

## IMPLEMENTAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS E AGRUPAMENTO DE DADOS PARA GERAÇÃO DE ZONAS DE MANEJO

ALAN GAVIOLI<sup>1</sup>, EDUARDO G. DE SOUZA<sup>2</sup>, CLAUDIO L. BAZZI<sup>3</sup>,  
KELYN SCHENATTO<sup>4</sup>, NELSON M. BETZEK<sup>5</sup>

<sup>1</sup> Doutor em Eng. Agrícola, Professor do Departamento de Computação da Universidade Tecnológica Federal do Paraná (UTFPR) – campus Medianeira / PR, Fone: (45) 3240-8000, alan@utfpr.edu.br.

<sup>2</sup> Doutor em Eng. Mecânica, Professor do Programa de Pós-Graduação em Eng. Agrícola da UNIOESTE – Cascavel / PR.

<sup>3</sup> Doutor em Eng. Agrícola, Professor do Departamento de Computação da UTFPR – campus Medianeira / PR.

<sup>4</sup> Doutora em Eng. Agrícola, Professora do Departamento de Computação da UTFPR – campus Santa Helena / PR.

<sup>5</sup> Doutorando em Eng. Agrícola, Professor do Departamento de Computação da UTFPR – campus Medianeira / PR.

Apresentado no  
XLVI Congresso Brasileiro de Engenharia Agrícola - CONBEA 2017  
30 de julho a 03 de agosto de 2017 - Maceió - AL, Brasil

**RESUMO:** Duas atividades fundamentais para a definição de zonas de manejo (ZMs) são a seleção das variáveis a serem utilizadas e a análise de agrupamento de dados. Existem diversos métodos para executá-las, que precisam ser disponibilizados por software. O objetivo deste trabalho foi apresentar dois módulos computacionais desenvolvidos para a execução eficiente dessas atividades. O módulo de seleção de variáveis disponibiliza 5 algoritmos baseados em análise da correlação espacial, análise de componentes principais (ACP) e no método denominado MULTISPATI-PCA. Já o módulo de geração de classes disponibiliza 17 algoritmos de agrupamento de dados: average linkage, bagged clustering, centroid linkage, clustering large applications, complete linkage, fuzzy analysis clustering, fuzzy c-means, hard competitive learning, hybrid hierarchical clustering, k-means, median linkage, método de McQuitty, método de Ward, neural gas, partitioning around medoids, spherical k-means e unsupervised fuzzy competitive learning. Para exemplificar o funcionamento desses módulos, foram empregados dados obtidos entre os anos de 2012 e 2015 de uma área agrícola localizada no município de Céu Azul - Paraná, na qual cultivou-se soja. Os módulos mostraram-se eficientes para a definição de ZMs. Além disso, são mais abrangentes que outros softwares de uso gratuito, como FuzME, MZA e SDUM, no que tange à diversidade de algoritmos de seleção de variáveis e de agrupamento de dados disponibilizados.

**PALAVRAS-CHAVE:** agricultura de precisão, análise de componentes principais, software para agricultura.

### IMPLEMENTATION OF VARIABLE SELECTION AND DATA CLUSTERING METHODS FOR GENERATION OF MANAGEMENT ZONES

**ABSTRACT:** Two basic activities for the definition of management zones (MZs) are the variable selection task and the cluster analysis task. There are several methods to execute them, which need to be made available by software. In this context, the objective of this study was to present two computational modules to enable the efficient execution of these activities. The variable selection module provides 5 algorithms based on spatial correlation analysis, principal component analysis (PCA) and the PCA-based method called MULTISPATI-PCA. The class generation module provides 17 data clustering algorithms, as follows: average linkage, bagged clustering, centroid linkage, clustering large applications, complete linkage, fuzzy analysis clustering, fuzzy c-means, hard competitive learning, hybrid hierarchical clustering, k-means, McQuitty's method, median linkage, neural gas, partitioning around medoids, spherical k-means, unsupervised fuzzy competitive learning, and Ward's method.

To exemplify the execution of these modules, data were obtained between 2012 and 2015 in an agricultural area in the municipality of Céu Azul - Paraná, where soybean was cultivated. The modules proved to be efficient to define MZs. Furthermore, they were more complete than other free-to-use software such as FuzME, MZA, and SDUM, in terms of the diversity of variable selection and data clustering methods.

**KEYWORDS:** precision agriculture, principal component analysis, software for agriculture.

## INTRODUÇÃO

Uma alternativa para tornar a agricultura de precisão (AP) economicamente mais atrativa para pequenos produtores é a implantação de zonas de manejo (ZMs) nas áreas agrícolas. Segundo Chang et al. (2014), cada ZM é uma subárea com características similares, geralmente limitantes da produtividade, para a qual pode-se adotar a aplicação a taxa fixa de insumos. Assim, cada ZM pode ser tratada como uma área homogênea para amostragem e gerenciamento, possibilitando reduzir a variabilidade espacial da produtividade das culturas e os danos ao meio ambiente provocados pela aplicação excessiva de determinados insumos.

