture*, 73(1), 219-229.
- Mazzini, P. L. F., & Schettini, C. A. F. (2009). Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quase-sinópticos. *Brazilian Journal of Aquatic Science and Technology*, 13(1), 53-64.
- Michelan, R., Souza, E. G., & Uribe-Opazo, M. A. (2007). Determinação e remoção do tempo de atraso em mapas de colheita de milho. *Acta Scientiarum. Agronomy*, 29(2), 147-155.
- Minasny, B., & Mcbratney, A. B. (2002). *FuzME version 3*. Sidney: Australian Centre for Precision Agriculture. The University of Sydney.
- Mingoti, S. A., & Lima, J. O. (2006). Comparing SOM neural network with Fuzzy C-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, 174(1), 1742-1759.
- Molin, J. P., & Castro, C. N. de C. (2008). Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. *Scientia Agrícola*, 65(6), 567-573.
- Moral, F. J., Terrón, J. M., & Marques da Silva, J. R. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil & Tillage Research*, 106(1), 335-343.
- Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., & MacDonald, L. (2005). Spatial variability of measured soil properties across site-specific management zones. *Soil Science Society of America*, 69(1), 1572-1579.
- Ortega, R. A., & Santibáñez, O. A. (2007). Determination of management zones in corn (*Zea mays* L.) based on soil fertility. *Computers and Electronics in Agriculture*, 58(1), 49-59.
- Peralta, N., Costa, J. L., Franco, M. C., & Balzarini, M. (2013). Delimitación de zonas de manejo con modelos de elevación digital y profundidad de suelo. *Interciencia*, 38(6), 418-424.
- Ping, J. L., & Dobermann, A. (2003). Creating spatially contiguous yield classes for site-specific management. *Agronomy Journal*, 95(1), 121-113.
- Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for

- mapping soil properties. *Computers and Electronics in Agriculture*, 50(2), 97-108.
- Suszek, G., Souza, E. G., Uribe-Opazo, M. A., & Nobrega, L. H. P. (2011). Determination of management zones from normalized and standardized equivalent productivity maps in the soybean culture. *Engenharia Agrícola*, 31(5), 895-905.
- Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., & Gemtos, T. A. (2013). Management zones delineation using Fuzzy clustering techniques in grapevines. *Precision Agriculture*, 14(1), 18-39.
- Valente, D. S. M., Queiroz, D. M., Pinto, F. A. C., Santos, N. T., & Santos, F. L. (2012). Definition of management zones in coffee production fields based on apparent soil. *Scientia Agrícola*, 69(3), 173-179.
- Vrindts, E., Mouazen, A. M., Reyniers, M., Maertens, K., Maleki, M. R., Ramon, H., & Baerdemaeker, J. D. (2005). Management zones based on correlation between soil compaction, yield and crop data. *Biosystems Engineering*, 92(4), 419-428.
- Wollenhaupt, N. C., Wolkonski, R. P., & Clayton M. K. (1994). Mapping soil test phosphorus and potassium for variable-rate fertilizer applications. *Journal of production agriculture*, 7(4), 441-448.
- Xiang, L., Yu-Chun, P., Zhong-Qiang, G., & Chun-Jiang, Z. (2007). Delineation and scale effect of precision agriculture management zones using yield monitor data over four years. *Agriculture Sciences*, 6(1), 180-188.
- Yan, L., Zhou, S., Feng, L., & Hong-Yi, L. (2007). Delineation of site specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture*, 56(2), 174-186.

Received on May 8, 2015.

Accepted on September 16, 2015.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.