



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Cerrados
Ministério da Agricultura, Pecuária e Abastecimento*

ISSN 1676-918X

Maio, 2005

Boletim de Pesquisa e Desenvolvimento 150

Classificação não- supervisionada no Delineamento de Zonas de Manejo

Luciano Shozo Shiratsuchi
Leonardo Ribeiro Queiros
Gustavo Camargo Faccioni

Planaltina, DF
2005

Exemplares desta publicação podem ser adquiridos na:

Embrapa Cerrados

BR 020, Km 18, Rod. Brasília/Fortaleza

Caixa Postal 08223

CEP 73310-970 Planaltina - DF

Fone: (61) 3388-9898

Fax: (61) 3388-9879

<http://www.cpac.embrapa.br>

sac@cpac.embrapa.br

Comitê de Publicações

Presidente: *José de Ribamar N. dos Anjos*

Secretária-Executiva: *Maria Edilva Nogueira*

Supervisão editorial: *Maria Helena Gonçalves Teixeira*

Revisão de texto: *Maria Helena Gonçalves Teixeira*

Normalização bibliográfica: *Shirley da Luz Soares*

Marilaine Schaun Pelufé

Capa: *Chaile Cherne Soares Evangelista*

Editoração eletrônica: *Leila Sandra Gomes Alencar*

Impressão e acabamento: *Divino Batista de Souza*

Jaime Arbués Carneiro

Impresso no Serviço Gráfico da Embrapa Cerrados

1ª edição

1ª impressão (2005): tiragem 100 exemplares

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

CIP-Brasil. Catalogação na publicação.

Embrapa Cerrados.

S558c Shiratsuchi, Luciano Shozo.

Classificação não-supervisionada no delineamento de zonas de manejo / Luciano Shozo Shiratsuchi, Leonardo Ribeiro Queiros, Gustavo Camargo Faccioni. – Planaltina, DF : Embrapa Cerrados, 2005.

16 p. — (Boletim de pesquisa e desenvolvimento / Embrapa Cerrados, ISSN 1676-918X ; 150)

1. Agricultura de precisão. 2. Mapeamento - manejo. I. Queiros, Leonardo Ribeiro. II. Faccioni, Gustavo Camargo. III. Título. IV. Série.

630.2018 - CDD 21

© Embrapa 2005

Sumário

Resumo	5
Abstract	6
Introdução	7
Material e Métodos	10
Descrição da área	10
Coleta e tratamento dos dados	10
Delineamento das zonas de manejo baseado na produtividade relativizada pela média	10
Delineamento automático das zonas de manejo	11
Resultados e Discussão	11
Conclusão	15
Referências	15

Classificação não-supervisionada no Delineamento de Zonas de Manejo¹

Luciano Shozo Shiratsuchi²

Leonardo Ribeiro Queiros³

Gustavo Camargo Faccioni⁴

Resumo – O objetivo deste trabalho foi utilizar o método de classificação não-supervisionada (*fuzzy k-means*) em dados de produtividade para delineamento de zonas de manejo. O estudo foi conduzido a partir de dados georreferenciados de produtividade de grãos coletados por sensores e DGPS. Foram mapeadas três safras (milho – soja – arroz) em área de plantio direto irrigada por pivô central. Os dados de produtividade foram interpolados pelo inverso da distância resultando em um mapa com resolução espacial de 5 x 5 m. Os valores relativizados em relação à média de produtividade foram comparados ao longo do tempo nos mesmos locais. Como referência, foram escolhidas três classes de produtividade para delinear as zonas de manejo segundo os seguintes critérios: produtividade alta estável (produtividade 10% acima da média e com coeficiente de variação (CV) menor que 30% ao longo das três safras); produtividade baixa estável (produtividade 10% abaixo da média e CV < 30%); produtividade Média Estável (produtividade média e CV < 30%) e produtividade inconsistente (produtividade com CV% > 30%). Para o delineamento automático de zonas de manejo, foi utilizado o método *fuzzy k-means* usando três classes. Obteve-se correlação de 0,98** entre a definição de zonas de manejo por meio de classificação automática pelo método da estabilidade temporal da produtividade em relação à média.

Termos para indexação: agricultura de precisão, mapas de produtividade, *fuzzy kmeans*.