Dentre os métodos mais utilizados para a geração de ZMs, estão os algoritmos de agrupamento de dados, que têm o propósito de dividir os pontos georreferenciados de uma área em certo número de classes. Na prática, essas classes ou grupos são empregados para delimitar as ZMs no campo. Além dos algoritmos k-means (MACQUEEN, 1967) e fuzzy c-means (BEZDEK, 1981), outros métodos também podem ser aplicados (DOBERMANN et al., 2003; GUASTAFERRO et al., 2010). Eles podem usar muitas variáveis na geração de ZMs, que denotem condições do solo, do relevo e/ou de plantas cultivadas. Entretanto, selecionar as variáveis necessárias é uma tarefa básica para assegurar que os algoritmos de agrupamento possam definir ZMs adequadas. Cohen et al. (2013) utilizaram análise de componentes principais (ACP) (HOTELLING, 1933) para selecionar variáveis para a geração de ZMs. Córdoba et al. (2013) e Peralta et al. (2015) aplicaram a análise multivariada espacial baseada no índice de Moran e em ACP (MULTISPATI-PCA) (DRAY; SAID; DÉBIAS, 2008). Bazzi et al. (2013) e Schenatto et al. (2016) empregaram uma abordagem fundamentada na análise de correlação espacial entre variáveis (REICH; CZAPLEWSKI; BECHTOLD, 1994). Já Gavioli et al. (2016) avaliaram 5 métodos, incluindo um novo algoritmo baseado em MULTISPATI-PCA e análise de correlação espacial.

A seleção de variáveis e o agrupamento de dados são atividades de execução complexa, que dependem do uso de software. Porém, os principais sistemas para definição de ZMs disponibilizam poucos métodos para essas duas atividades. É o caso de FuzME (MINASNY; MCBRATNEY, 2002), Management Zone Analyst (MZA) (FRIDGEN et al., 2004) e do Software para Definição de Unidades de Manejo (SDUM) (BAZZI et al., 2013). Neste contexto, o objetivo deste trabalho é apresentar dois módulos computacionais desenvolvidos para possibilitar a seleção de variáveis e a geração de classes para o delineamento de ZMs. Como diferencial, esses módulos disponibilizam 5 métodos de seleção de variáveis baseados em ACP, MULTISPATI-PCA e análise de correlação espacial, e 17 métodos de agrupamento.

## MATERIAL E MÉTODOS

Para o desenvolvimento dos dois módulos computacionais, empregaram-se softwares para a programação dos algoritmos de seleção de variáveis e de agrupamento de dados, e para a construção da interface gráfica. Todos os softwares puderam ser utilizados de forma gratuita. Os 5 algoritmos de seleção de variáveis e os 17 algoritmos de agrupamento foram programados em rotinas do software estatístico R (R CORE TEAM, 2016). Na Figura 1, representam-se atividades do processo de geração de ZMs, com destaque para os métodos incluídos nos módulos de seleção de variáveis e de agrupamento.

