

APLICAÇÃO DE ANÁLISE DE COMPONENTES PRINCIPAIS ESPACIAIS E ALGORITMO FUZZY C-MEANS PARA DEFINIÇÃO DE ZONAS DE MANEJO

ALAN GAVIOLI¹, EDUARDO G. SOUZA², GABRIELA K. MICHELON³, CLAUDIO L. BAZZI⁴, KELYN SCHENATTO⁵

¹Bacharel e Mestre em Ciência da Computação, Doutorando em Eng. Agrícola, Professor do Depto. de Computação, Universidade Tecnológica Federal do Paraná (UTFPR), Medianeira – PR, Fone: (45) 3240-8000, alan@utfpr.edu.br.

²Eng. Mecânico, Doutor em Eng. Mecânica, Prof. Associado, Universidade Estadual do Oeste do Paraná (UNIOESTE), Cascavel – PR.

³Acadêmica de Ciência da Computação, (UTFPR), Medianeira – PR.

⁴Tecnólogo em Proc. de Dados, Doutor em Eng. Agrícola, Prof. Adjunto, (UTFPR), Medianeira – PR.

⁵Tecnóloga em Análise e Desenv. de Sistemas, Doutoranda em Eng. Agrícola, Professora do Depto. de Computação, (UTFPR), Santa Helena – PR

Apresentado no
XLV Congresso Brasileiro de Engenharia Agrícola - CONBEA 2016
24 a 28 de julho de 2016 - Florianópolis - SC, Brasil

RESUMO: A definição de zonas de manejo (ZMs) corresponde à delimitação de sub-regiões dentro de um talhão que apresentem características do solo similares. Neste processo, costumam ser aplicados algoritmos de agrupamento, como fuzzy c-means. Três técnicas de seleção de variáveis para definição de ZMs que podem ser aplicadas com esse algoritmo de agrupamento são a análise de matriz de correlação espacial, análise de componentes principais (PCA), e uma extensão de PCA denominada *multivariate spatial analysis based on Moran's index PCA* (MULTISPATI-PCA). Neste trabalho, avaliou-se a eficiência destas três técnicas utilizadas com fuzzy c-means. Tais avaliações foram realizadas com dados coletados de 2010 a 2014 em duas áreas agrícolas localizadas no Estado do Paraná, para gerar duas, três e quatro classes para cada área. A partir dessas classes, foram elaborados mapas temáticos com as respectivas delimitações de ZMs a serem implantadas. A abordagem baseada em MULTISPATI-PCA proporcionou o melhor desempenho do algoritmo de agrupamento e ZMs mais viáveis para implantação, em termos de operações de campo.

PALAVRAS-CHAVE: agricultura de precisão, índice de Moran bivariado, MULTISPATI-PCA.

APPLICATION OF SPATIAL PRINCIPAL COMPONENT ANALYSIS AND FUZZY C-MEANS ALGORITHM FOR DEFINITION OF MANAGEMENT ZONES

ABSTRACT: The definition of site-specific management zones (MZs) corresponds to the definition of subregions inside an agricultural field that have similar soil characteristics. In this process, cluster algorithms, like fuzzy c-means, are usually applied. Three variable selection techniques that can be applied with this cluster algorithm are analysis of spatial correlation matrix, principal component analysis (PCA), and an extension of PCA technique – multivariate spatial analysis based on Moran's index PCA (MULTISPATI-PCA). In this study, the efficiency of each of these three techniques used with fuzzy c-means was evaluated. The data used in experiments were collected from 2010 to 2014 in two agricultural fields located in the State of Parana, generating two, three, and four classes for each field. Thematic maps with MZ delineations were designed from the classes. The MULTISPATI-PCA approach provided the best performance for the cluster algorithm, and MZs more viable for implementation from the viewpoint of field operations.

KEYWORDS: precision agriculture, Moran's bivariate index, MULTISPATI-PCA.

INTRODUÇÃO

A definição de zonas de manejo (ZMs) pode colaborar no processo de viabilizar economicamente a agricultura de precisão (AP) para um maior número de produtores, pois em cada ZM pode-se padronizar o manejo e a aplicação de insumos, variando-se formulações e/ou práticas de campo quando se passa de uma zona para outra. Essas subáreas também podem representar indicadores para amostragem do solo e das culturas plantadas, reduzindo o número de amostras que precisam ser analisadas. (FRIDGEN et al., 2004; MORAL et al., 2010).

