

Application of data grouping methods into vegetation indices to generate management zone maps.

Aplicação de métodos de agrupamento de dados em índices de vegetação para geração de mapas de zonas de manejo.

Fábio S. Ruver, Raul T. Santos, Cláudia A. Martins

Instituto de Computação
Universidade Federal do Mato Grosso (UFMT) - Cuiabá, MT – Brazil

fabio.ruver{@sou.ufmt.br}, raul,claudia{@ic.ufmt.br}

Abstract. *This article aims to evaluate the impact of using different agricultural indices to define management zones using pre-processing techniques to prepare the data for the use of clustering algorithms on spatial data from a cotton plantation. The methodology involved acquiring agricultural indices data by drone, pre-processing including removing not a number values, creating tables of comma separated values and selecting indices based on correlation. The Fuzzy C-Means algorithm was implemented to create maps of management zones with four groups, identifying flaws in the plantation. With this, we visually obtained areas with different health levels for cultivation.*

Keywords: *Management zones, Fuzzy C-Means, precision agriculture, pre-processing, vegetation indices*

Resumo. *Este artigo tem como objetivo avaliar o impacto do uso de diferentes índices agrícolas para a definição de zonas de manejo utilizando técnicas de pré-processamento visando preparar os dados para o uso de algoritmos de agrupamento em dados espaciais de uma plantação de algodão. A metodologia envolveu a aquisição de dados de índices agrícolas por drone, pré-processamento incluindo a remoção de valores not a number, criação de tabelas de valores separados por vírgula e seleção de índices com base na correlação. O algoritmo Fuzzy C-Means foi implementado para criar mapas de zonas de manejo com quatro grupos, identificando falhas na plantação. Com isso, obtivemos visualmente áreas com diferentes níveis de saúde para o cultivo.*

Palavras-chave: *Zonas de manejo, Fuzzy C-Means, agricultura de precisão, pré-processamento, índices de vegetação*

1. Introdução

A agricultura desempenha um papel fundamental na economia brasileira, com a produção de algodão posicionando o Brasil entre os maiores produtores mundiais [Hopewell 2013]. A integração de diferentes abordagens visa a potencialização da eficiência e produtividade desse setor vital, com contribuições significativas para seu crescimento constante no cenário agrícola global [Zarco 2005].

Este estudo aborda um campo de pesquisa multifacetado, repleto de técnicas e conceitos fundamentais que têm desempenhado um papel crucial em seu

desenvolvimento. Entre essas abordagens, destacam-se métodos de inteligência artificial aplicados à agricultura de precisão, como demonstrado nos trabalhos de Meneses (2021) e Nallan (2014). Esses estudos forneceram uma sólida base teórica, evidenciando o potencial da inteligência artificial na otimização das práticas agrícolas.

Da mesma forma, a utilização de índices agrícolas, conforme discutido nas pesquisas de Rodrigues (2021) e Trindade (2017), desempenha um papel central na seleção dos parâmetros analíticos para o estudo. Essas referências direcionaram a escolha criteriosa dos índices mais relevantes para a agricultura de precisão, destacando seu valor na pesquisa.

Além disso, a definição de zonas de manejo, conforme explorada nos estudos de Santos (2014) e Valente (2010), oferecem valiosas compreensões sobre a subdivisão do terreno com base em atributos consistentes ao longo do tempo. A noção de zonas de manejo é fundamental para este estudo, uma vez que visa melhorar a eficiência e a sustentabilidade da gestão agrícola.

Finalmente, técnicas de pré-processamento de dados espaciais, conforme elucidadas nas pesquisas de Devi (2021) e RUß (2010), desempenham um papel essencial no tratamento de dados de sensoriamento remoto e imagens. Essas técnicas forneceram orientações essenciais sobre como preparar os dados para análises subsequentes.

O objetivo deste estudo é investigar os benefícios da aplicação de técnicas de agrupamento em índices de vegetação. Ao fazê-lo, busca-se demonstrar como os resultados podem ser direcionados para aprimorar a tomada de decisões na agricultura, promovendo um manejo mais eficaz e sustentável das plantações de algodão. Essa abordagem não apenas reforça a liderança do Brasil na produção de algodão, mas também contribui para a evolução da agricultura de precisão, um campo de crescente relevância no cenário atual [Lindblom 2017] [Shuang *et al* 2021].