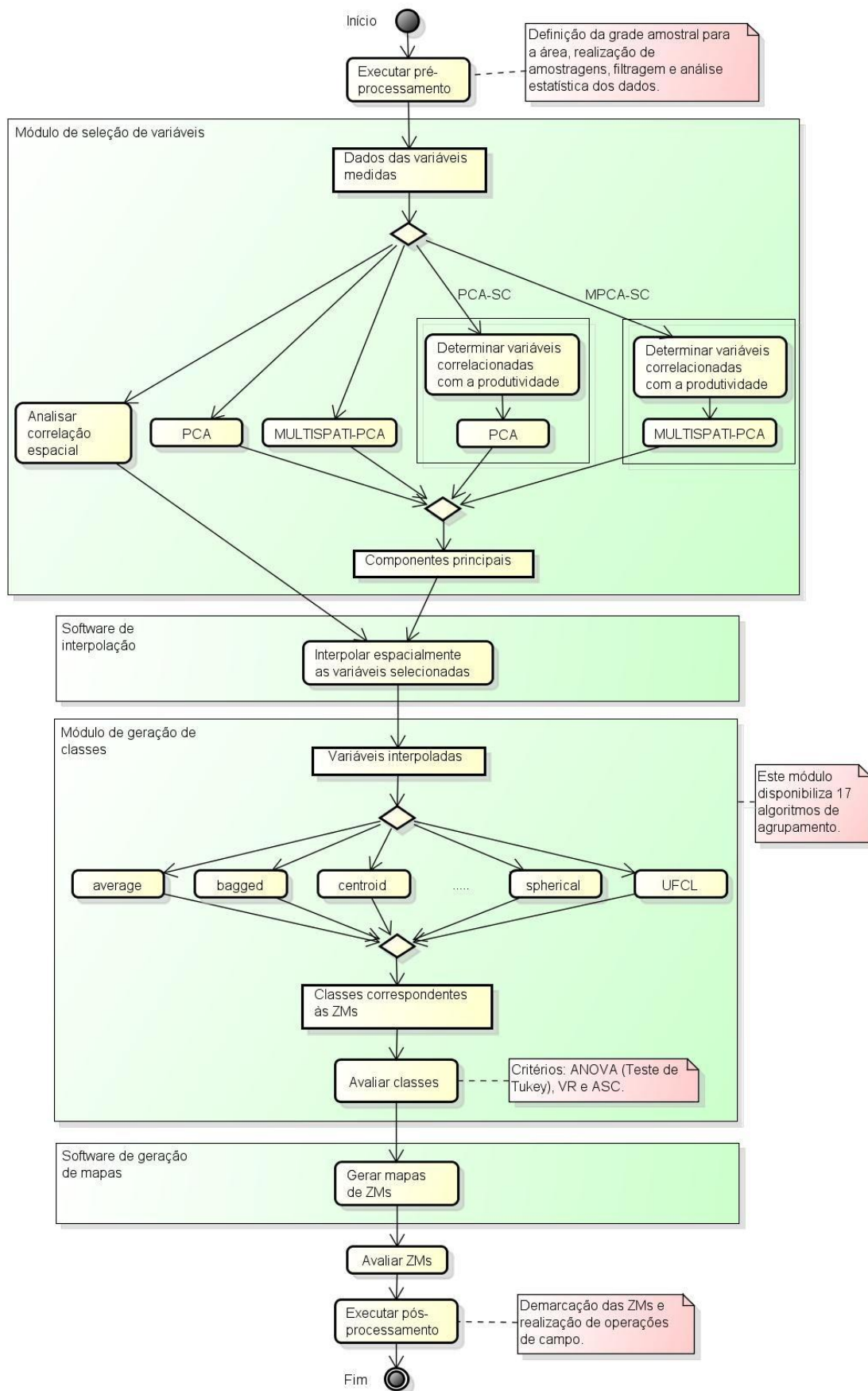


FIGURA 1. Atividades para a definição de zonas de manejo, com destaque para os módulos desenvolvidos para seleção de variáveis e geração de classes.

Para utilizar o módulo de seleção de variáveis, deve-se ter os dados georreferenciados correspondentes às variáveis do solo e/ou das plantas cultivadas de uma área. Esses dados devem estar armazenados em um arquivo de texto. Assim, podem-se executar as rotinas que constituem esse módulo para realizar a seleção das variáveis para a definição de ZMs. As rotinas implementadas correspondem aos 5 métodos comparados por Gavioli et al. (2016):

- PCA-All: executa a ACP sobre as variáveis originais disponíveis, para gerar as novas variáveis sintéticas, isto é, as componentes principais (CPs);
- MPCA-All: aplica a abordagem MULTISPATI-PCA sobre as variáveis originais disponíveis, a fim de definir as CPs correspondentes – que para MULTISPATI-PCA também são chamadas de CPs espaciais (CPEs);
- PCA-SC: verifica quais variáveis originais apresentam correlação espacial significativa com a produtividade da área, a um dado nível de significância; em seguida, sobre as variáveis que satisfizerem essa condição, aplica a ACP para gerar as CPs;
- MPCA-SC: executa-se o mesmo procedimento descrito para PCA-SC, mas substituindo a aplicação de ACP por MULTISPATI-PCA, para gerar as CPEs;
- Análise de correlação espacial: calcula a correlação espacial bivariada de Moran (REICH; CZAPLEWSKI; BECHTOLD, 1994) entre todas as variáveis; em seguida, seleciona variáveis pelo procedimento apresentado por Bazzi et al. (2013).

O resultado da execução das rotinas correspondentes aos métodos de seleção de variáveis é gravado em um arquivo de texto. Para executar a interpolação espacial das variáveis selecionadas, deve-se utilizar um software que disponibilize algoritmos de interpolação recomendados para a criação de ZMs, como, por exemplo, krigagem ordinária (aplicada neste trabalho). Para utilizar o módulo de geração de classes para ZMs, deve-se ter em um arquivo de texto dados georreferenciados resultantes da interpolação das variáveis ou CPs. Assim, pode-se executar as rotinas desse módulo para dividir os dados interpolados em certa quantidade de classes. As rotinas implementadas correspondem aos 17 métodos de agrupamento informados na Tabela 1. O resultado da execução de qualquer método de agrupamento também é gravado em um arquivo de texto.

TABELA 1. Métodos de agrupamento disponibilizados no módulo de geração de classes.