¹ Atividade componente do projeto da Embrapa 03.04.0.01.60: Áreas-piloto de Agricultura de Precisão em Sistemas Agrícolas do Cerrado, financiado pelo Clube de Plantio Direto do Oeste Baiano - CPDOB.

² Eng. Agrôn., M.Sc., Embrapa Cerrados, shozo@cpac.embrapa.br

³ Mestrando em Engenharia da Computação, Universidade Federal de Goiás, Goiânia - GO, rqlleonardo@gmail.com

⁴ Bolsista (Graduação), Estudante de Agronomia, União Pioneira de Integração Social – UPIS, Brasília - DF, gusfaccioni@yahoo.com.br

Delineating Management Zones Using *Fuzzy K-means* Unsupervised Clustering Algorithm

Abstract - *The objective of this work was to delineate yield management zones using fuzzy k-means unsupervised clustering algorithm in yield data. The study was done during 3 years using yield data collected with yield monitor and DGPS. The crop system was corn-soybean-rice successively in area of no-till system irrigated in central pivot at Cerrado's Region. The yield maps were interpolated by inverse distance weighting resulting in a map with space resolution of 5 x 5 m. The relativized yield values of each crop was compared along the time in the same places. They were chosen 3 yield classes to delineate the management zones with the following criteria: High yield stable (crop yield 10% above the yield average and with coefficient of variation (CV) smaller than 30% along the 3 years); Low Yield Stable (productivity 10% below the average and CV < 30%); Medium Yield Stable (average yield with CV < 30%) and inconsistent yield (yield with CV% > 30%). For automatic delineating of management zones with fuzzy k means was calculated using 3 classes. It was obtained a correlation of 0,98** between management zones definition through fuzzy logic and the method of temporal stability of crop yield.*

Index Terms: precision agriculture, yield mapping, fuzzy logic.

Introdução

Diversas pesquisas na linha de Agricultura de Precisão têm enfatizado esforços na amostragem em grades regulares para atributos da fertilidade do solo, plantas daninhas, pragas, compactação do solo, entre outros. O mapeamento desses atributos possibilita a geração de mapas de prescrição para intervenção com aplicação de insumos a taxas variáveis, diminuindo o custo e o impacto ambiental em relação ao manejo tradicional baseado numa dose média ([MULLA et al., 1992](#)).

O mapeamento de cada um dos diversos fatores que interferem na produtividade é de difícil execução e ainda não existem critérios técnicos bem definidos para o correto mapeamento e para a tomada de decisão considerando a interação dos vários fatores. Por sua vez, uma tecnologia de fácil implementação é o mapeamento da produtividade de grãos por meio de sensores acoplados nas colhedoras automotrizes. Esse processo é confiável e já foi validado na prática, tanto que diversos monitores de produtividade já estão disponíveis no mercado, máquinas já saem de fábrica equipadas, e a pesquisa já mostrou a viabilidade destes na quantificação da produtividade das culturas e na confecção de mapas de produtividade ([BIRRELL et al., 1996](#); [DOERGE, 1996](#); [GRISSE et al., 2002](#)).

Em trabalhos com mapeamento da produtividade das culturas ao longo do tempo, percebe-se que as culturas em geral apresentam locais com produtividades estáveis temporalmente, ou seja, existem subáreas bastante produtivas ou de baixa produtividade ao longo dos anos ([FRAISSE et al., 2001](#); [MOLIN, 2002](#); [BROCK et al., 2005](#)). Essa estabilidade espaço-temporal pode ser devida a vários fatores, mas condições de solo e relevo preponderam ([SCHEPERS et al., 2004](#)). Regiões que, dentro do talhão, possuem características semelhantes, necessitando da mesma quantidade de insumo, resultam no mesmo potencial produtivo e são denominadas Zonas de Manejo ([SCHEPERS et al., 2004](#)). Diferentemente da abordagem em grades regulares, as Zonas de Manejo podem ser tratadas como subáreas dentro do talhão, nas quais, nas intervenções, leva-se em consideração a homogeneidade em termos de manejo. Essa abordagem pode facilitar a adoção de uma estratégia de ação sem a necessidade do emprego de máquinas e implementos muito sofisticados.