2. Metodologia

2.1 Aquisição dos Dados

O trabalho foi realizado com índices de vegetação obtidos por drone em 4 diferentes fases de crescimento do algodão disponibilizados pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Os dados são de imagens multiespectrais em 6 bandas do espectro Red, Green, Blue, Near-Infrared (NIR), RedEdge e Thermal de talhões de algodão de uma fazenda em Mato-Grosso. A área mapeada pelo drone é um semicírculo de aproximadamente 100 hectares, onde metade dele foi dividida em 54 blocos de dimensão de 100 m por 100 m. Onde para cada bloco foram extraídos 6 índices. No total foram feitos 4 voos para aquisição das imagens, que são relativos a diferentes fases fenológicas do cultivo do algodão. Após obtido os ortomosaicos das 6 bandas devidamente georreferenciados, foi gerada a composição red, green, blue (RGB) e os

seguintes índices de vegetação no formato TIFF: Normalized Difference Vegetation Index (NDVI), Normalized Difference Vegetation Index (NDRE), Simple Chlorophyll Carotenoid Content Index (SCCCI), Soil-Adjusted Vegetation Index (SAVI), Modified Chlorophyll Absorption in Reflectance Index (MCARI) e Green Leaf Index (GLI).

2.2. Pré-processamento dos índices

2.2.1. Remoção de valores não numéricos

Primeiramente foi identificado que os índices apresentaram valores Not a Number (NaN). Esses valores estavam presentes pois as imagens do talhão foram divididas em blocos, e com isso partes de uma imagem estariam em outras. Essa sobreposição foi substituída por valores NaN. E para que não afetasse no desempenho do algoritmo de agrupamento, esses valores foram removidos através de um código em Python utilizando a função `dropna` da biblioteca Pandas.

2.2.2. Criação da tabela CSV

Para automatizar o processo de criação de tabelas e ter um processo mais modularizado, os valores de índices foram armazenados em tabelas de valor separados por vírgula (CSV). Para isso criou-se um código Python que percorre as imagens no formato TIFF, extrai informações de cada pixel dessas imagens e cria a tabela associando a localização dele à seus valores de índices. Utilizou-se as bibliotecas “numpy” para operações numéricas eficientes, “tiffio” para ler imagens TIFF e “pandas” para manipulação de dados tabulares.

2.2.3. Seleção de dados

Foi criada uma matriz de correlação para avaliar a correlação entre eles. Isso é importante, pois caso sejam utilizadas combinações de índices onde existam muitos índices fortemente correlacionados, os resultados podem ser tendenciosos [Javed 2021]. Como foram 6 índices fornecidos, precisava-se saber se era necessário utilizar todos ou não. Para se ter essa informação foi criada a matriz de correlação com todos os índices de um bloco no terceiro voo. Para ser gerada, foi calculada a média de cada índice por bloco. O coeficiente de correlação utilizado foi o de Pearson, pois é o padrão do *dataframe* Pandas. Isso permitiu saber se os índices têm alta ou baixa relação entre si. Como critério de descarte, foram escolhidos os índices que apresentavam correlação menor que 0.84 entre si.

2.3 Definição de mapas de zonas de manejo

A próxima etapa foi a execução do algoritmo Fuzzy C-Means (FCM). Existem 5 parâmetros que precisam ser ajustados para o Fuzzy: os dados utilizados, que no caso são mapas de índices; o número de zonas de manejo desejado; o fator de fuzziness, que foi mantido com o valor padrão de 2; o erro, que determina quando o algoritmo deve parar de fazer iterações, e também foi utilizado o valor padrão de 0.0005; e o número de máximo de iterações que mantivemos o padrão de 1000.

Na Agricultura de Precisão (AP), uma prática amplamente adotada é a criação das chamadas zonas de manejo (ZM). Nesse contexto, a extensão do terreno é subdividida em regiões menores, cada uma com seu próprio potencial produtivo distinto. Essa subdivisão se baseia em atributos que demonstram consistência ao longo do tempo, tais como a condutividade elétrica do solo, altitude, NDVI, textura do solo e resistência do solo à penetração, como exemplificado por Bottega (2014).