Métodos de agrupamento	Referências
average linkage	Jain e Dubes (1988)
bagged clustering	Leisch (1999)
centroid linkage	Jain e Dubes (1988)
complete linkage	Jain e Dubes (1988)
clustering large applications	Kaufman e Rousseeuw (1990)
fuzzy analysis clustering	Kaufman e Rousseeuw (1990)
fuzzy c-means	Bezdek (1981)
hard competitive learning	Jain e Dubes (1988)
hybrid hierarchical clustering	Chipman e Tibshirani (2006)
k-means	MacQueen (1967)
median linkage	Jain e Dubes (1988)
método de McQuitty	McQuitty (1966)
método de Ward	Ward (1963)
neural gas	Martinetz, Berkovich e Schulten (1993)
partitioning around medoids	Kaufman e Rousseeuw (1990)
spherical k-means	Dhillon e Modha (2001)
unsupervised fuzzy competitive learning	Pal, Bezdek e Hathaway (1996)

Após a execução do módulo de geração de classes, os dados referentes à classificação dos pontos podem ser utilizados para a geração do mapa temático das ZMs. Para isso, deve-se utilizar um software como, por exemplo, o SDUM, que foi empregado neste trabalho. Para avaliar o desempenho dos métodos de agrupamento em relação à qualidade das classes geradas, utilizaram-se a ANOVA (teste de Tukey), o índice de redução da variância (VR) (LI et al., 2007) e o coeficiente de silhueta médio (ASC) (KAUFMAN; ROUSSEEUW, 1990).

Para exemplificar o funcionamento dos dois módulos, utilizaram-se dados coletados entre 2012 e 2015 de uma área agrícola de 15,5 ha, localizada no município de Céu Azul – Paraná, onde cultivou-se soja. Para essa área, gerou-se uma grade amostral irregular com 40 pontos (2,67 pontos ha<sup>-1</sup>), de modo que os pontos ficassem posicionados na linha central imaginária entre as curvas de nível. Seguindo recomendação de Doerge (2000), no estudo de caso empregaram-se somente variáveis consideradas temporalmente estáveis para a geração de ZMs: altitude, declividade, densidade, textura do solo (areia, argila e silte), e resistência do solo à penetração (RSP) para as profundidades de 0 – 0,1 m, 0,1 – 0,2 m e 0,2 – 0,3 m. Para validação das ZMs definidas, utilizaram-se dados da produtividade de soja do período de 2012 a 2015, determinados para os mesmos pontos onde as outras variáveis foram avaliadas.

## RESULTADOS E DISCUSSÃO

A tela inicial projetada para o módulo de seleção de variáveis (Figura 2) apresenta, na parte superior, um campo para seleção da área agrícola de interesse. Após a escolha da área, o arquivo de texto com os dados das variáveis correspondentes a ela será automaticamente lido pelo software, para que os nomes dessas variáveis sejam exibidos na parte inferior da tela.