A partir do delineamento dessas áreas homogêneas de produtividade, fatores que interferem na produtividade delas podem ser investigados e corrigidos segundo critérios técnicos ([SCHEPERS et al., 2004](#)). Estudos já mostraram a viabilidade

dessa técnica. Entretanto, o delineamento de zonas de manejo baseado em mapas de produtividade ainda é feito empiricamente, dependendo das classes de alta ou baixa produtividade definidas pelo técnico. Dessa forma, a divisão arbitrária do talhão pode variar segundo critérios subjetivos ([FRAISSE et al., 2001](#)). Contudo, existem técnicas, como a lógica nebulosa (*fuzzy k-means*), que podem realizar o zoneamento automaticamente formando agrupamentos ou *clusters* de dados semelhantes, sem a subjetividade da divisão em classes escolhidas pelo técnico ([FRIDGEN et al., 2004](#)).

Um algoritmo de agrupamento (*clusterização*) tem o objetivo de particionar um conjunto de dados em *k-clusters* ([LUZ, 2004](#)), tendo como referência um centro de massa ou centróide para cada *cluster* formado. Esse processo foi implementado em vários programas comerciais de Sistemas de Informações Geográficas e recentemente melhorado para uso específico em Agricultura de Precisão ([FRIDGEN et al., 2004](#); [MANAGEMENT ZONE ANALYST 1.0, 2005](#)). A grande vantagem em relação aos delineamentos que vêm sendo feitos é que na escolha do valor das classes não existe o viés da escolha arbitrária feita pelo técnico.

Nesses algoritmos de classificação não-supervisionada, o usuário inicialmente escolhe o número de zonas de manejo que deseja trabalhar, o qual representa o grupo inicial e arbitrário de médias de valores de produtividade que são específicas para cada um dos *k* grupos preestabelecidos (zonas de manejo). Em seguida e de uma forma iterativa, todos os dados de produtividade são agrupados para a classe mais próxima, pelo critério de mínima distância ([VENTURIERI, 1996](#)). A letra *k* representa o número de agrupamentos que é o único parâmetro determinado pelo usuário. Um novo conjunto de médias é obtido para cada iteração, e os dados de produtividade são reclassificados de acordo com essas novas médias, a menos que um desvio-padrão, um limiar ou um número máximo de iterações sejam especificados pelo técnico.

Estatisticamente, o *fuzzy k-means* minimiza a soma dos quadrados dos erros dentro de cada classe seguindo os seguintes critérios ([Australian Centre for Precision Agriculture \(2005\)](#)):

$$\sum_{k=1}^k m_{ik} = 1 \quad i = 1, 2, \dots, n$$

$$\sum_{i=1}^n m_{ik} > 0 \quad k = 1, 2, \dots, c$$

$$m_{ik} \in \{0, 1\} \quad i = 1, 2, \dots, n; \quad k = 1, \dots, c \quad 1$$

É definido pela seguinte função:

$$J = \sum_{i=1}^{\pi} \sum_{k=1}^{\tau} m_{ik}^{\phi} d^2(x_i, c_k) \quad 2$$

Onde:

n são os dados; c é o número de classes escolhido; c_k é o vetor que representa o centróide da classe k ; x_i é o vetor que representa cada dado i ; e $d^2(x_i, c_k)$ é o quadrado da distância entre x_i e c_k seguindo a escolha da definição da distância que por simplicidade é denotada como d_{ik}^2 . ϕ é o expoente fuzzy que varia de $(1, \infty)$. Ele determina o grau de confusão ou sobreposição entre os diversos grupos ou *clusters* criados. Quando $\phi = 1$, a solução é bem dividida, ou seja, existem clusters bem definidos. Quando o $\phi = \infty$, a solução apresenta alto grau de confusão.

A minimização da função J possibilita a solução para a função definida por Bezdek (1981):

$$m_{ik} = \frac{d_{ik}^{2/(\phi-1)}}{\sum_{j=1}^{\tau} d_{ij}^{2/(\phi-1)}} \quad i = 1, 2, \dots, n; \quad k = 1, \dots, c \quad 3$$

$$c_k = \frac{\sum_{i=1}^n m_{ik}^{\phi} x_i}{\sum_{i=1}^n m_{ik}^{\phi}} \quad k = 1, 2, \dots, c \quad 4$$

Partindo do pressuposto que o mapeamento espaço-temporal da produtividade pode auxiliar no delineamento de zonas de manejo, o objetivo deste trabalho foi utilizar a lógica nebulosa (algoritmo de classificação não-supervisionada *fuzzy k-means*) em dados de produtividade para delineamento de zonas de manejo.

