THE INTERNATIONAL SOCIETY OF PRECISION AGRICULTURE PRESENTS THE
13th INTERNATIONAL CONFERENCE ON
PRECISION AGRICULTURE
July 31-August 4, 2016 • St. Louis, Missouri USA

Delineation of site-specific management zones using spatial principal components and cluster analysis

A. Gavioli[1,2], E. G. Souza[2], C. L. Bazzi[1], N. M. Betzek[1], K. Schenatto[3], H. Beneduzzi[4]

[1] Department of Computer Science, Federal University of Technology of Paraná (UTFPR), Medianeira, Brazil.
[2] PGEAGRI, Technological and Exact Sciences Center, State University of West Paraná (UNIOESTE), Cascavel, Brazil.
[3] Department of Computer Science, Federal University of Technology of Paraná (UTFPR), Santa Helena, Brazil.
[4] Informatics Department, Federal Institute of Paraná (IFPR), Foz do Iguacu, Brazil.

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract. *The delineation of site-specific management zones (MZs) can enable economic use of precision agriculture for more producers. In this process, many variables, including chemical and physical (besides yield data) variables, can be used. After selecting variables, a cluster algorithm like fuzzy c-means is usually applied to define the classes. Selection of variables comprise a difficult issue in cluster analysis because these will often influence cluster determination. The goal of this study was to assess the effectiveness of the variable selection techniques - spatial correlation analysis, principal component analysis (PCA) and multivariate spatial analysis based on Moran's index PCA (MULTISPATI-PCA) - when used with the fuzzy c-means algorithm to generate MZs. The data used in experiments were collected from 2012 to 2014 in two agricultural fields with corn and soybean crops, located in Brazil. The variables selected were used as input for the fuzzy c-means, generating two, three, and four classes. The performance of the three techniques was assessed by applying analysis of variance (ANOVA), variance reduction index, fuzziness performance index, and modified partition entropy index. The delineated MZs were different according to the variable selection approach used along with fuzzy c-means. For the two agricultural fields, it was possible to* *define two classes with potential yields that showed statistically significant differences. The MULTISPATI-PCA technique resulted in classes with higher internal homogeneity, better performance of the clustering algorithm, the best variance reduction values, and the most viable MZs to be implemented in terms of field operations.*

Keywords. *fuzzy c-means, Moran's index, MULTISPATI-PCA, PCA, precision agriculture.*

Introduction

Management zones (MZs) have similar characteristics of soil and topography, and therefore, require management and similar amounts of agricultural supplies (Moral et al. 2010; Schepers et al. 2004). This approach can contribute significantly to enable precision agriculture for a larger number of producers, because in each MZ, one can standardize the management and application of supplies, varying formulations and field practices only when transitioning from one zone to another. Therefore, one can use the same mechanized systems already employed in traditional agriculture.

The MZs can also reduce the number of samples to be analyzed without compromising on the reliability of the results. Yield data, chemical and physical data of the soil, topographic data and apparent electrical conductivity of the soil, vegetation indexes, and combinations of these data, may be used to define MZs (Fraisse et al. 2001). However, it is recommended that stable variables correlated with yield be used for delimiting the sub-regions (Doerge 2000). This is so because the variables used for the definition are intended to be used for several years; hence, chemical attributes are eliminated. For this process of delimitation, is also customary to employ clustering algorithms such as fuzzy c-means (Fridgen et al. 2004; Fu et al. 2010; Hornung et al. 2006; Li et al. 2013; Zhang et al. 2013).

The selection of variables is a difficult task in cluster analysis (Gnanadesikan et al. 1995). Three variable selection techniques that can be applied in combination with clustering algorithms are as follows: spatial correlation analysis (Reich 2008; Schepers et al. 2004), applied as described by Bazzi et al. (2013) and Schenatto et al. (2016); principal component analysis (PCA) (Hotelling 1933), used by Fraisse et al. (2001), Li et al. (2007), Moral et al. (2010), and Cohen et al. (2013); and multivariate spatial analysis based on Moran’s index PCA (MULTISPATI-PCA) (Dray et al. 2008), applied by Córdoba et al. (2012), Córdoba et al. (2013), and Peralta et al. (2015).