Para criação dos mapas com as zonas de manejo (ZMs), foi criado um código em Python que utiliza a técnica de agrupamento FCM para agrupar os dados representativos. Primeiramente, ele lê os valores de índices da tabela CSV criada. Em seguida são passados os parâmetros necessários para o FCM, apresentados no parágrafo, depois os dados dessas imagens são submetidos ao algoritmo FCM, que gera centróides para os grupos e uma matriz de pertinência que associa cada pixel a um grupo. Com base nesses grupos, são atribuídas cores aos pixels e, finalmente, um mapa com diferentes cores é criado em que o eixo x e y representam a localização de cada pixel.

2.3.1. Definição da quantidade ideal das ZMs

Para a definição da quantidade ideal das ZMs, foi consultada a biblioteca skfuzzy do Python, a documentação pode ser encontrada no site Python Software Foundation (PSF). Nela, podemos ver que a função de agrupamento retorna o valor fuzzy partition coefficient (FPC), que é utilizado como um índice para determinar se a quantidade de ZMs escolhida é a ideal. Esse retorno é um valor entre 0 e 1 e quanto maior, melhor. Para definir o valor ideal de zonas de manejo, foi criado um algoritmo em python para calcular o FCM com diferentes valores de grupos (ZMs). Foram testados valores de 2 até 8 grupos.

3. Resultados e discussões

3.1 Aquisição dos dados

A Figura 1 mostra o índice MCARI do bloco 23 do terceiro voo. Pode-se observar como os índices são representados e como estão dispostos, ficando claro o tipo de dados utilizado.

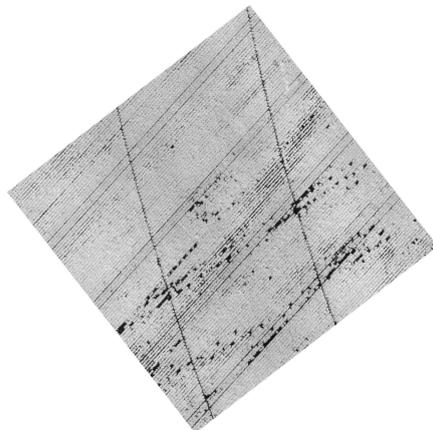


Figura 1. Índice MCARI do bloco 23 do terceiro voo.

3.2. Pré-processamento dos índices

3.2.1 Remoção de valores NaN

A Figura 2 mostra, em vermelho, a localização dos valores NaN, enquanto que os demais valores estão em preto. Com isso podemos ter uma noção visual de como foram divididos os grids e a remoção de áreas sobrepostas.

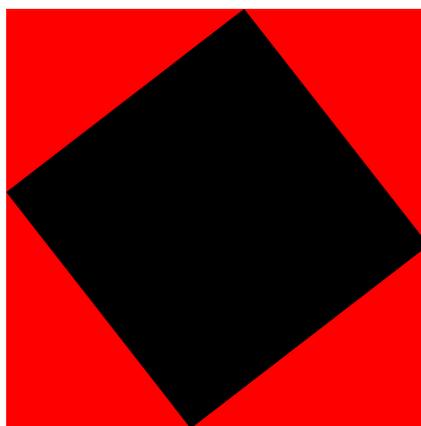


Figura 2. Índice com identificação dos valores NaN.

3.2.2 Criação da tabela CSV

A Tabela 1 contém uma amostra dos valores dos índices no bloco 10 do terceiro voo. A tabela completa é composta por 4.723.527 linhas.

Linha	Coluna	NDVI	SAVI	SCCCI	GLI	MCARI	NDRE
0.0	1716.0	0.877303	0.713112	0.613917	0.322856	0.509550	0.538592
1.0	1715.0	0.890129	0.738010	0.585658	0.309897	0.686824	0.521312
1.0	1716.0	0.887635	0.766853	0.600684	0.304492	0.304492	0.533188

Tabela 1. Valores dos índices no bloco 10 do terceiro voo

3.2.3 Seleção de dados

A Figura 3 mostra a matriz de correlação. Nela pode ser observado que os índices NDVI e NDRE estão bastante correlacionados, assim como MCARI e SAVI. Esse resultado foi semelhante para outros blocos do voo 3 e 4.