Material e Métodos

Descrição da área

O estudo foi conduzido no Município de Luís Eduardo Magalhães, BA, na Fazenda Maria das Águas Santas, em uma área experimental de 22 ha alocada num pivô central de 100 ha, cultivada no sistema de plantio direto em Neossolo Quartzorênico. O clima é tipo Aw, segundo a classificação climática de Köppen, com temperatura média entre 19°C e 28°C e pluviosidade média inferior a 2000 mm/ano ([AMBIENTE BRASIL, 2005](#)).

Coleta e tratamento dos dados

Os dados georreferenciados de produtividade coletados por sensores e DGPS, equipados em colhedora automotriz MF 34/38, foram mapeados em três safras (milho – soja – arroz). Foi selecionada para o estudo uma subárea de 22 ha, porção do pivô onde foram obtidos dados de produtividade nas três safras. O esquema de sucessão de culturas foi: milho – feijão – algodão – soja – arroz, sendo mapeados somente o milho, a soja e o arroz.

Os dados brutos de produtividade foram filtrados segundo metodologia descrita por [Menegatti e Molin \(2004\)](#), já que erros de posicionamento do receptor GPS e outros erros operacionais durante a colheita, tais como: largura e altura da plataforma, colheita parcial da largura da plataforma, manobras de cabeceiras, velocidade, tempo de enchimento podem embutir erros grosseiros nos dados reais de produtividade. Feita a filtragem de erros, fez-se a estatística descritiva dos dados das três culturas para se avaliar a qualidade dos dados utilizados na confecção dos mapas de produtividade. Esses mapas foram confeccionados em programa especializado de Sistema de Informações Geográficas (Arc View 3.2®).

Os dados de produtividade foram interpolados pelo método do inverso da distância, utilizando peso dois, resultando em um mapa com resolução espacial de 5 x 5 m. Os valores relativizados em relação à média da produtividade de cada cultura foram comparados ao longo do tempo nos mesmos locais, sendo que cada local representou aproximadamente 25 m² que corresponde à resolução do mapa final.

Delineamento das zonas de manejo baseado na produtividade relativizada pela média

Numa primeira abordagem, foi utilizada a metodologia descrita por Molin (2002) em que, para cada quadrícula do mapa, foi calculado quanto cada valor

de produtividade representava em relação à média. Nesse caso, cada valor de produtividade interpolado representou uma quadrícula de 5 x 5 m.

Para delinear as zonas de manejo, foram escolhidas três classes de produtividade segundo os critérios de estabilidade temporal em relação à média:

- a) Produtividade alta estável (produtividade 10% acima da média e com coeficiente de variação (CV) menor que 30% ao longo das três safras).
- b) Produtividade baixa estável (produtividade 10% abaixo da média e $CV < 30\%$).
- c) Produtividade média estável (produtividade média e $CV < 30\%$).
- d) Produtividade inconsistente (produtividade com $CV\% > 30\%$).

Dessa forma, os pontos com valores iguais ou superiores a 110% e $CV < 30\%$ constituiram a classe 1, os pontos com 100% e $CV < 30\%$ a classe 2 e os pontos com 90% ou menos a classe 3.

Delineamento automático das zonas de manejo

Numa segunda abordagem, foi feito um delineamento automático de zonas de manejo, para o qual se usou o algoritmo de classificação não-supervisionada *fuzzy k-means* com a utilização de um programa protótipo em desenvolvimento pelos autores feito em *Java Machine* denominado AESP ([QUEIROS et al., 2005](#)). Foram escolhidos inicialmente três valores de centróides para cada classe e iniciaram-se as iterações. Posteriormente, foram escolhidas três zonas de manejo para se trabalhar, e o programa agrupou-as da melhor forma com o emprego da técnica de *fuzzy k-means*.

Resultados e Discussão

Pela análise dos mapas de produtividade gerados ([Figura 1](#)), percebeu-se que na área existe alta variabilidade espacial da produção de grãos, sendo que o arroz apresentou maior variabilidade em relação ao milho e à soja pela análise de coeficiente de variação. Os coeficientes de variação da produtividade das culturas ao longo desses três anos foram 15%, 20% e 33%, e os valores médios de produtividade foram 7023, 2159 e 2730 kg/ha para milho, soja e arroz respectivamente.