For spatial correlation analysis, Moran's bivariate spatial autocorrelation statistic (Ord, 1975) is used to assess whether the variables have correlation and spatial autocorrelation. Thereafter, the variables without spatial dependence, those with no correlation with yield, and redundant variables are eliminated (Bazzi et al. 2013).

When using PCA, a new set of synthetic variables (principal components - PCs), which are uncorrelated among themselves and commonly denoted as linear combinations of the original variables, are obtained from the original variables through some transformations (Johnson and Wichern 2007).

MULTISPATI-PCA aims to add a spatial restriction on the traditional PCA, enabling it to be executed considering the existence of spatial dependence in sets of georeferenced data. This technique relies on introducing a spatial weighting matrix, which is constructed using Moran's bivariate spatial autocorrelation statistic, to the PCA. Its advantage over the PCA is that the scores obtained with MULTISPATI-PCA maximize the spatial autocorrelation between points, while those obtained with PCA maximize the total variance (Córdoba et al. 2013).

The aim of this study was to evaluate the effectiveness of the spatial correlation analysis, PCA, and MULTISPATI-PCA techniques, when used with the fuzzy c-means algorithm to generate MZs.

Materials and methods

Data collected between 2012 and 2014 from two commercial agricultural fields with corn and soybean crops, located in Paraná State, Brazil, were used (Figure 1). The soils were classified as typical dystroferric Red Latosol (Embrapa 2006) and grown in a no-till system. Field A extends for 15.5 ha, and is located in the municipality of Céu Azul (central geographical location 25°06'32" S and 53°49'55" W, and an average elevation of 460 m). Field B extends for 9.9 ha, and is located in the municipality of Serranópolis do Iguaçu (central geographical location 25°24'28" S and 54°00'17" W, and an average elevation of 355 m).

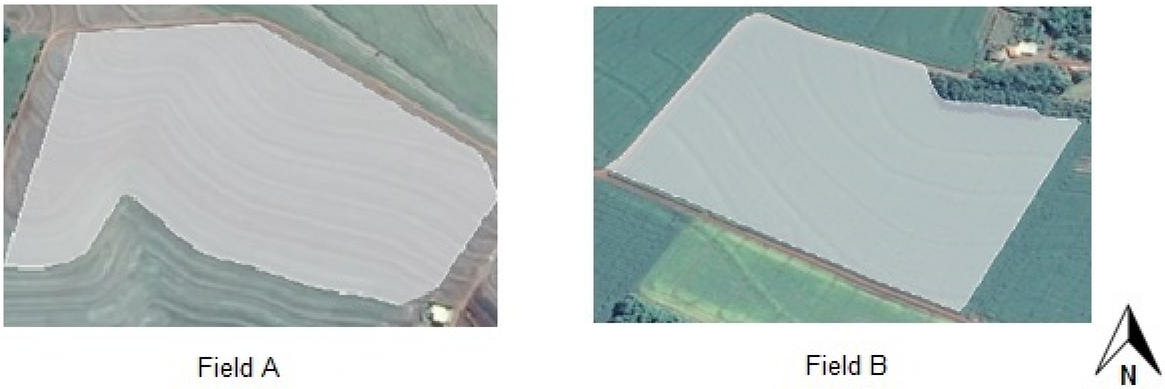

Fig. 1. The two experimental fields: field A: Céu Azul, Paraná, Brazil; field B: Serranópolis do Iguaçu, Paraná, Brazil.

Only those variables considered stable (Table 1) were used for defining the MZs, to meet the recommendation of Doerge (2000). Irregular sampling grids were used to assign 40 (2.67 points ha^{-1}) and 42 (4.24 points ha^{-1}) sample points to fields A and B, respectively, with the sampling points located in the central imaginary line between the contours present in each area.

