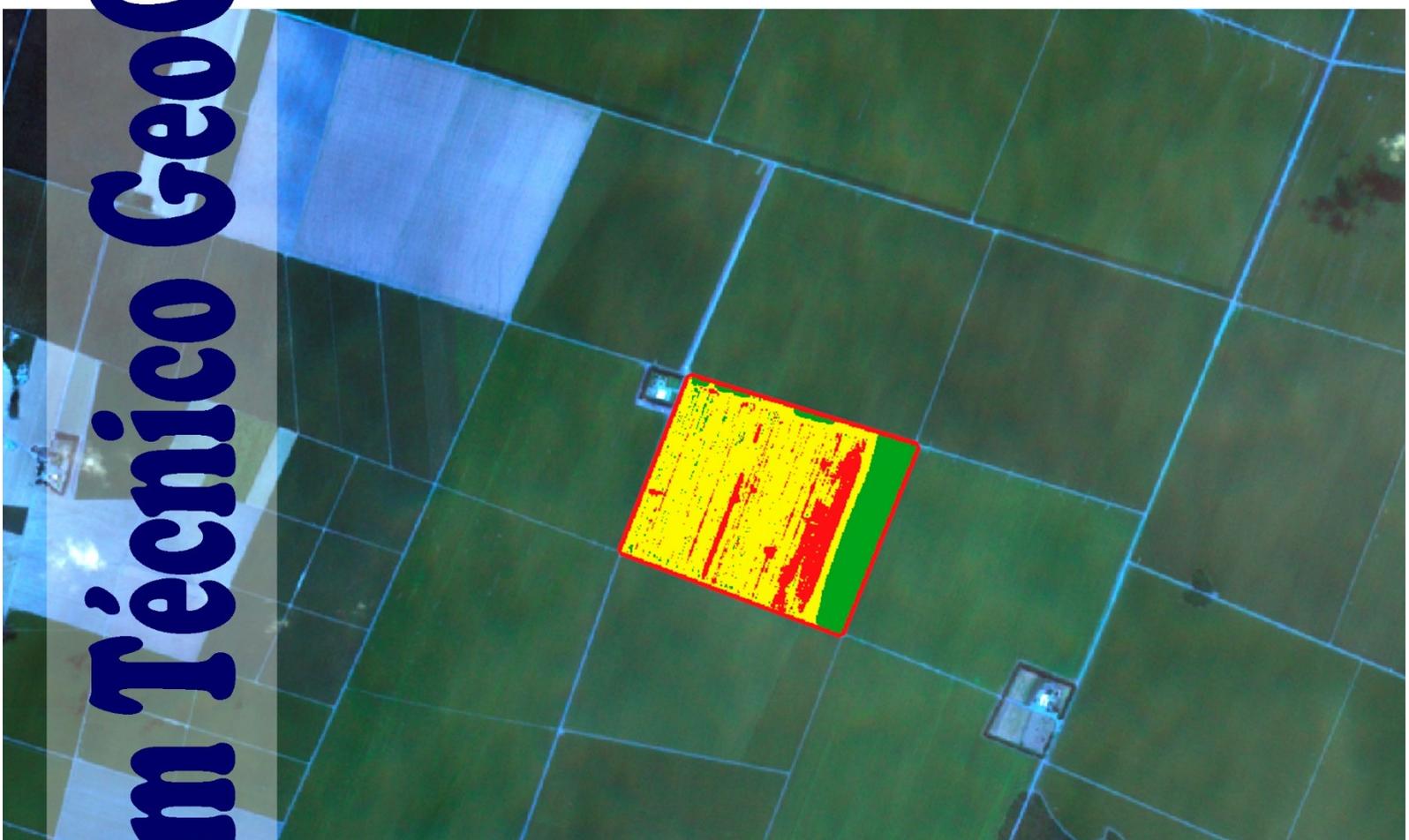


**Classificação e agrupamento de dados
em mapas de produtividade para
agricultura de precisão.**



Boletim Técnico GeoClimaMT
<http://pesquisa.unemat.br/geoclimamt/>

v. 6, n. 1, 2020

Universidade do Estado de Mato Grosso – UNEMAT
Centro Tecnológico de Geoprocessamento e
Sensoriamento Remoto– CETEGEO-SR
Grupo de estudos em geotecnologias aplicadas às mudanças
climáticas e agricultura digital – GeoClimaMT