Dentre diversas variáveis que podem ser utilizadas no processo de definição de ZMs, há recomendações de que sejam considerados os atributos estáveis que estejam correlacionados com a produtividade (DOERGE, 2000), para que as subáreas possam ser utilizadas por vários anos. Nesse processo, também costumam ser usados algoritmos de agrupamento como o fuzzy c-means (FU et al., 2010; HORNUNG et al., 2006; LI et al., 2013; ZHANG et al., 2013). Três técnicas de seleção de variáveis que podem ser aplicadas junto com esse algoritmo são: análise de matriz de correlação espacial (BAZZI et al., 2013; REICH, 2008; SCHEPERS et al., 2004), aplicada conforme descrito por Schenatto et al. (2016); análise de componentes principais (PCA) (HOTELLING, 1933), utilizada por Fraisse et al. (2001), Li et al. (2007) e Cohen et al. (2013); e *Multivariate Spatial Analysis based on Moran's Index* PCA (MULTISPATI-PCA) (DRAY et al., 2008), empregada por Córdoba et al. (2016) e Peralta et al. (2015).

A estatística de autocorrelação espacial bivariada de Moran (ORD, 1975), utilizada na técnica de seleção por análise de matriz de correlação espacial, possibilita avaliar se variáveis possuem correlação e autocorrelação espacial. Com o uso desta estatística, pode-se eliminar do processo de geração de ZMs as variáveis sem dependência espacial, ou sem correlação com a produtividade, ou redundantes (BAZZI et al., 2013).

PCA permite identificar as variáveis que explicam a maior parte da variância total contida em conjuntos de dados. Ao utilizá-la, são realizadas transformações a partir das variáveis originais que resultam em um novo conjunto de variáveis sintéticas, as componentes principais (CPs), não correlacionadas entre si (JOLLIFFE, 2002).

Já MULTISPATI-PCA visa adicionar uma restrição espacial à técnica PCA tradicional, possibilitando que PCA seja executada considerando a existência de dependência espacial em conjuntos de dados georreferenciados (ARROUAYS et al., 2011). Segundo Dray et al. (2008), a aplicação de MULTISPATI-PCA facilita a seleção das componentes principais necessárias, além de possibilitar a maximização da autocorrelação espacial entre pontos, algo que não é possível com PCA.

O objetivo deste trabalho foi avaliar o desempenho da análise de matriz de correlação espacial, de PCA e de MULTISPATI-PCA, quando aplicadas juntamente com fuzzy c-means para a definição de ZMs em áreas com cultivo de soja e milho.

MATERIAL E MÉTODOS

Foram utilizados dados coletados no período de 2010 a 2014 em duas áreas agrícolas comerciais (Figura 1), com solos classificados como LATOSSOLO VERMELHO Distroférico típico (EMBRAPA, 2006), com cultivo de soja e milho em sistema de plantio

direto. As áreas estão localizadas no Estado do Paraná: área A1: com 9,9 ha, localizada no município de Serranópolis do Iguaçu, localização geográfica central de 25°24'28" S e 54°00'17" O e altitude média de 355 m; e área A2: com 19,8 ha, localizada no município de Cascavel, localização geográfica central de 24°57'08" S e 53°33'59" O e altitude média de 650 m.

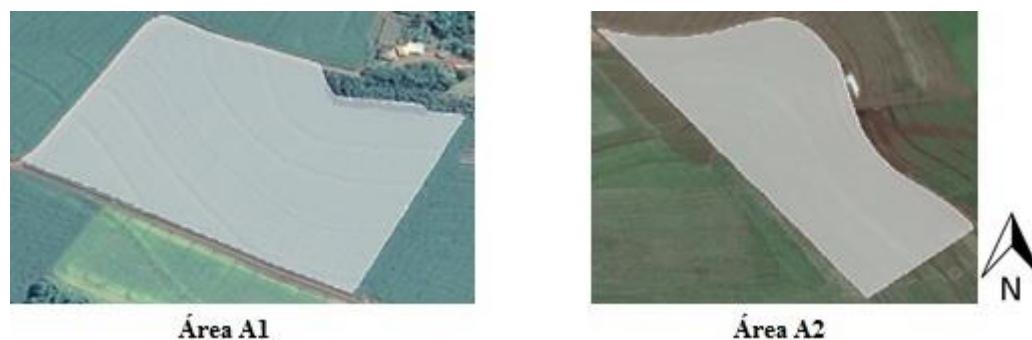


FIGURA 1. Áreas experimentais: A1, localizada em Serranópolis do Iguaçu – PR; A2, localizada em Cascavel – PR.