Table 1. Data collected by year, for each experimental field.

Variable (Attribute)	Field A			Field B		
	2012	2013	2014	2012	2013	2014
SPR 0 - 0.1 m (MPa)	X	X	X	X	X	X
SPR 0.1 - 0.2 m (MPa)	X	X	X	X	X	X
SPR 0.2 - 0.3 m (MPa)	X	X	X	X	X	X
pH	X			X		
Elevation (m)	X			X		
Slope (°)	X					
Density (g cm^{-3})	X					
Sand (%)	X			X		
Silt (%)	X			X		
Clay (%)	X			X		
OM (%)	X			X		
Soybean yield (t ha^{-1})	X	X	X	X	X	X
Corn yield (t ha^{-1})					X	X

SPR: soil penetration resistance; OM: organic matter.

Soil samples were collected at depths of 0 – 0.2 m. The soil penetration resistance (SPR) was determined for the depths 0 – 0.1 m, 0.1 – 0.2 m, and 0.2 – 0.3 m, using an electronic meter of soil compaction Falker PenetroLOG PLG1020. The data of elevation of the two areas were obtained using an electronic total station of high precision Topcon GPT-7505, and subsequently, the slopes were calculated depending on the elevation of the sampling points.

Soybean yield data for field A was determined by means of a harvesting monitor attached to a CASE IV harvester. As for field B, yield was determined by hand harvesting of a 1 m^2 sample area in each of the sample points. In all cases, yield values were corrected to 13% water content.

To meet the requirement of stability of the yield data, which is normally heavily influenced by climate and rainfall, the data of soybean yield for the two fields, and data of corn yield for field B, were standardized through the standard score technique (Larscheid and Blackmore 1996) (Equation 1). Then, the arithmetic average of the standardized values of available years was calculated, generating a single variable corresponding to the average of standard yield.

$$P_{iN} = \frac{(P_i - \overline{P})}{S} \quad (1)$$

where P_{iN} is the standardized value for the sample point i; P_i is the original value of the sample point i; $\overline{P}$ corresponds to the arithmetic average of all the original values of the points to be standardized; and *S* corresponds to the standard deviation of the original values.

Four approaches for selecting variables for defining MZs were compared:

- All-Attrib: no disposal of stable variables;
- SCM (spatial correlation matrix): after calculating Moran's bivariate spatial autocorrelation statistic (Czaplewski and Reich 1993) (Equation 2) among all the variables by using the software for management zones definition (SDUM) (Bazzi et al. 2013), variables were selected by the procedure proposed by Bazzi et al. (2013): 1) elimination of variables with no significant spatial autocorrelation at 95% significance; 2) removal of the variables that were not correlated with yield; 3) decreasing ordination of the remaining variables, considering the degree of correlation with yield; and 4) elimination of variables which are correlated with each other, with preference to the withdrawal of those variables with lower correlation with yield.

  $$I_{XY} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij} * X_i * Y_j}{W\sqrt{m_X^{2} * m_Y^{2}}} \quad (2)$$

  where W_{ij} is the spatial association matrix, calculated by $W_{ij} = (1/(1+D_{ij}))$; D_{ij} is the distance between points *i* and *j*; X_i is the value of variable X transformed, at point *i*; Y_j is the value of the variable Y transformed, at point *j*; W corresponds to the sum of the degrees of spatial association, obtained from the W_{ij} matrix, for *i ≠ j*; m_X^{2} corresponds to the sample variance of X; and m_Y^{2} corresponds to the sample variance of Y. Note that the transformation of a variable Z should be interpreted as the procedure performed on their values so that it is on an average equal to zero, applying $Z_k = (z_k - \overline{Z})$, where $\overline{Z}$ is the sample mean of Z;
- PCA-All (PCA of all stable variables): calculation of PCs such that the amount of components selected was based on the criterion of representation of at least 70% of the total variability of the data associated with the original variables (Johnson and Wichern 2007);
- MPCA-All (MULTISPATI-PCA of all stable variables): calculation of spatial principal components (SPCs) such that the amount of components selected was also based on the criterion of representation of at least 70% of the total variability of the original data.