Classificação e agrupamento de dados em mapas de produtividade para agricultura de precisão

Rafael Cesar Tieppo
Daniel Gonçalves Riselo
Gabriel Vergilio Barboza
Vitor Alfeu Guedes Moreira Vieira



Governo do Estado de Mato Grosso
Secretaria do Estado de Ciência e Tecnologia - SECITECI
Universidade do Estado de Mato Grosso – UNEMAT
Centro Tecnológico de Geoprocessamento e Sensoriamento
Remoto aplicado à produção de Biodiesel – CETEGEO-SR
Grupo de estudos em geotecnologias aplicadas às mudanças climáticas e
agricultura digital – GeoClimaMT

Autor corporativo:

Centro Tecnológico de Geoprocessamento e Sensoriamento Remoto aplicado à produção de Biodiesel – CETEGEO-SR

Endereço:

Rod. MT 358, km 07, s/n, Jardim Aeroporto, Tangará da Serra, Mato Grosso, Brasil
CEP: 78.300-000 Fone: (65) 3311-4937
e-mail: geoclimamt@unemat.br homepage: <http://pesquisa.unemat.br/geoclimamt/>

Corpo Editorial:

Cornelio Alberto Zolin (Embrapa Agrossilvipastoril)
Edenir Maria Serigatto (UNEMAT)
Erivelto Mercante (UNIOESTE)
Marco Antonio Camillo de Carvalho (UNEMAT)
Maria Carolina da Silva Andrea (UNEMAT)
Paulo Sergio Lourenço de Freitas (UEM)
Rafael Cesar Tieppo (UNEMAT)
Rivanildo Dallacort (UNEMAT)
Thiago Libório Romanelli (Esalq-USP)
Vanessa Rakel de Moraes dias (UNEMAT)

Revisão de texto: Renato Cardoso de Moraes

Ilustração da capa: Vitor Alfeu Guedes

O conteúdo deste boletim e seus dados em sua forma e confiabilidade são de responsabilidade exclusiva dos autores.

Tieppo, Rafael Cesar (et al.).

Classificação e agrupamento de dados em mapas de produtividade para agricultura de precisão/ Rafael Cesar Tieppo; Daniel Gonçalves Riselo; Gabriel Vergilio Barboza; Vitor Alfeu Guedes Moreira Vieira. – Tangará da Serra: UNEMAT, 2020.

v.06, n.01. 12p. Il. (Boletim Técnico Grupo de estudos em geotecnologias aplicadas às mudanças climáticas e agricultura digital – GeoClimaMT)

ISSN: 2595-8550

1. Agricultura de precisão. 2. Mapas. 3. Sensoriamento remoto. I. Título. II. Autor.

CDU 528.711.7:528.9:553.3/.9 (817.2)

WALTER CLAYTON DE OLIVEIRA CRB 1/2049



Autores

Rafael Cesar Tieppo

Eng. Agrícola. D.Sc.

Universidade do Estado de Mato Grosso

e-mail: rafaelt@unemat.br

Gabriel Vergilio Barboza

Graduando em Agronomia

Universidade do Estado de Mato Grosso

e-mail: gabrielvergiliobarboza@hotmail.com

Vitor Alfeu Guedes Moreira Vieira

Graduando em Agronomia

Universidade do Estado de Mato Grosso

e-mail: vitoralfeuguedes@gmail.com

Daniel Gonçalves Riselo

Graduando em Agronomia

Universidade do Estado de Mato Grosso

e-mail: daniel.riselo@hotmail.com



Introdução

A agricultura de precisão (AP) é uma técnica para gerenciamento de sistemas agrícolas, que utiliza a tecnologia de informação para evidenciar as particularidades de uma determinada área, com o intuito de maximizar os resultados a serem obtidos (BONGIOVANNI; LOWENBERG-DEBOER, 2004).

Mesmo assumindo que o uso do mapa de produtividade é uma fração do potencial do sistema de AP, sua utilização é uma opção de ponto de partida da referida tecnologia, sendo que possibilita representar o reflexo do efeito combinado de fontes de variabilidade espacial e temporal, permitindo diagnosticar as possíveis causas de modificação às respostas econômicas e de impacto ambiental do sistema de produção (QUEIROZ, 2000), (MOLIN, 2002), (MILANI et al., 2006), (TIEPPO et al., 2007) .