Para a definição das classes e, conseqüentemente, das ZMs, foram utilizadas apenas variáveis consideradas estáveis (Tabela 1), visando atender à recomendação de Doerge (2000). Utilizando grades amostrais irregulares, realizou-se coleta em 42 (4,24 pontos ha⁻¹) e 68 (3,43 pontos ha⁻¹) pontos amostrais para as áreas A1 e A2, respectivamente. Os pontos amostrais estavam localizados na linha imaginária central entre as curvas de nível presentes em cada área.

TABELA 1. Identificação das variáveis e períodos de coleta, para cada área experimental.

Variável	Área A1 (42 pontos)			Área A2 (68 pontos)	
	2012	2013	2014	2010	2011
RSP 0 – 0,1 m (MPa)	X	X	X	X	
RSP 0,1 – 0,2 m (MPa)	X	X	X	X	
RSP 0,2 – 0,3 m (MPa)	X	X	X	X	
pH	X			X	
Elevação (m)	X			X	
Declividade (°)				X	
Densidade (g cm ⁻³)				X	
Areia (%)	X			X	
Silte (%)	X			X	
Argila (%)	X			X	
MO (%)	X				
Produtividade soja (t ha ⁻¹)	X	X	X	X	X
Produtividade milho (t ha ⁻¹)		X	X		

RSP: Resistência do solo à penetração; MO: Matéria orgânica.

As amostras de solo foram coletadas à profundidade de 0 – 0,2 m. A resistência mecânica do solo à penetração (RSP) foi determinada para as profundidades de 0 – 0,1 m, 0,1 – 0,2 m e 0,2 – 0,3 m, utilizando-se um medidor eletrônico de compactação do solo Falker PenetroLOG PLG1020. Os dados referentes à altitude foram obtidos utilizando-se uma estação total eletrônica de alta precisão Topcon GPT-7505, sendo posteriormente calculados os valores de declividade em função da altitude dos pontos amostrais.

Os dados de produtividade foram determinados por meio da colheita manual de uma

área de amostragem de 1 m² em cada um dos pontos amostrais. Em todos os casos, os valores de produtividade foram corrigidos para um teor de água de 13%. Visando atender ao requisito de estabilidade dos dados de produtividade de soja e de milho, normalmente bastante influenciados pelo clima e pela precipitação pluviométrica, realizou-se a normalização desses dados por meio da técnica *escore padrão* (Equação 1) (LARSCHEID; BLACKMORE, 1996). Em seguida, calculou-se a média aritmética dos valores normalizados dos anos disponíveis, gerando-se assim uma única variável correspondente à média das produtividades normalizadas.

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

em que, P_{iN} é o valor normalizado para o ponto amostral i ; P_i é o valor original do ponto amostral i ; \bar{P} corresponde à média aritmética de todos os valores originais dos pontos a serem normalizados; e S corresponde ao desvio padrão dos valores originais.

Foram comparadas quatro técnicas de seleção de variáveis para definição de ZMs: a) All-Attrib: técnica sem aplicação de algoritmos e sem descarte de variáveis; b) SCM: análise de matriz de correlação espacial; c) PCA-All: PCA executada a partir das variáveis originais; d) MPCA-All: MULTISPATI-PCA executada a partir das variáveis originais.

A técnica SCM foi aplicada da seguinte forma: após o cálculo da estatística de correlação espacial bivariada de Moran (Equação 2) entre todas as variáveis, utilizando o software SDUM (software para definição de unidades de manejo) (BAZZI et al., 2013), foram selecionadas variáveis pelo procedimento proposto por Bazzi et al. (2013): a) eliminação das variáveis com autocorrelação espacial não significativa a 95% de significância; b) remoção das variáveis que não apresentaram correlação com a produtividade; c) ordenação decrescente das variáveis restantes, considerando o módulo do grau de correlação com a produtividade; e d) eliminação das variáveis redundantes (com correlação significativa entre si), dando preferência para a retirada das que apresentaram menor correlação com a produtividade.