The PCA-All and MPCA-All approaches were applied to the data of each field by developing a routine in the statistical software R (R Core Team 2014), including the packages geoR, gstat, ade4 (Chessel et al. 2004), and spdep (Bivand 2012). For MPCA-All, the euclidean distance was used to compute the distance of a point from another and to return a list of neighbors for each point, based on the value set to neighborhood radius. This distance was determined experimentally for each location: field A, 240 m radius; and field B, 120 m radius.

The object-relational database system PostgreSQL 9.0.5, maintained by the PostgreSQL Global Development Group, was used for data storage. The software PostGIS 1.5.5, a spatial database extender for PostgreSQL maintained by the PostGIS Project Steering Committee, was also applied. Furthermore, the software pgAdmin III, maintained by the pgAdmin Development

Team, was used for managing the databases.

Data of the selected variables were interpolated using the inverse square distance method with pixels in an area of 5×5 m and 10 neighbors. After interpolation, resulting data were used as the input for the fuzzy c-means algorithm, considering error parameter equals to 0.0001 and weight index equals to 1.3, thus generating two, three, and four classes. For interpolating data and defining MZs, the software SDUM was used. For All-Attrib and SCM approaches, variables were standardized before interpolation (Equation 3) (Mielke Jr and Berry 2007), with the objective of maintaining the same data range, regardless of the used variable.

$$P_{in} = \frac{P_i - Median}{Range} \tag{3}$$

where P_i is the value of the pixel i to be standardized, and P_{in} is the standardization result.

The performance of the four variable selection approaches was evaluated using analysis of variance (ANOVA), variance reduction index (VR) (Li et al. 2007; Ping and Dobermann 2003), fuzziness performance index (FPI) (Fridgen et al. 2004), and modified partition entropy index (MPE) (Boydell and McBratney 2002).

The yield values were compared between classes by using the normalized average yield, and performing the Tukey's range test (ANOVA) to identify whether the generated sub-regions showed significant differences in normalized average yield (there was no spatial dependence within each class).

The variance reduction index is calculated for the standardized average yield, with the expectation that the sum of the variances of the data from MZs generated is smaller than the total variance (Equation 4).

$$VR = \left(1 - \frac{\sum_{i=1}^{c} W_i * V_{mz_i}}{V_{field}} \right) * 100 \tag{4}$$

where c is the number of MZs; W_i is the proportion of the area of i-th MZ to the total area; V_{mzi} is the data variance of the i-th MZ; and V_{field} is the data variance corresponding to the area as a whole.

Fuzziness performance index allows determining the degree of separation between the fuzzy c groups generated from a data set. FPI varies between 0 and 1, such that the closer this value to 0, the lower is the degree of sharing of elements among the generated groups (Equation 5).

$$FPI = 1 - \frac{c}{(c-1)} \left[1 - \sum_{j=1}^{n} \sum_{i=1}^{c} (m_{ij})^2 / n \right] \tag{5}$$

where c is the number of groups; n is the number of elements in the data set; and m_{ij} is the element of the fuzzy pertinence matrix $\boldsymbol{M}$.

Modified partition entropy is an estimate of the level of difficulty of organization of c groups, such that the closer the value to 0, the lower is the difficulty of organizing groups (Equation 6).

$$MPE = \frac{-\sum_{j=1}^{n} \sum_{i=1}^{c} m_{ij} \log(m_{ij}) / n}{\log c} \tag{6}$$

where c is the number of groups; n is the number of elements in the data set; and m_{ij} is the element of the fuzzy pertinence matrix $\boldsymbol{M}$.

Results and discussion

The variables selected for defining MZs by the SCM approach and the values of Moran's bivariate spatial autocorrelation statistic, between each variable and the normalized average yield value, are listed in Table 2.