Antes da utilização dos dados brutos oriundos dos monitores de colheita, os mesmos devem ser tratados devido às possíveis informações inconsistentes contidas no arquivo (PING; DOBERMANN, 2005). Com o intuito de realizar a filtragem dos arquivos gerados pelo monitor de colheita (BLACKMORE, 1999) (JURSCHIK; GIEBEL; WENDROTH, 1999), (MOLIN; GIMENEZ, 2004), (MENEGATTI; MOLIN, 2004), (SPEKKEN; ANSELMINI; MOLIN, 2013), entre outros autores, desenvolveram rotinas para a realização de filtragem dos dados.

Após o processo de filtragem, com o intento de averiguar a distribuição dos dados da produtividade da cultura, inicia-se uma análise descritiva, em que os parâmetros média, desvio padrão e coeficiente de variação podem ser obtidos como indicadores de uniformidade da produtividade.

Além de averiguar o comportamento da variável produtividade, torna-se interessante realizar o seu agrupamento. Uma das primeiras ocorrências do termo agrupamentos de dados na ciência ocorreu em um estudo na área de saúde (TYRON, 1939). Os métodos de agrupamento são algoritmos de classificação, que têm por objetivo organizar as observações em estruturas que apresentem informação sobre o conjunto de dados, ou como desenvolver classes para os dados observados.

No caso de mapas de produtividade em AP, o agrupamento de dados consiste em identificar conjuntos de pontos de produtividade que se assemelham ao longo do talhão, que por sua vez delimitam regiões com níveis de produtividade distintas.

Entre os métodos de agrupamentos hierárquicos ou particionais há uma variedade de opções que podem ser implementadas. Com finalidade didática na introdução à classificação e agrupamento de dados de mapas de produtividade, este material teve por objetivo apresentar a classificação de dados com uso de quartis, e agrupamento de dados por meio de do algoritmo k-means.

Material e Métodos

Os dados de produtividade utilizados na classificação e no agrupamento são oriundos de uma simulação da cultura do milho de segunda safra no estado de MT. A área analisada foi composta por quatro talhões denominados T08, T09, T10 e T11, com área de 127,2; 131,6; 132,6 e 127,0 ha, respectivamente (Figura 1).

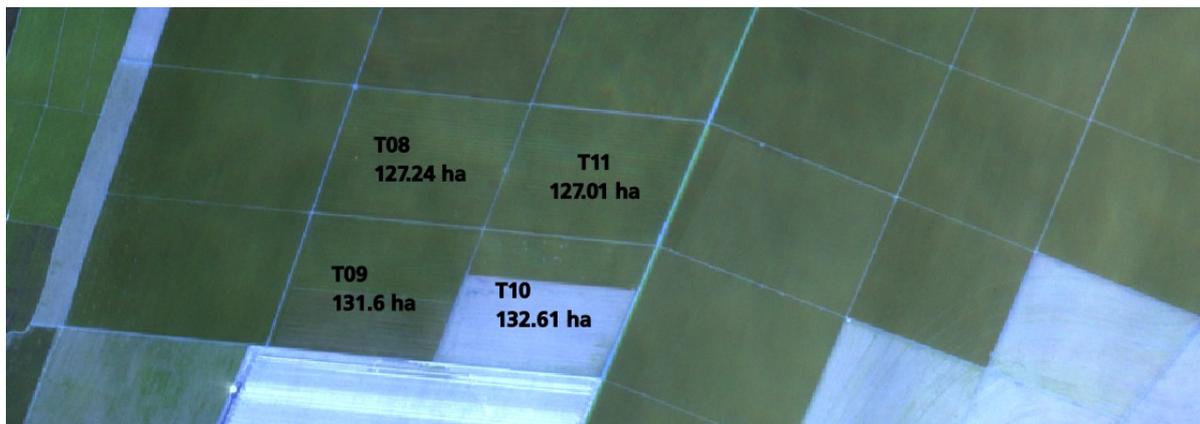


Figura 1: Área utilizada para simulação dos dados de produtividade, imagem de fundo composição RGB (Sentinel). Escala 1:26000. Datum: EPSG:4326.