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

em que, W_{ij} é a matriz de associação espacial, calculada por $W_{ij} = (1/(1 + D_{ij}))$, sendo D_{ij} a distância entre os pontos i e j ; X_i é o valor da variável X transformada, no ponto i ; Y_j é o valor da variável Y transformada, no ponto j ; W corresponde à soma dos graus de associação espacial, obtidos da matriz W_{ij} , para $i \neq j$; m_x^2 corresponde à variância amostral de X ; m_y^2 corresponde à variância amostral de Y . Deve-se interpretar a transformação de uma variável Z como o procedimento executado sobre seus valores para que ela fique com média igual a zero, aplicando-se $Z_k = (z_k - \bar{Z})$, em que \bar{Z} representa a média amostral de Z .

A técnica PCA-All foi aplicada construindo-se as CPs a partir das variáveis originais, tal que a quantidade de CPs selecionadas foi baseada no critério da representação de ao menos 70% da variabilidade total dos dados associados às variáveis originais (JOHNSON; WICHERN, 2007).

Já MPCA-All foi executada construindo-se as componentes principais espaciais (CPEs) a partir das variáveis originais, tal que a quantidade de CPEs selecionadas também foi baseada no critério da representação de ao menos 70% da variabilidade total dos dados originais.

Para executar PCA-All e MPCA-All sobre os dados de cada área, foi desenvolvida uma rotina no software estatístico R (R CORE TEAM, 2014), incluindo os pacotes *geoR*, *gstat*, *ade4* (CHESSEL et al., 2004) e *spdep* (BIVAND, 2012). Especificamente para MPCA-All, a

distância euclidiana foi usada para calcular a distância de cada ponto em relação aos demais e retornar como resultado uma lista de vizinhos para cada ponto. Esta lista de vizinhos foi elaborada com base no valor definido como maior distância permitida para que um ponto seja considerado vizinho de outro (raio de vizinhança). Neste trabalho, foi determinado experimentalmente que o raio adequado para a área A1 era de 120 m e para a área A2 era de 200 m.

Além dos softwares SDUM e R, outros aplicativos também foram necessários. O sistema gerenciador de bancos de dados objeto-relacionais PostgreSQL 9.0.5, mantido por *PostgreSQL Global Development Group*, foi utilizado para o armazenamento dos dados. O software PostGIS 1.5.5, considerado uma extensão do PostgreSQL para bancos de dados espaciais e mantido por *PostGIS Project Steering Committee*, também foi aplicado. Além disso, o aplicativo pgAdmin III, mantido por *pgAdmin Development Team*, foi usado para gerenciar as bases de dados criadas.

Os dados das variáveis selecionadas foram interpolados utilizando-se o método do inverso do quadrado da distância, com pixel de 5x5 m e dez vizinhos. Depois de interpolados, os dados resultantes foram usados como entrada para o algoritmo fuzzy c-means, considerando os parâmetros de erro = 0,0001 e índice de ponderação = 1,3, gerando-se então duas, três e quatro classes. A seleção do interpolador e dos valores citados para os demais parâmetros foi feita após diversos testes com os dados das duas áreas. Este trabalho não tem como objetivo descrever esses testes.

A definição das classes e a geração de mapas temáticos com as ZMs correspondentes às classes foram realizadas após interpolação, feita utilizando-se o software SDUM. Para as abordagens All-Attrib e SCM, as variáveis foram normalizadas antes de passarem pelo processo de interpolação, empregando-se a Equação 3 (MIELKE; BERRY, 2007), tendo em vista o objetivo de manter a mesma amplitude dos dados independentemente da variável utilizada.

$$P_{in} = \frac{P_i - Mediana}{Amplitude} \quad (3)$$

em que, P_i é o valor do pixel i a ser normalizado, e P_{in} é o valor resultante da normalização.