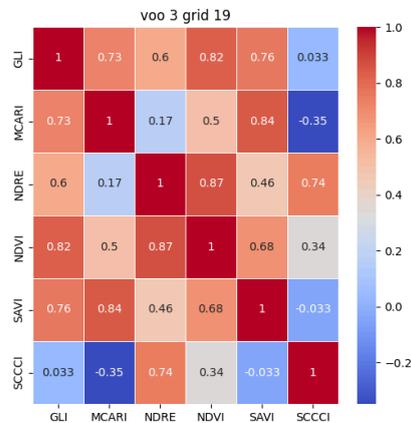


Figura 3. Matriz de correlação entre os índices de vegetação.

Fica claro, então, que não se deve utilizar o NDVI juntamente com o NDRE para evitar possíveis vieses nos resultados finais, assim como SAVI e MCARI. Por isso, foram processados os índices do conjunto GLI, SAVI, NDRE e SCCCI.

3.3 Definição de mapas de zonas de manejo

Foi executado o Fuzzy c-means com todos os parâmetros definidos e obteve-se o resultado da Figura 4. Esse mapa dividiu os grupos entre valores que representam a

planta e valores que representam o solo, o que não ajuda na identificação de falhas na plantação. Por isso foi redefinido o número de grupos.

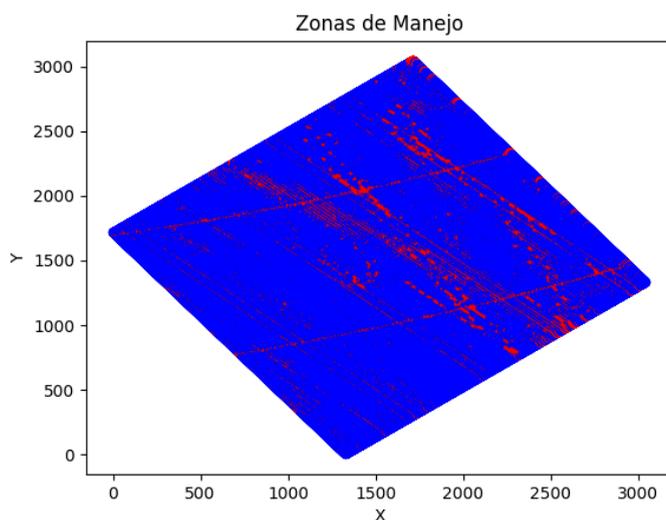


Figura 4. Zona de manejo do bloco 23 do terceiro voo.

3.3.1 Definição da quantidade ideal das ZMs

Com os resultados do FPC obtidos para valores de 2 até 8 ZMs, o que obteve o melhor resultado, foi para duas zonas de manejo. A Tabela 2 mostra o resultado do FPC para os blocos 23 e 19, um com maior e outro com menor produtividade respectivamente.

Número de grupos	FPC Bloco 23	FPC Bloco 19
2	0.80	0.76
3	0.70	0.69
4	0.62	0.59
5	0.55	0.53
6	0.49	0.47
7	0.44	0.42
8	0.40	0.38

Tabela 2. FPC de dois blocos para diferentes números de grupos

De acordo com os valores observados na Tabela 2, o número de grupos escolhidos a princípio foi de 2. Contudo a Figura 4, mostrou que utilizando dois grupos, o mapa destaca apenas falhas na área de cultivo, desta forma foram gerados mapas com 4 e 3 grupos, mostrados nas Figuras 5 e 6 respectivamente, para que possa destacar áreas com falhas e diferentes níveis de vitalidade do cultivo.

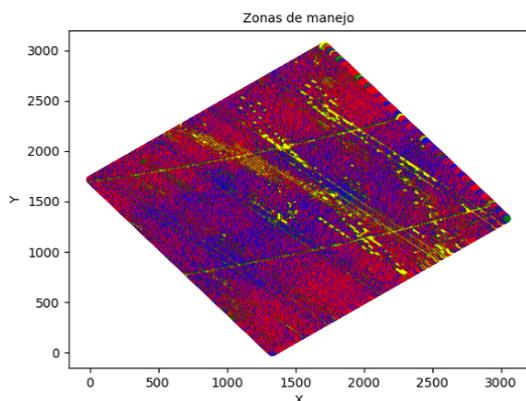


Figura 5. ZMs com 4 grupos.