As rotinas utilizadas para o processamento dos dados foram desenvolvidas na linguagem Python (ROSSUM, 1995) e no SIG (Sistema de Informação Geográfica) QGIS (QGIS DEVELOPMENT TEAM, 2015).

Os valores de quartis compreendem parte da estatística descritiva que permite extrair informações do conjunto amostral, e usualmente são divididos em primeiro quartil (Q1), segundo quartil (Q2), terceiro quartil (Q3) e quarto quartil (Q4).

A obtenção dos valores se dá a partir da disposição em ordem crescente do conjunto de dados e posterior divisão da distribuição em quatro partes iguais. O Q1 corresponde aos primeiros 25% das observações, o Q2 (mediana) corresponde ao intervalo entre 25 e 50%, o Q3 entre 50 e 75%, e o Q4 ao intervalo 75 e 100% (ou o valor máximo) das observações.

A posição de cada quartil no conjunto de dados ordenado é obtida pela relação do número de amostras e ordem do quartil (equação 1).

$$Q_i = X_j + \left(\frac{i(n+1)}{4} - j \right) * (X_{j+1} - X_j) \quad \text{sendo que } i \in \mathbb{N} \wedge 1 \leq i \leq 4$$

em que Q é o quartil, i é a posição do quartil, X é o elemento do conjunto ordenado, n o número de amostras, j é a posição do elemento X sendo que j é o maior inteiro desde que $j \leq i(n+1)/4$.

Como exemplo, considere o conjunto de dados $S = [3, 4, 5, 5, 5, 6, 7, 8, 9]$.

$$Q_1 = X_2 + \left(\frac{1(9+1)}{4} - 2 \right) * (X_{2+1} - X_2) = 4 + \left(\frac{1(9+1)}{4} - 2 \right) * (5 - 4) = 4,5$$

$$Q_2 = X_5 + \left(\frac{2(9+1)}{4} - 5 \right) * (X_{5+1} - X_5) = 5 + \left(\frac{2(9+1)}{4} - 5 \right) * (6 - 5) = 5,0$$

$$Q_3 = X_7 + \left(\frac{3(9+1)}{4} - 7 \right) * (X_{7+1} - X_7) = 7 + \left(\frac{3(9+1)}{4} - 7 \right) * (8 - 7) = 7,5$$

Uma vez calculado os valores dos quartis torna-se possível gerar um gráfico boxplot (Figura 2).



Figura 2: Boxplot do conjunto S (linha em vermelho indica mediana).

Nota-se que o limite inferior é igual a 3, o qual representa o menor valor do conjunto de dados S . A posição do Q_1 calculado é igual a 4,5, sendo que o valor correspondente em S é 5, que inclusive possui o mesmo valor de Q_2 (mediana). O Q_3 calculado indica que o terceiro quartil é composto por valores até 7,5, cujo o valor correspondente no conjunto S é 7. Pelos resultados obtidos observa-se que 50,0% dos elementos de S estão entre 5 e 7.

O agrupamento K-means é um dos métodos utilizados em técnicas de machine learning (aprendizado de máquina). Como observação, salienta-se que apesar da ciência de machine learning ser tratada como tema atual, a mesma já era aplicada por volta de 1960 (MCCARTHY et al., 2006). O k-means é um método não supervisionado e atua como agrupador lógico de segmentação, com foco na investigação por similaridade entre os dados da amostra. Esta por sua vez é definida pela distância entre as amostras.

O K , de K-Means, indica a quantidade de centroides (pontos centrais dos grupos) que são criados para dar suporte na definição da similaridade dos dados. De forma sucinta, o algoritmo funciona em quatro fases:

- Inicialização - algoritmo gera de forma aleatória k centroides, onde o número de centroides é representado ao parâmetro k ;
- Atribuição ao Cluster - cálculo da distância entre todos os pontos de dados e cada um dos centroides, cálculo da distância (euclidiana), divisão dos dados conforme o número de centroides estipulado pelo argumento k .
- Movimentação de Centroids - cálculo da média dos valores dos pontos de dados de cada cluster, sendo que o valor médio será a posição do novo centroide (Figura 3).
- Otimização do K-means – a fase de movimentação de centroides se repete até que o cluster se torne estático.