O desempenho das técnicas de seleção de variáveis foi avaliado empregando-se análise de variância (ANOVA), e os índices *fuzziness performance index* (FPI) (FRIDGEN et al., 2004) e *modified partition entropy* (MPE) (BOYDELL; MCBRATNEY, 2002).

Os valores de produtividade foram comparados entre as classes definidas pelo algoritmo fuzzy c-means, considerando a produtividade média normalizada de cada classe. O teste de Tukey (ANOVA) foi aplicado para identificar se as classes geradas apresentaram diferenças significativas de produtividade média normalizada, após a constatação de que não havia dependência espacial dentro de cada classe.

Para a escolha do número ideal de classes, foram utilizados os índices FPI e MPE. O FPI possibilita determinar o grau de separação entre os c grupos *fuzzy* gerados a partir de um conjunto de dados. O FPI varia entre 0 e 1, tal que quanto mais próximo o valor for de 0, menor é o grau de compartilhamento de elementos entre os grupos gerados (Equação 4).

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

em que, c é o número de grupos; n é a quantidade de elementos no conjunto de dados; e m_{ij} é o elemento da matriz de pertinência *fuzzy* M .

O índice MPE é uma estimativa do nível de dificuldade para organização dos c grupos, tal que quanto mais próximo o valor for de 0, menor foi a dificuldade do algoritmo para a organização dos grupos (Equação 5).

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

em que, c é o número de grupos; n é a quantidade de elementos no conjunto de dados; e m_{ij} é o elemento da matriz de pertinência *fuzzy* M .

RESULTADOS E DISCUSSÃO

As variáveis selecionadas para definição de ZMs pela técnica SCM, bem como os valores da estatística de correlação espacial bivariada de Moran (índice de Moran) entre cada variável e a produtividade média normalizada (PMN), estão apresentados na Tabela 2. Como estes valores do índice de Moran não foram padronizados, mesmo pequenos valores podem ser estatisticamente significativos. Neste caso, foram considerados importantes quando foram significativos ao nível de 5%.

TABELA 2. Variáveis selecionadas para definição de ZMs pela abordagem SCM, e índice de Moran entre cada variável e a PMN.

Área	Variável	Índice de Moran com PMN	Selecionada por SCM
A1	RSP 0 - 0,1 m (MPa)	0,039*	S
	RSP 0,1 - 0,2 m (MPa)	0,044*	N
	RSP 0,2 - 0,3 m (MPa)	-0,014	N
	pH	-0,029*	N
	Elevação (m)	0,051*	S
	Areia (%)	0,007	N
	Silte (%)	-0,013	N
	Argila (%)	0,012	N
	MO (%)	-0,037*	N
A2	RSP 0 - 0,1 m (MPa)	-0,002	N
	RSP 0,1 - 0,2 m (MPa)	0,114*	N
	RSP 0,2 - 0,3 m (MPa)	0,102*	N
	pH	0,024	N
	Elevação (m)	0,137*	S
	Declividade (°)	0,011	N
	Densidade (g cm ⁻³)	-0,029	N
	Areia (%)	0,078*	N
	Silte (%)	0,021	N
Argila (%)	-0,082*	N	

*: valor significativo a 5%; RSP: resistência do solo à penetração; MO: matéria orgânica; PMN: produtividade média normalizada; SCM: *spatial correlation matrix*; S: sim; N: não.

Verificou-se que a elevação foi a variável que apresentou correlação espacial mais forte com a PMN nas duas áreas. Estas constatações estão alinhadas com Jaynes et al. (2005) e Peralta et al. (2013), que sugeriram que há associação espacial entre essa variável e a produtividade de soja e de milho. Pelo critério da matriz de correlação espacial, as variáveis selecionadas para a área A1 foram elevação e RSP 0 – 0,1 m, enquanto que para a área A2 somente a elevação foi selecionada.

Comparando-se as estatísticas das primeiras CPs geradas com PCA-All e MPCA-All (Tabela 3), constatou-se que MPCA-All proporcionou maior redução da dimensionalidade dos dados sem perda significativa de informação, pois garantiu as maiores porcentagens acumuladas de representação da variância original com as menores quantidades de CPs, nas duas áreas. Estes resultados estão de acordo com os obtidos por Córdoba et al. (2012), que realizaram experimentos similares em áreas com cultivo de soja e trigo na Argentina.