Table 2. Moran's index with the normalized average yield and variables selected by SCM approach.

Field	Variable	Moran's index with NAY	Selected by SCM
A	SPR 0 - 0.1 m (MPa)	-0.053*	Y
	SPR 0.1 - 0.2 m (MPa)	-0.017	N
	SPR 0.2 - 0.3 m (MPa)	-0.022	N
	pH	-0.034*	N
	Elevation (m)	0.100*	Y
	Slope (˚)	-0.016	N
	Density (g cm^{-3})	0.023	N
	Sand (%)	-0.075*	N
	Silt (%)	0.028	N
	Clay (%)	-0.040*	N
B	SPR 0 - 0.1 m (MPa)	0.039*	Y
	SPR 0.1 - 0.2 m (MPa)	0.044*	N
	SPR 0.2 - 0.3 m (MPa)	-0.014	N
	pH	-0.029*	N
	Elevation (m)	0.051*	Y
	Sand (%)	0.007	N
	Silt (%)	-0.013	N
	Clay (%)	0.012	N
	OM (%)	-0.037*	N

*: significative value at 5%; SPR: soil penetration resistance; OM: organic matter; NAY: normalized average yield; SCM: spatial correlation matrix; Y: yes; N: no.

Because the values are not standardized, even small values of Moran's index can be statistically significant. In this case, the values are important if the statistic is significant at 0.05 level.

It was found that elevation was the variable with a strong spatial correlation with normalized average yield in the two areas. These findings agree with those of Jaynes et al. (2005) and Peralta et al. (2013), which suggests that there is a spatial association between this variable and yield of soybeans and corn. According to the criterion of the SCM approach, the variables selected for fields A and B were elevation and SPR 0 – 0.1 m.

When comparing PCA-All and MPCA-All approaches, from the viewpoint of variance and spatial autocorrelation (Table 3), the first spatial component (SPC1) obtained with MPCA- All had lower variance and higher spatial autocorrelation (Moran's index) than the first component (PC1) generated with PCA- All in the two fields. This indicates that the spatial autocorrelation values increase with the use of the MULTISPATI-PCA technique. Furthermore, MPCA-All had the best performance in reducing the dimensionality of data without significant loss of information, and therefore, MPCA-All ensured the highest cumulative percentage representation of the original variance with smaller number of components in the two fields. These findings agree with those of Córdoba et al. (2012) and Córdoba et al. (2013), who performed similar experiments in fields with soybean and wheat crops, in Argentina.

In the analysis of the coefficients of PCs and SPCs, which act as weights for the original variables in that components, the first component for PCA-All and MPCA-All had higher weighting coefficients for the variables as follows: elevation and clay to field A; and elevation and SPR 0 – 0.1 m to field B. Similar results for PCA-All were found by Fraisse et al. (2001) and Moral et al. (2010), and for MPCA-All approach were found by Córdoba et al. (2012) and Córdoba et al. (2013).

The MZs generated with the application of the four approaches along with fuzzy c-means were different according to the variable selection approach used (Figure 2). When using the All-Attrib approach for defining three or four classes in field A, field operations are difficult to perform in at least one of the classes owing to its small size and format. However, similar problem did not arise when using SCM, PCA-All, and MPCA-All.

Table 3. Statistics of the principal components for PCA-All and MPCA-All approaches.

Field	Variable (PC)	Variance	Percentage of variance	Sum of percentages	Moran's index
	PCA-All				
	PC1	2.98	27	27	0.23
	PC2	2.57	23	50	0.15
A	PC3	1.50	14	64	-0.05
	PC4	1.15	10	74	-0.05
	MPCA-All				
	SPC1	2.81	53	53	0.29
	SPC2	2.45	47	100	0.15
	PCA-All				
	PC1	3.20	32	32	0.01
	PC2	1.93	19	51	0.01
	PC3	1.33	13	64	0.07
B	PC4	1.18	12	76	0.03
	MPCA-All				
	SPC1	1.66	35	35	0.19
	SPC2	1.50	32	67	0.11
	SPC3	0.68	15	82	0.08

PCi: principal component *i*; SPCi: spatial principal component *i*.