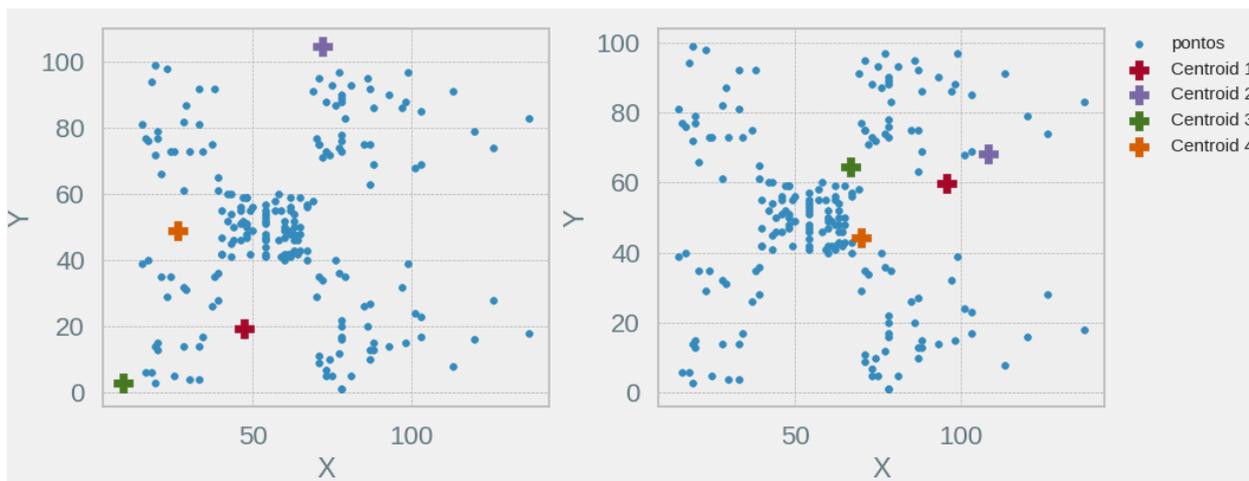


Figura 3: Processo iterativo de posicionamento dos centroides

Uma dúvida comum no uso de técnicas de agrupamentos de dados é definir o número de agrupamentos (cluster) para cada conjunto de dados. Em situações em que não se conhece o número de agrupamentos desejado é possível utilizar o método do cotovelo (elbow) que, de modo simplista, realiza testes da variância dos dados em relação ao número de agrupamentos (Figura 4).

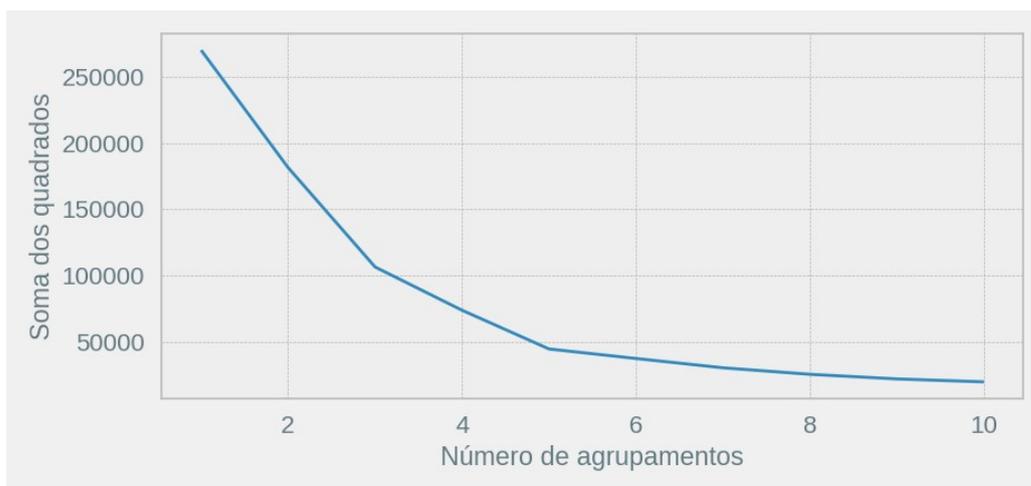


Figura 4: Definição do número de agrupamentos pelo método “Elbow”

Resultados e Discussão

O conjunto de dados analisados totalizaram 195.248 pontos amostrais de produtividade, com médias de produtividade similares entre os talhões (Tabela 1).