Visualmente, tanto as zonas de produtividades altas quanto as baixas apresentaram certa correspondência espacial ao longo do tempo (Figura 1). No entanto, quando foram escolhidas classes arbitrárias para cada cultura, essas tendências não ficaram evidentes (Figura 2).

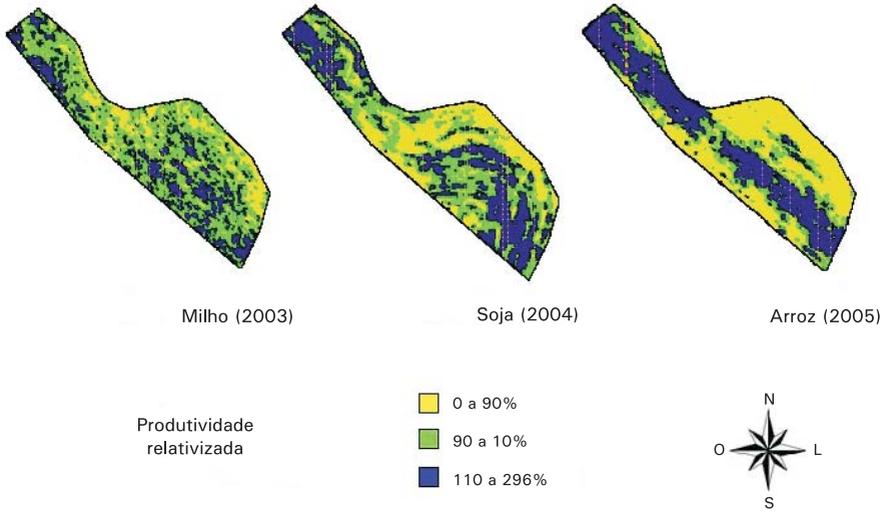


Figura 1. Mapas de produtividade relativizada (%) em relação à produtividade média da área.

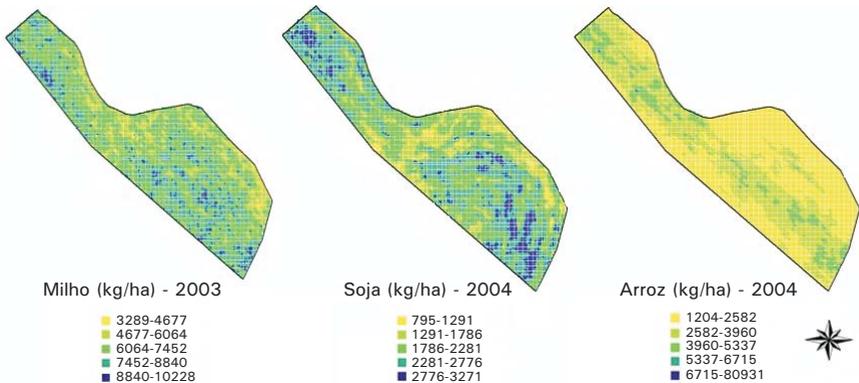


Figura 2. Mapas obtidos com escolha arbitrária das classes de produtividade.

Na Figura 3, observa-se que a proporção de áreas dentro de cada classe varia segundo a proporção adotada para classificação em relação à média. Quando a proporção se afasta da média (85% a 115%), a área de produtividade média aumenta (produtividade igual à média, com $CV < 30\%$). Portanto, as zonas de manejo mudam de local de acordo com a classificação adotada. Isso demonstra que a metodologia descrita por [Molin \(2002\)](#) permite definir zonas de manejo, porém, é sensível à escolha arbitrária de classes definidas pelo usuário.