TABELA 3. Estatísticas das primeiras componentes principais geradas com PCA-All e MPCA-All, para as duas áreas experimentais.

Área	Variável (CP)	Variância	Porcentagem da variância	Soma das porcentagens	Índice de Moran	
A1	PCA-All					
	CP1	2,98	27	27	0,23	
	CP2	2,57	23	50	0,15	
	CP3	1,50	14	64	-0,05	
	CP4	1,15	10	74	-0,05	
	MPCA-All					
	CPE1	2,81	53	53	0,29	
	CPE2	2,45	47	100	0,15	
A2	PCA-All					
	CP1	3,20	32	32	0,01	
	CP2	1,93	19	51	0,01	
	CP3	1,33	13	64	0,07	
	CP4	1,18	12	76	0,03	
	MPCA-All					
		CPE1	1,66	35	35	0,19
		CPE2	1,50	32	67	0,11
	CPE3	0,68	15	82	0,08	

CP: componente principal *i*; CPE: componente principal espacial *i*.

Na análise dos coeficientes das CPs e CPEs, que atuam como ponderações sobre as variáveis originais nessas componentes, verificou-se que as primeiras componentes possuem coeficientes maiores para as seguintes variáveis: elevação e argila, para a área A1; e elevação, RSP 0,1 – 0,2 m, RSP 0,2 – 0,3 m e argila, para a área A2. Resultados similares utilizando PCA-All foram encontrados por Fraisse et al. (2001) e Moral et al. (2010), e com MPCA-All foram obtidos por Córdoba et al. (2013) e Peralta et al. (2015).

As ZMs geradas com a aplicação das quatro abordagens de seleção de variáveis em conjunto com fuzzy c-means foram influenciadas pela abordagem utilizada (Figura 2).

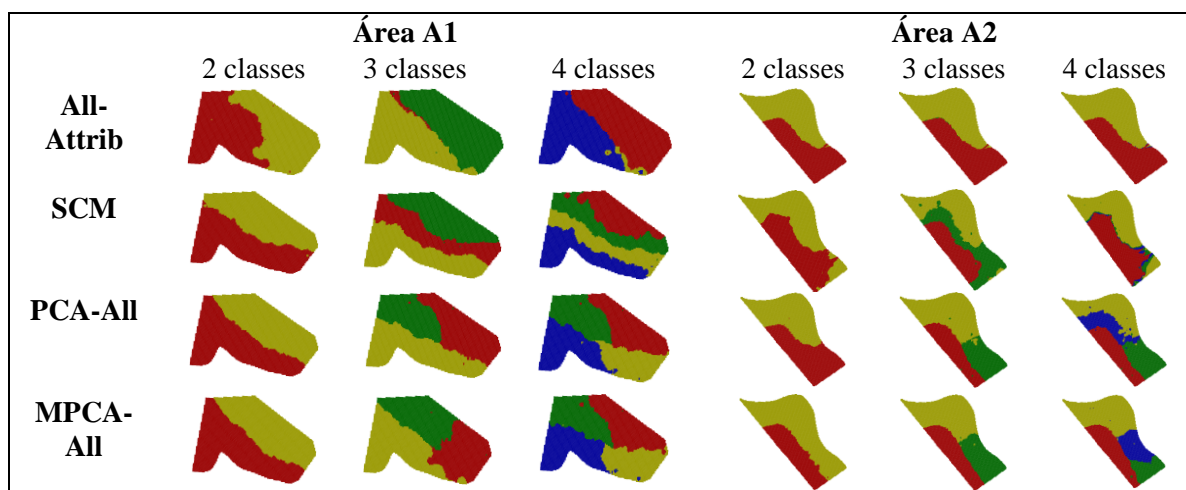


FIGURA 2. Mapas temáticos representando as zonas de manejo geradas pelas abordagens All-Attrib, SCM, PCA-All e MPCA-All.