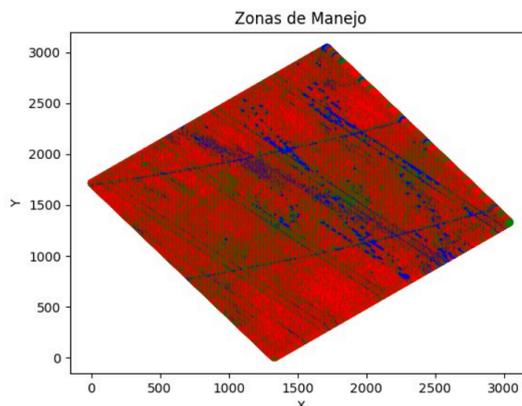


Figura 6. ZMs com 3 grupos.

De acordo com os mapas obtidos, é possível observar que na Figura 5 existem manchas na plantação que foram pintadas de azul e outras de amarelo. As amarelas representam terra e as azuis uma área com menor vitalidade. Já na Figura 6 essas manchas foram pintadas de verde e azul, em que as azuis são terra e as verdes são áreas com menor vitalidade.

As Figuras 7 e 8 apresentam os mapas das zonas de manejo em 4 diferentes blocos e considerando 4 grupos. Todos os demais parâmetros são iguais e foram citados na seção 2.3. Podemos observar que o algoritmo de agrupamento FCM, foi eficaz em mostrar falhas na plantação em alguns blocos.

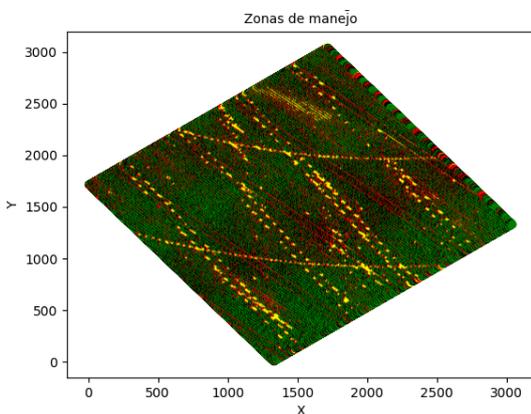


Figura 7. ZMs do bloco 10 no terceiro voo

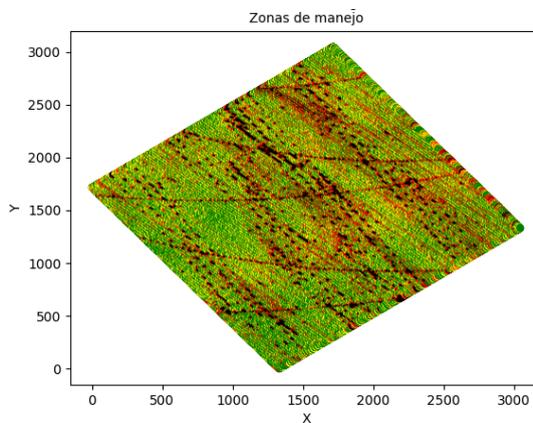


Figura 8. ZMs do bloco 11 no terceiro voo

Na Figura 7 as falhas na plantação receberam a cor amarela, enquanto que os diferentes níveis de vitalidade receberam as cores vermelho e preto, ao passo que o plantio está na cor verde. Já na Figura 8, as falhas são da cor preta, enquanto que os diferentes níveis de vitalidade receberam as cores amarelo e vermelho, ao passo que o plantio está na cor verde.

4. Conclusão

A conclusão deste estudo reforça a importância das etapas de pré-processamento de dados em análises de imagens aéreas para a agricultura de precisão. A remoção cuidadosa de valores NaN e a criação de tabelas em formato CSV surgiram como componentes fundamentais na garantia da qualidade e confiabilidade das análises subsequentes. Essas ações preparatórias não apenas simplificaram os processos analíticos, mas também contribuíram para a precisão geral dos resultados, permitindo um exame mais minucioso dos índices de vegetação.