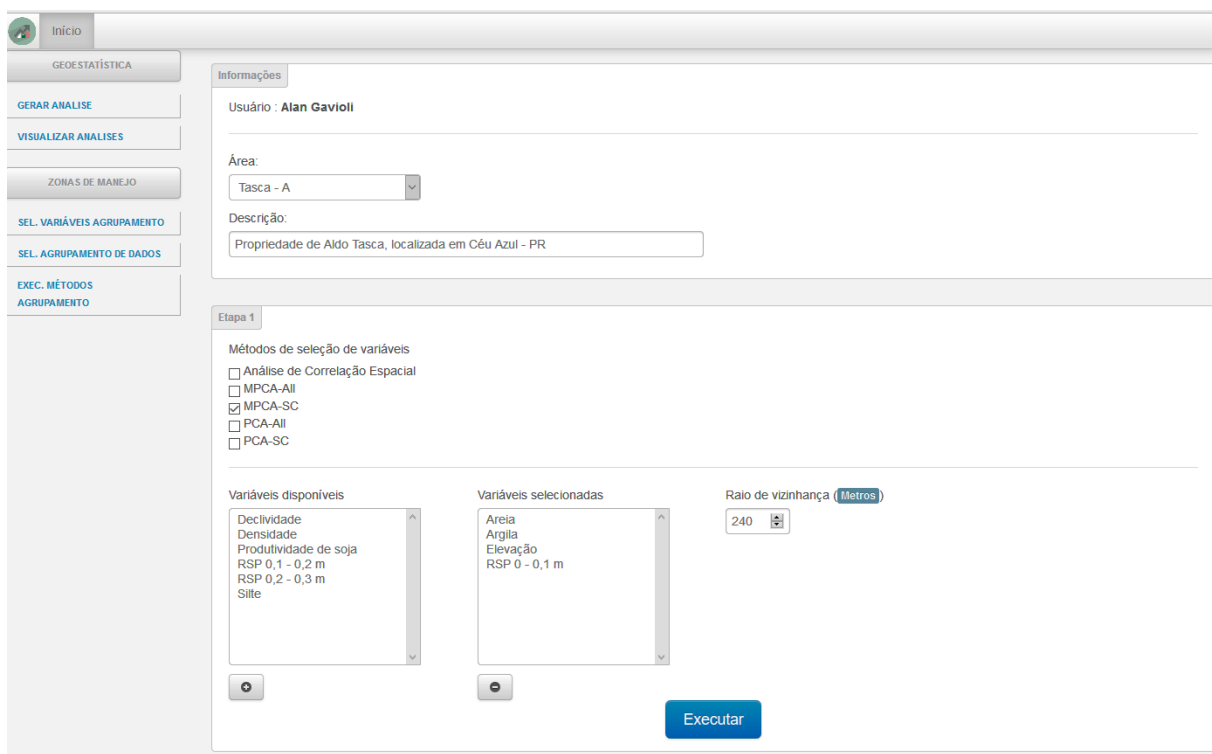


FIGURA 2. Tela inicial do módulo de seleção de variáveis, com os dados do estudo de caso considerado: nesse exemplo, optou-se pelo método MPCA-SC.

Assim, pode-se escolher nessa tela um dos 5 métodos de seleção de variáveis para ser utilizado. De acordo com essa escolha, o módulo automaticamente analisará as variáveis disponíveis e preencherá o campo correspondente às variáveis selecionadas, também na parte inferior da tela. No exemplo mostrado na Figura 2, a opção pelo método MPCA-SC fez com

que o módulo selecionasse quatro das dez variáveis disponíveis para serem empregadas na geração de CPs. Além disso, determinou-se experimentalmente que 240 m era um valor satisfatório para o raio de vizinhança (exigido pelos métodos MPCA-SC e MPCA-All).

Após a execução dos métodos MPCA-All, MPCA-SC, PCA-All e PCA-SC, as CPs geradas serão apresentadas na parte inferior da tela de resultados (Figura 3), com os respectivos valores do percentual de representação da variância original dos dados e do percentual acumulado dessa variância. Em relação ao exemplo mostrado na Figura 3, ao encerrar a execução de MPCA-SC, as coordenadas dos pontos amostrais da área e os respectivos escores das componentes CPE1 e CPE2 foram gravados em um arquivo de texto.

**Etapa 1**

Métodos de seleção de variáveis

Análise de Correlação Espacial

MPCA-All

MPCA-SC

PCA-All

PCA-SC

---

Variáveis disponíveis

- Declividade
- Densidade
- Produtividade de soja
- RSP 0,1 - 0,2 m
- RSP 0,2 - 0,3 m
- Silte

Variáveis selecionadas

- Areia
- Argila
- Elevação
- RSP 0 - 0,1 m

Raio de vizinhança (Metros): 240

**Executar**

---

**Etapa 2**

MPCA-SC	Componente Principal	Percentual da Variância Original	Percentual Acumulado da Variância
<input type="checkbox"/>	CPE1	71	71
<input type="checkbox"/>	CPE2	29	100

**Gravar CPs**

FIGURA 3. Tela com os resultados da execução dos métodos de seleção de variáveis baseados na criação de componentes principais.

O módulo de seleção de variáveis também permite selecionar simultaneamente dois ou mais métodos de geração de CPs, para serem executados com os mesmos dados de entrada. Para isso, deve-se selecionar, na tela inicial desse módulo (Figura 2), os métodos que deverão ser comparados. Neste caso, para facilitar a interpretação dos resultados, os métodos executados são exibidos pelo software em ordem decrescente de desempenho (Figura 4). O critério de comparação empregado é o maior percentual possível de representação da variância original dos dados nas primeiras CPs. Assim, no exemplo mostrado na Figura 4, MPCA-SC obteve o melhor desempenho.

A tela inicial projetada para o módulo de geração de classes (Figura 5) exhibe, na parte superior, um campo para seleção da área agrícola de interesse. Após a escolha da área, os arquivos de texto com os dados interpolados das variáveis correspondentes a ela serão automaticamente lidos pelo software, para que os nomes dessas variáveis sejam exibidos na parte inferior da tela. Deste modo, será possível selecionar um dos 17 algoritmos de agrupamento e as variáveis de interesse para que sejam utilizados. Também será necessário informar as quantidades mínima e máxima de ZMs desejadas, além de possíveis parâmetros

específicos do método de agrupamento selecionado. Para facilitar o uso, os parâmetros específicos dos métodos serão automaticamente preenchidos com valores recomendados.