Para a geração de três ou quatro classes para a área A1 utilizando All-Attrib, as operações de campo seriam de difícil execução para ao menos uma das classes, devido à

quantidade de pontos muito inferior às demais classes e ao formato da respectiva ZM delimitada. Esta mesma dificuldade poderia ocorrer no caso da aplicação de SCM para gerar quatro classes para a área A2. Outra situação relacionada a All-Attrib é que esta não foi capaz de gerar três ou quatro classes para a área A2. Por outro lado, nenhum destes problemas ocorreu nas definições de classes resultantes da aplicação de PCA-All e MPCA-All. Córdoba et al. (2012) relataram resultados similares, principalmente sobre a superioridade das subáreas definidas a partir de MPCA-All no que tange à facilitação das operações de campo.

Os resultados das avaliações das ZMs mostradas, segundo o teste de comparação de médias (ANOVA) e os índices FPI e MPE (Tabela 4), possibilitam afirmar que é possível a divisão de cada área em duas classes com potenciais médios de produtividade estatisticamente distintos. Resultados similares foram obtidos por Peralta et al. (2015), que também se basearam nestes índices para optar pela definição de duas classes para dados provenientes de áreas com cultivo de trigo.

TABELA 4. Resultados obtidos para ANOVA (Teste de Tukey), FPI e MPE, para as duas áreas experimentais.

Área	Classes	Método	ANOVA (Teste de Tukey)				FPI	MPE
			C ₁	C ₂	C ₃	C ₄		
A1	2	All-Attrib	a	a			0,500	0,079
		SCM	a	b			0,091	0,018
		PCA-All	a	b			0,185	0,035
		MPCA-All	a	b			0,161	0,030
	3	All-Attrib	a	a	a		0,667	0,125
		SCM	a	b	b		0,156	0,032
		PCA-All	a	a	b		0,287	0,058
		MPCA-All	a	a	b		0,212	0,043
	4	All-Attrib	a	a	a	a	0,750	0,158
		SCM	a	b	b	a	0,213	0,044
		PCA-All	a	b	b	a	0,314	0,069
		MPCA-All	a	ab	b	a	0,215	0,048
A2	2	All-Attrib	a	b			0,500	0,077
		SCM	a	b			0,495	0,076
		PCA-All	a	b			0,206	0,037
		MPCA-All	a	b			0,162	0,030
	3	All-Attrib	a	a	a		0,667	0,122
		SCM	a	a	b		0,108	0,023
		PCA-All	a	b	a		0,189	0,040
		MPCA-All	a	b	b		0,147	0,031
	4	All-Attrib	a	a	a	a	0,750	0,154
		SCM	a	b	bc	ac	0,535	0,111
		PCA-All	a	ab	b	c	0,286	0,061
		MPCA-All	a	b	ac	bc	0,166	0,037

C_i: classe *i*.

A análise dos valores dos índices FPI e MPE permitiram constatar que, entre os três métodos que consideraram todas as variáveis para a análise de agrupamento (All-Attrib, PCA-All e MPCA-All), o que proporcionou ao algoritmo fuzzy c-means o melhor desempenho foi MPCA-All. Isto porque este método obteve os menores valores para esses dois índices. Estes resultados estão de acordo com os que foram apresentados por Córdoba et al. (2013), mesmo estes tendo realizado seus experimentos em áreas com cultivo de trigo.

CONCLUSÕES

As duas abordagens de seleção de variáveis baseadas em PCA e MULTISPATI-PCA conduziram à formação de zonas de manejo mais fáceis de serem gerenciadas, sob o ponto de vista das operações de campo. A abordagem SCM (análise de matriz de correlação espacial) também proporcionou resultados satisfatórios, exceto no caso da definição de quatro classes para a área A2.

A abordagem MPCA-All (MULTISPATI-PCA executada a partir das variáveis originais) proporcionou melhor redução de dimensionalidade dos dados originais do que PCA-All (PCA executada a partir das variáveis originais), o que significa que com MPCA-All é possível representar melhor a variância de conjuntos de dados multivariados utilizando-se menos variáveis sintéticas. Como consequência, a execução de métodos de agrupamento a partir das variáveis definidas por esta abordagem tende a ser mais rápida e com resultados de melhor qualidade.

Portanto, mostrou-se que a abordagem MPCA-All, aplicada em conjunto com o algoritmo fuzzy c-means, pode melhorar consideravelmente a qualidade do processo e do resultado da definição de zonas de manejo.