The results of the evaluations of the shown MZs, according to ANOVA (Tukey's test), VR, FPI, and MPE indexes (Table 4), make it possible to state that the division of each field is possible in two classes with statistically different potential yields (not possible using All-Attrib). For field B, these results could be obtained only with MPCA-All approach.

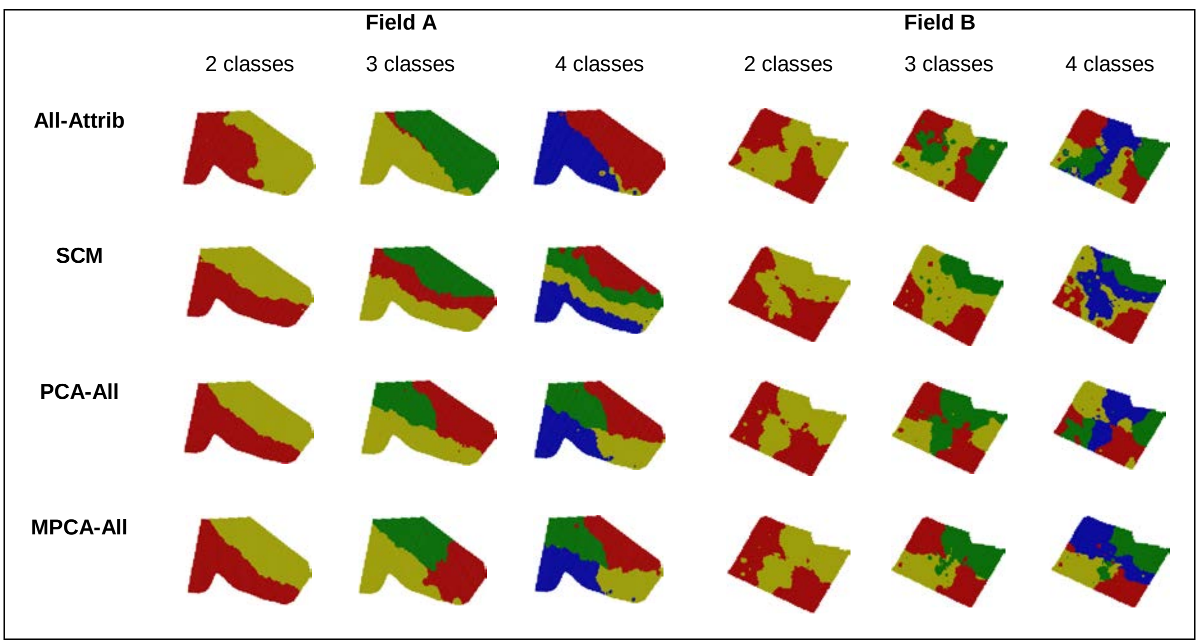

Fig. 2. Thematic maps generated by the four approaches: All-Attrib, SCM, PCA-All, and MPCA-All.

MPCA-All yielded the best results in terms of the variance reduction index for field B; in other words, this approach identified classes with larger differences between the respective normalized average yields and lower internal residual values. Differences in the normalized average yield between classes indicate that soil conditions influence the crop response. As previously mentioned, for the two areas, elevation was the variable that had the greatest influence among all variables on SPC1, and therefore, this variable was crucial to the results obtained with MPCA-All, as found by Cordoba et al. (2013) and Peralta et al. (2015).

Analysis of the values of the FPI and MPE indexes in the Table 4 showed that the MPCA-All method was, among the approaches that used all variables, the one that provided the best performance in combination with fuzzy c-means algorithm when defining the classes. This is so because this method yielded the lowest values of these two indexes. Similar results were found by Córdoba et al. (2013).