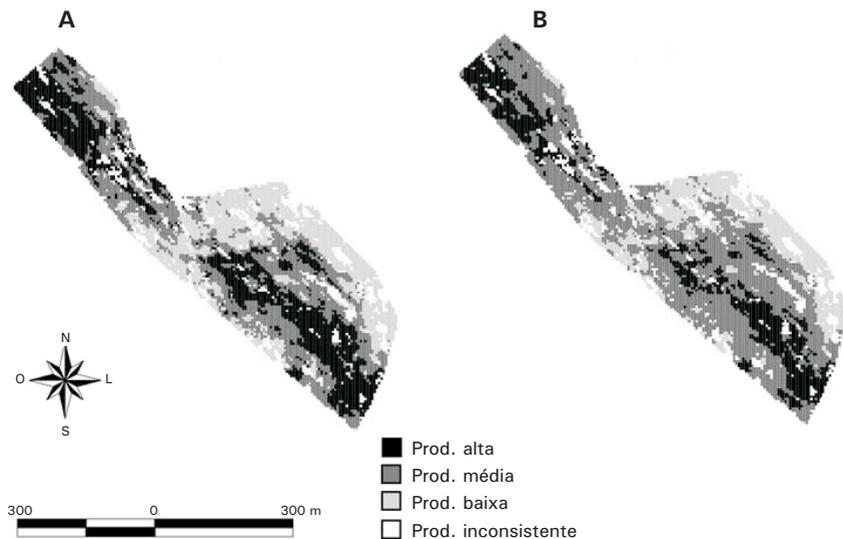


Figura 3. Delineamento de zonas de manejo baseado na produtividade relativizada com classificação considerando 90% a 110% (A) ou 85% a 115% (B) em relação à média da área.

Foi obtida correlação de $0,98^{**}$ quando se comparou o mapa gerado pela abordagem de definição de zonas de manejo pela classificação *k-means* (adotando três classes) com o mapa gerado pela abordagem de relativização em relação à média (adotando-se proporções de 10% acima e abaixo da média) ([Figura 4](#)). Isso demonstra que, nessa área, existe estabilidade espacial da produtividade ao longo do tempo, podendo-se dividi-la em duas zonas de manejo bem distintas, possibilitando futura investigação direcionada dos possíveis fatores que vêm interferindo na produtividade das culturas.

Outros trabalhos têm indicado que a estabilidade temporal da produtividade é susceptível ao clima (veranicos) e às características do solo (FRAISSE et al., 2001). Consideram, também, que em anos com distribuição normal de chuva, os mapas de produtividade reproduzem melhor essa tendência de zonas com produtividades semelhantes. É interessante frisar que, nas três safras monitoradas, a área experimental apresentou tendências temporais estáveis com zonas de produtividade muito contrastantes, apesar de ser uma área irrigada. Isso pode ser um indicativo de que fatores, como pragas de solo, plantas daninhas, compactação, possam também apresentar estabilidade espaço-temporal nesta área.

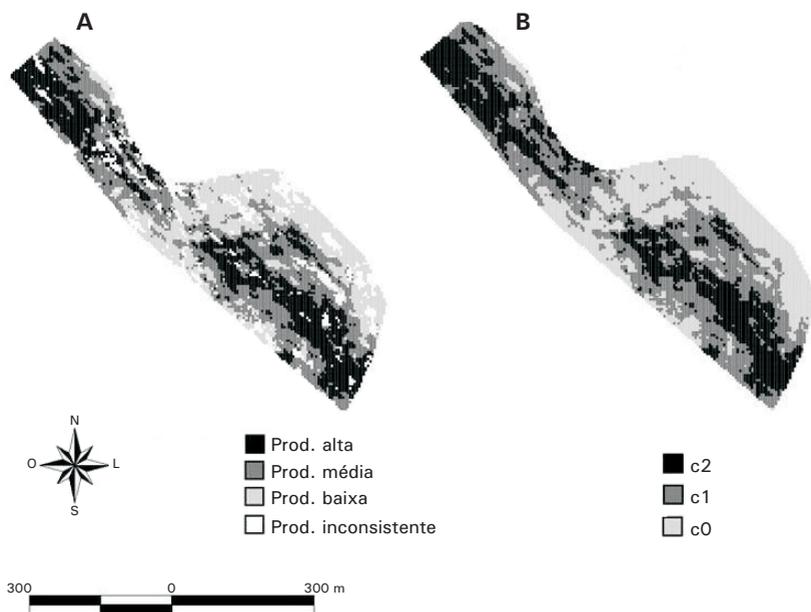


Figura 4. Zonas de manejo delineadas pelo método da estabilidade temporal da produtividade relativizada em relação à média do talhão (A) e por meio de *fuzzy k-means* (B).

Diversos fatores interferem na produtividade das culturas. Portanto, o uso exclusivo de mapas de produtividade pode não levar ao correto delineamento das zonas de manejo. Todavia, o mapeamento desses fatores sem o diagnóstico da produtividade inviabiliza a possibilidade de aferição das operações de manejo localizado realizadas pelo produtor.