REFERÊNCIAS

- ARROUAYS, D. et al. Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis. **Geoderma**, n. 161, p. 107-114, 2011.
- BAZZI, C. L. et al. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, v. 33, n. 5, p. 952-964, 2013.
- BIVAND, R. **spdep: Spatial Dependence: Weighting Schemes, Statistics and Models**. Vienna, Austria: R Foundation for Statistical Computing, 2012.
- BOYDELL, B.; MCBRATNEY, A. B. Identifying potential within-field management zones from cotton-yield estimates. **Precision Agriculture**, v. 3, n. 1, p. 9-23, 2002.
- CHELSEL, D.; DUFOUR, A. B.; THIOULOUSE, J. The ade4 package-I-one-table methods. **R News**, n. 4, p. 5-10, 2004.
- COHEN, S. et al. Combining spectral and spatial information from aerial hyperspectral images for delineating homogenous management zones. **Biosystems Engineering**, v. 114, n. 4, p. 435-443, 2013.
- CÓRDOBA, M. et al. Análisis de componentes principales com datos georreferenciados: Una aplicación en agricultura de precisión. **Revista de la Facultad de Ciencias Agrarias UNCUIYO**, v. 44, n. 1, p. 27-39, 2012.
- _____. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. **Computers and Electronics in Agriculture**, v. 97, p. 6-14, 2013.
- _____. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, n. 143, p. 95-107, 2016.
- DOERGE, T. A. **Site-specific management guidelines**. Norcross: Potash & Phosphate Institute, 2000. 135p.
- DRAY, S.; SAÏD, S.; DÉBIAS, F. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. **Journal of Vegetation Science**, v. 19, n. 1, p. 45-56, 2008.
- EMBRAPA. **Sistema brasileiro de classificação de solo**. Rio de Janeiro: CNPSO, 2006. 412p.
- FRAISSE, C. W.; SUDDUTH, K. A.; KITCHEN, N. R. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical

conductivity. **International Journal of the American Society of Agricultural and Biological Engineers**, v. 44, n. 1, p. 155-166, 2001.

FRIDGEN, J. J. et al. Management Zone Analyst (MZA): Software for subfield management zone delineation. **Agronomy Journal**, v. 96, p. 100-108, 2004.

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

HORNUNG, A. et al. Comparison of Site-Specific Management Zones: Soil-Color-Based and Yield-Based. **Agronomy Journal**, v. 98, n. 1, p. 407-415, 2006.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of educational psychology**, v. 24, n. 6, p. 417-441, 1933.

JAYNES, D. B.; COLVIN, T. S.; KASPAR, T. C. Identifying potential soybean management zones from multi-year yield data. **Computers and Electronics in Agriculture**, v. 46, n. 1, p. 309-327, 2005.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6. New Jersey: Pearson, 2007. 800p.

JOLLIFFE, I. T. **Principal Component Analysis**. 2. New York: Springer, 2002. 488p.

LARSCHEID, G.; BLACKMORE, B. S. Interactions between farm managers and information systems with respect to yield mapping. International Conference on Precision Agriculture, 1996, Minneapolis. American Society of Agronomy. p. 1153-1163.

LI, Y. et al. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, v. 56, n. 2, p. 174-186, 2007.

_____. Definition of management zones for enhancing cultivated land conservation using combined spatial data. **Environmental Management**, v. 52, n. 1, p. 792-806, 2013.

MIELKE JR, P. W.; BERRY, K. J. **Permutation methods: a distance function approach**. New York: Springer, 2007. 446p.

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

ORD, J. K. Estimation methods for models of spatial interaction. **Journal of the American Statistical Association**, v. 70, n. 349, p. 120-126, 1975.

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

_____. Delimitación de zonas de manejo con modelos de elevación digital y profundidad de suelo. **Interciencia**, v. 38, n. 6, p. 418-424, 2013.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2014.

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

SCHENATTO, K. et al. Data interpolation in the definition of management zones. **Acta Scientiarum**, v. 38, n. 1, p. 31-40, 2016.

SCHEPERS, A. R. et al. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal**, v. 96, p. 195-203, 2004.

ZHANG, Z. et al. Defining agricultural management zones using GIS techniques: Case study of drip-irrigated cotton fields. **Information Technology Journal**, v. 12, n. 1, p. 6241-6246, 2013.