Table 4. Results for ANOVA (Tukey's range test), VR, FPI, and MPE indexes, for the two experimental fields.

Field	Classes	Approach	ANOVA				FPI	MPE	VR(%)
			C_1	C_2	C_3	C_4			
A	2	All-Attrib	a	a			0.500	0.079	0.0
		SCM	a	b			0.091	0.018	42.7
		PCA-All	a	b			0.185	0.035	42.5
		MPCA-All	a	b			0.161	0.030	25.5
	3	All-Attrib	a	a	a		0.667	0.125	0.0
		SCM	a	b	b		0.156	0.032	22.6
		PCA-All	a	a	b		0.287	0.058	39.8
		MPCA-All	a	a	b		0.212	0.043	16.7
	4	All-Attrib	a	a	a	a	0.750	0.158	0.0
		SCM	a	b	b	a	0.213	0.044	39.1
		PCA-All	a	b	b	a	0.314	0.069	28.1
		MPCA-All	a	ab	b	a	0.215	0.048	20.8
B	2	All-Attrib	a	a			0.285	0.054	4.1
		SCM	a	a			0.146	0.029	5.2
		PCA-All	a	a			0.292	0.054	1.7
		MPCA-All	a	b			0.255	0.048	15.1
	3	All-Attrib	a	a	a		0.667	0.132	8.5
		SCM	a	a	a		0.153	0.034	11.6
		PCA-All	a	a	a		0.357	0.076	2.2
		MPCA-All	a	ab	b		0.333	0.071	21.7
	4	All-Attrib	a	b	ab	ab	0.536	0.119	22.8
		SCM	a	ab	b	ab	0.239	0.052	15.3
		PCA-All	a	a	a	a	0.415	0.095	7.7
		MPCA-All	a	ab	b	ab	0.290	0.068	23.1

C_i: class *i*

From the viewpoint of the values of the variance reduction index, the MPCA-All approach showed the best results for field B, and PCA-All and SCM approaches showed the best results for field A. Finally, from the viewpoint of flatness of the generated MZs, it was found that the MPCA-All and SCM provided the best results, mainly for four classes. In other words, MPCA-All and SCM yielded MZs more viable in terms of field operations.

The use of MPCA-All allowed identification of the variables that account for global spatial variation. By using this approach, the part of the multidimensional variance that is spatially structured was analyzed. In addition to the works mentioned above, similar discussion about the treatment of multidimensional spatially structured variance by using MULTISPATI-PCA was addressed in the context of ecological data by Dray et al. (2008).

Conclusion

Among the PCs-based approaches, MPCA-All provided the best dimensionality reduction of the original data without significant loss of information. It also facilitated the generation of the MZs, and allowed them to be deployed more viable in terms of field operations.

The MPCA-All approach conducted, in most situations, distinguished the classes with larger differences between the respective normalized average yield values and lower internal residual values.

Therefore, the application of MULTISPATI-PCA technique on stable variables of fields can greatly improve the quality of bounded MZs with fuzzy c-means, and it is suitable for large multivariate data sets.

Acknowledgements

The authors would like to thank the State University of Western Paraná (UNIOESTE), the Federal University of Technology of Paraná (UTFPR), the Araucária Foundation, and the National Council for Scientific and Technological Development (CNPq) for funding.

References

Bazzi, C.L., Souza, E.G., Uribe-Opazo, M.A., Nóbrega, L.H.P., Rocha, D.M. (2013). Management zones definition using soil chemical and physical attributes in a soybean area. *Engenharia Agrícola*, 33, 952-964.

Bivand, R. (2012). *spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. Vienna: R Foundation for Statistical Computing.

Boydell, B., & McBratney, A.B. (2002). Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture*, 3, 9-23.

Chessel, D., Dufour, A.B., Thioulouse, J. (2004). The ade4 package-I-one-table methods. *R News 2004*, 5-10.