A metodologia de *fuzzy k-means*, é uma ferramenta útil e de fácil utilização para o delineamento de zonas de manejo baseadas em dados de produtividade.

Conclusão

A utilização do algoritmo *k-means* para classificação não-supervisionada de dados espaço temporais de produtividade permitiu definir zonas de manejo sem depender de classificações arbitrárias definidas pelo usuário.

Referências

AMBIENTE BRASIL. **Classificação dos climas do Brasil**. Disponível em: < <http://www.ambientebrasil.com.br/composer.php3?base=./natural/index.html&conteudo=./natural/clima.html> >. Acesso em: 01 nov. 2005.

AUSTRALIAN CENTRE FOR PRECISION AGRICULTURE. **Fuzzy k-means with extragrades program**. Disponível em: < <http://www.usyd.edu.au/su/agric/acpa/pag.htm> > Acesso em: 10 nov. 2005.

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

BIRRELL, S. J.; SUDDUTH, K. A.; BORGELT, S. C. Comparison of sensors and techniques for crop yield mapping. **Computers and Electronics in Agriculture**, Amsterdam, v. 14, n. 2/3, p. 215-233, 1996.

BROCK, A.; BROUDER, S. M.; BLUMHOFF, G.; HOFMANN, B. S. Defining yield-based management zones for corn-soybean rotations. **Agronomy Journal**, Madison, v. 97, p. 1115-1128, 2005.

DOERGE, T. Weigh wagon vs. yield monitor comparison. **Crop Insights**, Johnston, v. 7, n. 17, p. 1-5, 1996.

FRAISSE, C. W.; SUDDUTH, K. A.; KITCHEN, N. R. Delineation site-specific management zone by unsupervised classification of topographic attributes and soil electrical conductivity. **Transactions of the ASAE**, St. Joseph, v. 44, n. 1, p. 155-166, 2001.

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**, Madison, v. 96, p. 100-108, 2004.

GRISSE, R. D.; JASA, P. J.; SCHROEDER, M. A.; WILCOX, J. C. Yield monitor accuracy: successful farming magazine case study. **Applied Engineering in Agriculture**, St. Joseph, v. 18, p. 147-151, 2002.

LUZ, N. C. **Algoritmos de clusterização como apoio à modelagem de preditores de sítios arqueológicos baseados em variáveis ambientais**. 2004. 87 f. Dissertação (Mestrado) - Escola de Engenharia Elétrica e de Computação, Universidade Federal de Goiás, Goiânia, 2004.

MANAGEMENT ZONE ANALYST 1.0. Disponível em: <http://www.fse.missouri.edu/ars_software/mza_reg.asp> Acesso em: 22 nov. 2005.

MENEGATTI, L. A. A.; MOLIN, J. P. Remoção de erros em mapas de produtividade via filtragem de dados brutos. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 8, n. 1, p. 126-134, 2004.

MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**, Jaboticabal, v. 22, n. 1, p. 83-92, 2002.

MULLA, D. J.; BHATTI, A. U.; HAMMOND, M. W.; BENSON, J. A. A comparison of winter wheat yield and quality under uniform versus spatially variable fertilizer management. **Agriculture Ecosystems and Environment**, Amsterdam, v. 38, p. 301-311, 1992.

QUEIROS, L. R.; SHIRATSUCHI, L. S.; VINHAL, C. D. N. Desenvolvimento de um sistema protótipo para o mapeamento da altura de plantas de algodão. In: SIMPÓSIO INTERNACIONAL DE AGRICULTURA DE PRECISÃO, 3., 2005, Sete Lagoas. **Anais... Sete Lagoas: Embrapa Milho e Sorgo**, 2005. 1 CD-ROM.

SCHEPERS, A. R.; SHANAHAN, J. F.; LIEBIG, M. A.; SCHEPERS, J. S.; JOHNSON, S. H.; LUCHIARI JR, A. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal**, Madison, v. 96, p. 195-203, 2004.

VENTURIERI, A. **Segmentação de imagens e lógica nebulosa para treinamento de uma rede neural artificial na caracterização do uso da terra na Região de Tucuruí (PA)**. 1996. 115 f. Dissertação (Mestrado) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.