Tabela 1: Análise descritiva dos dados de produtividade (kg ha⁻¹)

Talhão	n	Mín.	Máx.	Média	Q1	Mediana	Q3	Sd
T08	48353	220,3	16063,6	8337,2	7765,1	8322,3	8933	975,4
T09	49202	2942,7	14541,8	8087,9	7497,9	8047,3	8717,1	971,4
T10	49856	1,0	15897,1	8209,9	7667,45	8240,2	8894,5	1076,2
T11	47837	7,8	15783	8228,5	7676,7	8203	8816,2	908,4
Todos	195248	1	16063,6	8215,29	7646,9	8205	8844,6	989,57

n = número de amostras; Q1 = primeiro quartil; Q2 = segundo quartil; Sd desvio padrão

Além de valores médios de produtividade similares, nota-se pela estatística descritiva que os talhões são relativamente homogêneos, com coeficiente de variação de aproximadamente 11,0%. Pelos resultados encontrados, no que tange às técnicas da agricultura convencional, os talhões seriam tratados de maneira uniforme. Todavia, com o advento da agricultura de precisão, há possibilidade da realização de tratamentos personalizados ao longo do talhão. Para que isto possa ser feito, algumas das alternativas é a classificação dos níveis de produtividade e o agrupamento de dados (Figura 5).

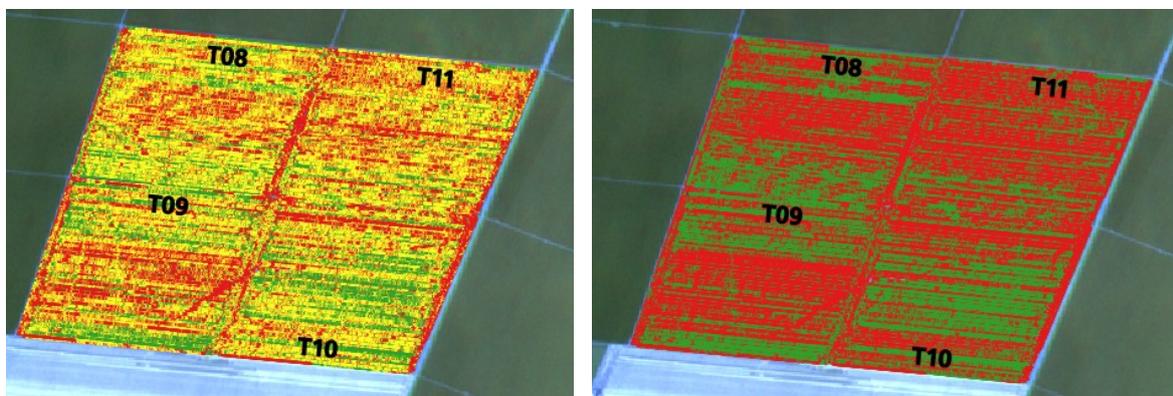


Figura 5: Classificação dos dados de produtividade por quartis (esq), sendo vermelho, amarelo, verde para produtividades baixa, média e alta, respectivamente. Agrupamento (cluster) dos dados pelo método k-means (dir.), sendo grupo inferior vermelho e grupo superior verde.

Em ambos os métodos foi possível identificar as áreas que proporcionaram níveis de produtividade distintos. Na classificação pelos quartis (Figura 5 esq.) nota-se a presença de áreas vermelhas (produtividade baixa) na bordaduras, que de certa

forma é esperado, e no centro dos talhões T08, T09 e T11. O resultado obtido no agrupamento por k-means (Figura 5 dir.) apresentou um comportamento similar, pois os locais menos produtivos estão nas regiões centrais de cada talhão. Dessa forma, a presença de possíveis anomalias (compactação, nematóides, etc) deveria ser investigada na região central dos talhões com produtividade inferior.

No agrupamento por k-means (Figura 5 dir.) há dois grupos, a região em vermelho tem produtividade média de 7490,2 kg ha⁻¹ contra e 8963,0 kg ha⁻¹ da região em verde, o que indica uma possibilidade de incremento de produtividade em uma próxima safra.

Comparando os intervalos obtidos em cada método, na classificação por quartis metade dos valores ficam entre 7646.9 e 8844.6 kg ha⁻¹ (região em amarelo), as demais regiões acima e abaixo do referido intervalo representam 25,0% cada. No agrupamento por k-means, para este caso, os grupos são seccionados em um valor próximo da mediana, o que de certa forma é esperado, pois uma das dimensões no modelo é constante.