Cohen, S., Cohen, Y., Alchanatis, V., Levi, O. (2013). Combining spectral and spatial information from aerial hyperspectral images for delineating homogenous management zones. *Biosystems Engineering*, 114, 435-443.

Córdoba, M., Balzarini, M., Bruno, C., Costa, J.L. (2012). Principal component analysis with georeferenced data: An application in precision agriculture. *FCA UNCUYO*, 44, 27-39.

Córdoba, M., Bruno, C., Costa, J.L., Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics in Agriculture*, 97, 6-14.

Czaplewski, R.L., & Reich, R.M. (1993). Expected value and variance of Moran's bivariate spatial autocorrelation statistic under permutation. Fort Collins: U. S. Department of Agriculture.

Doerge, T.A. (2000). *Site-specific management guidelines*. Norcross: Potash & Phosphate Institute.

Dray, S., Saïd, S., Débias, F. (2008). Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *Journal of Vegetation Science*, 19, 45-56.

Embrapa - Brazilian Agricultural Research Corporation (2006). *Brazilian system of soil classification*. Rio de Janeiro: CNPSO.

Fraisse, C.W., Sudduth, K.A., Kitchen, N.R. (2001). Delineation of site–specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. *International Journal of the ASABE*, 44, 155-166.

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, 100-108.

Fu, Q., Wang, Z., Jiang, Q. (2010). Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO. *Mathematical and Computer Modelling*, 51, 1299-1305.

Gnanadesikan, R., Kettenring, J., Tsao, S. (1995). Weighting and selection of variables for cluster analysis. *Journal of Classification*, 12, 113-136.

Hornung, A., Khosla, R., Reich, R.M., Inman, D., Westfall, D.G. (2006). Comparison of Site-Specific Management Zones: Soil-Color-Based and Yield-Based. *Agronomy Journal*, 98, 407-415.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24, 417-441.

Jaynes, D.B., Colvin, T.S., Kaspar, T.C. (2005). Identifying potencial soybean management zones from multi-year yield data. *Computers and Electronics in Agriculture*, 46, 309-327.

Johnson, R.A., & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. 6 ed. New Jersey: Pearson.

Larscheid, G., & Blackmore, B.S. (1996). Interactions between farm managers and information systems with respect to yield mapping. In Proceedings of the International Conference on Precision Agriculture - ICPA 1996 (pp. 1153-1163). Minneapolis: American Society of Agronomy.

Li, Y., Shi, Z., Li, F., Li, H.Y. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture*, 56, 174-186.

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, 792-806.

Mielke Jr, P.W., & Berry, K.J. (2007). *Permutation methods: a distance function approach.* New York: Springer.

Moral, F.J., Terrón, J.M., Silva, J.R.M. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research*, 106, 335-343.

Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120-126.

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, 418-424.

Peralta, N.R., Costa, J.L., Balzarini, M., Franco, M.C., Córdoba, M., Bullock, D. (2015). Delineation of management zones to improve nitrogen management of wheat. *Computers and Electronics in Agriculture*, 110, 103-113.

Ping, J.L., & Dobermann, A. (2003). Creating spatially contiguous yield classes for site-specific management. *Agronomy Journal*, 95, 1121-1131.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reich, R. M. (2008). *Spatial statistical modeling of natural resources*. Fort Collins: Colorado State University.

Schenatto, K., Souza, E.G., Bazzi, C.L., Bier, V.A., Betzek, N.M., Gavioli, A. (2016). Data interpolation in the definition of management zones. *Acta Scientiarum*, 38, 31-40.

Schepers, A.R., Shanahan, F, J., Liebig, M.A., Schepers, J.S., Johnson, S.H., Luchiari, J.A. (2004). Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. *Agronomy Journal*, 96, 195-203.

Zhang, Z., Lu, X., Lv, N., Chen, J., Feng, B., Li, X.W., Ma, L. (2013). Defining agricultural management zones using GIS techniques: Case study of drip-irrigated cotton fields. *Information Technology Journal*, 12, 6241-6246.