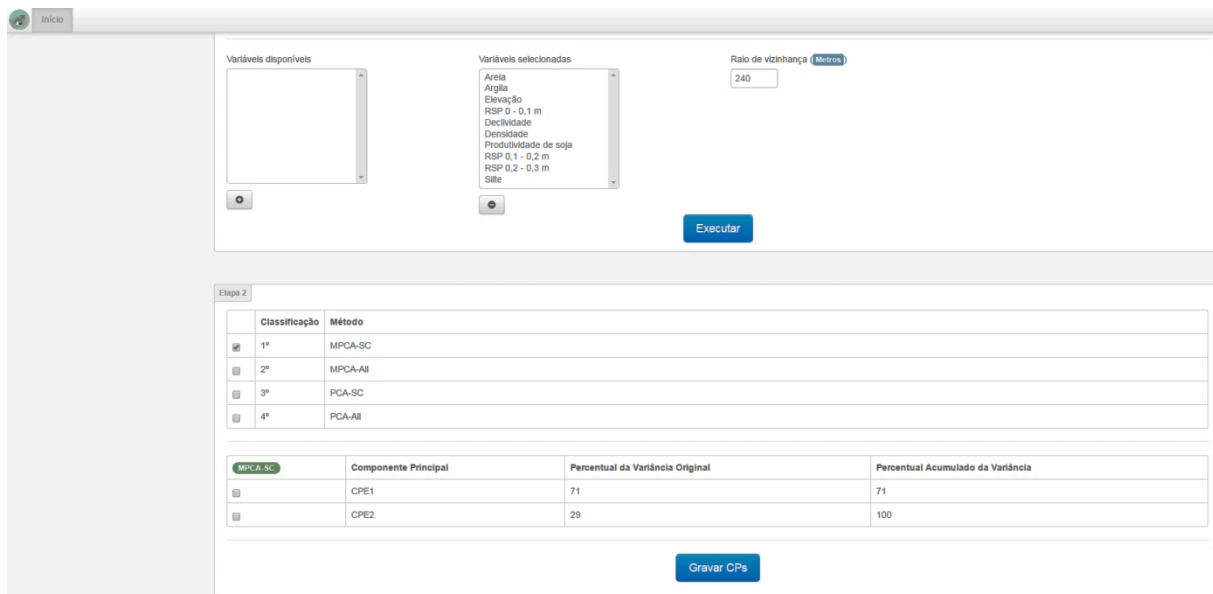


FIGURA 4. Tela com resultado da comparação de métodos de seleção baseados em componentes principais, com destaque para o desempenho de MPCA-SC.