Considerações Finais

A utilização do método de classificação pelos quartis, assim como o agrupamento por k-means, possibilitou identificar espacialmente a produtividade nas áreas predeterminadas, permitindo assim caracterizar os níveis de produtividade em cada talhão. Por ambos os métodos foi possível indicar regiões com prováveis anomalias.

Havendo disponibilidade de dados espaciais de outras variáveis (teor de nutriente, compactação, etc), pelo método de agrupamento k-means torna-se possível agrupar os dados em função da produtividade e as referidas variáveis.

Referências bibliográficas

BLACKMORE, Simon. Remedial Correction of Yield Map Data. **Precision Agriculture**, [S. l.], v. 1, n. 1, p. 53–66, 1999. DOI: 10.1023/A:1009969601387. Disponível em: <http://dx.doi.org/10.1023/A:1009969601387>.



BONGIOVANNI, R.; LOWENBERG-DEBOER, J. Precision Agriculture and Sustainability. **Precision Agriculture**, [S. l.], v. 5, n. 4, p. 359–387, 2004. DOI: 10.1023/B:PRAG.0000040806.39604.aa. Disponível em: <http://dx.doi.org/10.1023/B:PRAG.0000040806.39604.aa>.

JURSCHIK, P. K.; GIEBEL, A.; WENDROTH, O. P. Processing of point data from combine harvesters for precision farming. In: STAFFORD, J. (org.). **Precision Agriculture 99**. [s.l.] : Sheffield Academic Press, 1999. p. 297–307.

MCCARTHY, John; MINSKY, Marvin L.; ROCHESTER, Nathaniel; SHANNON, Claude E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. **AI Magazine**, [S. l.], v. 27, n. 4, p. 12, 2006. DOI: 10.1609/aimag.v27i4.1904.

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

MILANI, Lairton; SOUZA, Eduardo Godoy; URIBE-OPAZO, Miguel Angel; GABRIEL FILHO, Antonio; JOHANN, Jerry Adriany; PEREIRA, Joaquim Odilon. Unidades de manejo a partir de dados de produtividade. **Acta Scientiarum: Agronomy**, Maringá, v. 28, n. 4, p. 591–598, 2006. DOI: 10.4025/actasciagron.v28i4.937. Disponível em: <http://periodicos.uem.br/ojs/index.php/ActaSciAgron/article/view/937>.

MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia agrícola**, Jaboticabal-SP, v. 22, n. 11, p. 83–92, 2002.

MOLIN, J. P.; GIMENEZ, Leandro Maria. Algoritmo para redução de erros em mapas de produtividade para agricultura de precisão. **Revista Brasileira de Agrocomputação**, [S. l.], v. 2, n. 1, p. 5–10, 2004.

PING, J. L.; DOBERMANN, A. Processing of Yield Map Data. **Precision Agriculture**, [S. l.], v. 6, n. 2, p. 193–212, 2005. DOI: 10.1007/s11119-005-1035-2. Disponível em: <http://link.springer.com/10.1007/s11119-005-1035-2>.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System**, 2015. Disponível em: <http://qgis.osgeo.org>.

QUEIROZ, D. .. Agricultura de precisão na produção de grãos. In: BORÉM, A. (org.). **Agricultura de precisão**. Viçosa: UFV, 2000. p. 1–42.

ROSSUM, G. van (. Centrum voor Wiskunde en Informatica (CWI)). **Python tutorialPython**. Amsterdam.

SPEKKEN, M.; ANSELMINI, A. A.; MOLIN, J. P. A simple method for filtering spatial data. In: IX EUROPEAN CONFERENCE ON PRECISION AGRICULTURE 2013, **Anais [...]**. : Fundação Universitária de Lleida, 2013.

TIEPPO, R. C.; SOUZA, E. G.; URIBE-OPAZO, M. A.; SILVEIRA, J. C. M. Avaliação de diferentes interpoladores na geração de mapas temáticos da produtividade de soja em agricultura de precisão. In: IV SIMPÓSIO INTERNACIONAL DE AGRICULTURA DE PRECISÃO 2007, Viçosa. **Anais [...]**. Viçosa p. 1–4.

TYRON, R. C. **Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality**. Ann Arbor. v. 3