Além disso, a determinação do número ideal de Zonas de Manejo e a análise da matriz de correlação entre os índices revelaram-se cruciais na escolha dos índices a serem empregados. Essas abordagens forneceram uma visão profunda sobre a importância de selecionar os índices mais relevantes, evitando possíveis vieses nos resultados e fornecendo diretrizes valiosas para a análise da vitalidade das culturas.

Por meio da aplicação do algoritmo Fuzzy C-Means nos índices de vegetação, este estudo alcançou com sucesso a identificação de falhas nas plantações de algodão. Inicialmente, a execução do algoritmo com dois grupos revelou-se limitada na detecção de diferentes níveis de vitalidade do cultivo. Como resposta a esse desafio, a análise foi refinada ao optar por quatro grupos, o que possibilitou uma identificação mais precisa das áreas com diferentes níveis de vitalidade do cultivo.

Essa abordagem não apenas fortalece a posição do Brasil como líder na produção de algodão, mas também gera benefícios significativos para os agricultores envolvidos. Os mapas resultantes das Zonas de Manejo forneceram informações práticas e valiosas para orientar as decisões agrícolas. Ao identificar áreas com problemas, os agricultores podem implementar ações corretivas específicas, como ajustes na irrigação, aplicação direcionada de fertilizantes e outras medidas personalizadas. Isso, por sua vez, resulta em economias substanciais de recursos e, ao mesmo tempo, no aumento da produtividade das plantações.

Referências

- Bottega, E. L. (2014). Use of site-specific management for soybean production in the Brazilian Savannah. 2014. 92 f. Tese (Doutorado em Construções rurais e ambiente; Energia na agricultura; Mecanização agrícola) - Universidade Federal de Viçosa, Viçosa.
- Nenos, L. *et al.* (2021) Machine learning in agriculture: A comprehensive updated review. *Sensors*, v. 21, n. 11, p. 3758.

- Rodrigues, L. D. S; PEREIRA, D. R. (2021) Aprendizado de Máquina Aplicado Em Imagem Ndvi Para Previsão Da Produtividade Da Cana-De-Açúcar. In: **Colloquium Exactarum**. ISSN: 2178-8332. p. 82-98.
- Devi, N. B.; Kavida, A. C. RETRACTED ARTICLE (2021): Pre-processing on remotely sensed data with unsupervised classification analysis. *J Ambient Intell Human Comput* 12, 6825–6839.
- Hopewell, K. (2013) New protagonists in global economic governance: Brazilian agribusiness at the WTO. **New Political Economy**, v. 18, n. 4, p. 603-623.
- Javed, T. *et al.* (2021) Performance and relationship of four different agricultural drought indices for drought monitoring in China's mainland using remote sensing data. **Science of the total environment**, v. 759, p. 143530.
- Li, S. *et al.* High-quality vegetation index product generation (2021): A review of NDVI time series reconstruction techniques. **International Journal of Applied Earth Observation and Geoinformation**, v. 105, p. 102640.
- Lindblom, J. *et al.* (2017). Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies. **Precision agriculture**, v. 18, p. 309-331.
- Meneses, K. C. D. (2021). Rumo à agricultura inteligente: previsão de produtividade agrícola com dados agrometeorológicos usando machine learning.
- Muruganantham, P. *et al.*(2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. **Remote Sensing**, v. 14, n. 9, p. 1990.
- Nallan, S. *et al.* (2014). Geospatial data pre-processing on watershed datasets: A GIS approach.

RUß, G. (2012). **Spatial data mining in precision agriculture**. Tese de Doutorado. Universitätsbibliothek.

Sharma, A. *et al.* (2020). Machine learning applications for precision agriculture: A comprehensive review. **IEEE Access**, v. 9, p. 4843-4873.

Santos, R. T. D. (2014). **Um modelo de referência para o processo de definição de zonas de manejo em agricultura de precisão**. Tese de Doutorado. Universidade de São Paulo.

Valente, D. S. M. (2010). Desenvolvimento de um sistema de apoio à decisão para definir zonas de manejo em cafeicultura de precisão.