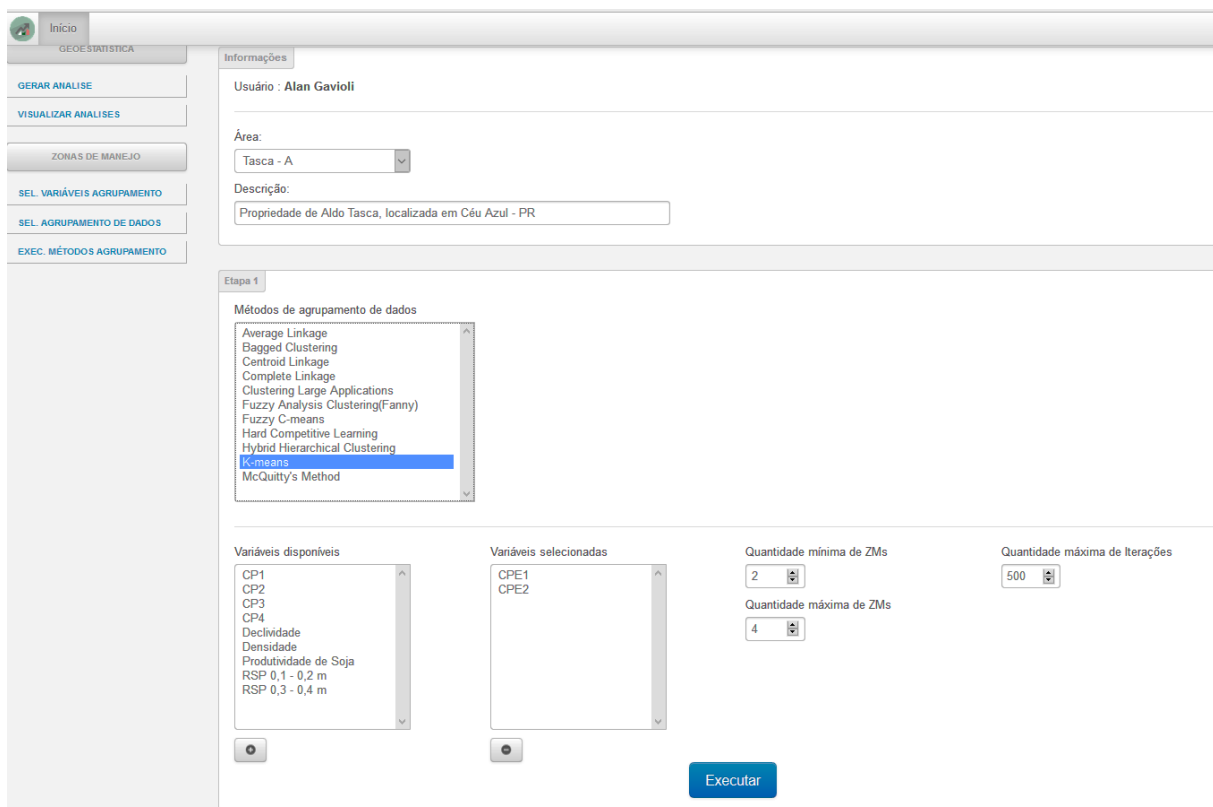


FIGURA 5. Tela inicial do módulo de geração de classes, mostrando a escolha do algoritmo K-means e das variáveis CPE1 e CPE2 para gerar 2, 3 e 4 zonas de manejo.

No exemplo mostrado na Figura 5, selecionou-se o algoritmo K-means para gerar duas, três e quatro classes, a partir dos valores resultantes da interpolação por krigagem ordinária das componentes CPE1 e CPE2. Para K-means, foi necessário informar a quantidade máxima

de iterações que podem ocorrer durante a sua execução. A geração de diversas quantidades de classes, empregando o mesmo algoritmo de agrupamento com as mesmas variáveis de entrada, é importante quando não se sabe, a priori, em quantas ZMs uma área deve ser dividida para se obter os melhores resultados. Na tela de resultados (Figura 6), exibe-se o desempenho do algoritmo para cada quantidade de classes. Para facilitar a visualização dos resultados, estes são exibidos em ordem crescente de quantidade de classes. No exemplo da Figura 6, o resultado do teste de Tukey mostrou que é possível dividir a área somente em duas ZMs com potenciais produtivos estatisticamente distintos, ao nível de 5% de significância. Essa divisão promoveu redução satisfatória da variância da produtividade, já que VR = 33,8%, com subáreas que apresentaram elevada homogeneidade interna (ASC = 0,59).

Desempenho do método selecionado:			
Quantidade de classes	ANOVA(Teste de Tukey)	VR	ASC
2	a b	33,8	0,59
3	a b a	23,8	0,46
4	a a b b	35,8	0,39

FIGURA 6. Tela com os resultados da avaliação do método de agrupamento selecionado: valores do teste de Tukey, do índice VR e do coeficiente ASC.

Ao concluir a execução de K-means para esse estudo de caso, as coordenadas geográficas dos pontos e os respectivos números identificadores das classes associadas a esses pontos foram gravados em um arquivo de texto. Com essa rotina, foram criados três arquivos de resposta, correspondentes à geração de duas, três e quatro classes. Utilizando o software SDUM, leram-se dos três arquivos de texto as coordenadas dos pontos e suas respectivas classes, para definirem-se os mapas temáticos com duas, três e quatro ZMs (Figura 7).

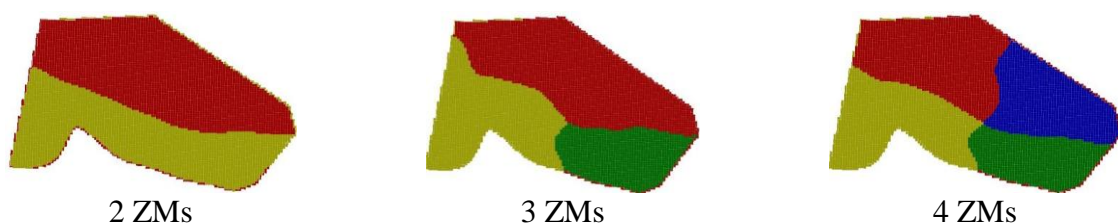


FIGURA 7. Mapas temáticos com 2, 3 e 4 zonas, correspondentes ao estudo de caso.

Existem outros softwares que possibilitam aplicar os métodos de seleção de variáveis baseados em ACP, como ArcGIS (Esri Software), Matlab (The MathWorks), SAS (SAS Institute), Scilab (Scilab Enterprises) e Statistica (Statsoft). No entanto, quase todos exigem o pagamento de algum valor para serem utilizados. Além disso, são de propósito mais geral que o módulo de seleção desenvolvido, isto é, nenhum deles foi projetado especificamente para a definição de ZMs. Por isso, requerem mais tempo de aprendizado para serem aplicados em uma tarefa que pode ser feita com simplicidade no módulo apresentado. Já para a execução dos métodos baseados em MULTISPATI-PCA, tem-se apenas o software R disponível.

Comparando os dois módulos desenvolvidos aos softwares FuzME e MZA, notou-se que estes dois não disponibilizam métodos para seleção das variáveis. Em relação a algoritmos de agrupamento, FuzME e MZA disponibilizam apenas o método fuzzy c-means. Já ao serem comparados os dois módulos e o software SDUM, este viabiliza a seleção de variáveis apenas por meio da análise de correlação espacial e disponibiliza dois métodos de agrupamento: fuzzy c-means e k-means.

## CONCLUSÕES

Os módulos computacionais desenvolvidos proporcionam flexibilidade na seleção de variáveis para a geração de classes, por meio do uso dos 5 algoritmos de seleção disponibilizados, assim como no agrupamento de dados que efetivamente define as ZMs, por meio da aplicação dos 17 métodos implementados. Esses módulos são mais abrangentes que outros softwares de uso gratuito, como FuzME, MZA e SDUM, em relação à diversidade de algoritmos para seleção de variáveis e agrupamento de dados disponibilizados.

## REFERÊNCIAS

- BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, v. 33, n. 5, p. 952-964, 2013.
- BEZDEK, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. New York: Plenum Press, 1981. 256 p.
- CHANG, D.; ZHANG, J.; ZHU, L.; GE, S. H.; LI, P. Y.; LIU, G. S. Delineation of management zones using an active canopy sensor for a tobacco field. **Computers and Electronics in Agriculture**, v. 109, p. 172-178, 2014.
- CHIPMAN, H.; TIBSHIRANI, R. Hybrid Hierarchical Clustering with Applications to Microarray Data. **Biostatistics**, v. 7, p. 302-317, 2006.
- COHEN, S.; COHEN, Y.; ALCHANATIS, V.; LEVI, O. 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.; BRUNO, C.; COSTA, J. L.; BALZARINI, M. 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.
- DHILLON, I. S.; MODHA, D. S. Concept decompositions for large sparse text data using clustering. **Machine Learning**, v. 42, p. 143-175, 2001.
- DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of crop yield variability in irrigated production fields. **Agronomy Journal**, v. 95, n. 1, p. 1105-1120, 2003.
- DOERGE, T. A. **Site-specific management guidelines**. Norcross: Potash & Phosphate Institute, 2000. 135 p.
- DRAY, S.; SAID, 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.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management zone analyst (MZA): software for subfield management zone delineation. **Agronomy Journal**, v. 96, p. 100-108, 2004.

GAVIOLI, A.; SOUZA, E. G.; BAZZI, C. L.; GUEDES, L. P. C.; SCHENATTO, K. Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, v. 127, p. 302-310, 2016.

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

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

JAIN, A. K.; DUBES, R. **Algorithms for clustering data**. Englewood Cliffs: Prentice-Hall, 1988. 320 p.

KAUFMAN, L.; ROUSSEEUW, P. J. **Finding groups in data**. Hoboken: John Wiley & Sons, 1990. 342 p.

LEISCH, F. Bagged clustering. In: SFB ADAPTIVE INFORMATION SYSTEMS AND MODELLING IN ECONOMICS AND MANAGEMENT SCIENCE, 51, 1999, Vienna. **Anais...** Vienna: Vienna University of Economics and Business, 1999. p. 1-11.

LI, X.; PAN, Y.; GE, Z.; ZHAO, C. Delineation and scale effect of precision agriculture management zones using yield monitor data over four years. **Agricultural Sciences in China**, v. 6, n. 2, p. 180-188, 2007.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: PROCEEDINGS OF 5<sup>TH</sup> BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 1967, Berkeley. **Anais...** Berkeley: University of California Press, 1967. p. 281-297.

MARTINETZ, T. M.; BERKOVICH, S. G.; SCHULTEN, K. J. "Neural-gas" network for vector quantization and its application to time-series prediction. **IEEE Transactions on Neural Networks**, v. 4, n. 4, p. 558-569, 1993.

MCQUITTY, L. L. Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. **Educational and Psychological Measurement**, v. 26, p. 825-831, 1966.

MINASNY, B.; MCBRATNEY, A. B. **FuzME 3.0**. Australian Centre for Precision Agriculture. The University of Sydney. Sydney. 2002.

PAL, N. R.; BEZDEK, J. C.; HATHAWAY, R. J. Sequential competitive learning and the fuzzy c-means clustering algorithm. **Neural Networks**, v. 9, n. 5, p. 787-796, 1996.

PERALTA, N. R.; COSTA, J. L.; BALZARINI, M.; FRANCO, M. C.; CÓRDOBA, M.; BULLOCK, D. Delineation of management zones to improve nitrogen management of wheat. **Computers and Electronics in Agriculture**, v. 110, p. 103-113, 2015.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2016. 99 p.

REICH, R. M.; CZAPLEWSKI, R. L.; BECHTOLD, W. A. Spatial cross-correlation of undisturbed, natural shortleaf pine stands in northern Georgia. **Environmental and Ecological Statistics**, v. 1, p. 201-217, 1994.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BIER, V. A.; BETZEK, N. M.; GAVIOLI, A. Data interpolation in the definition of management zones. **Acta Scientiarum**, v. 38, n. 1, p. 31-40, 2016.

WARD, J. H. Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, v. 58, n. 301, p. 236-244, 1963.