

## ESTIMATIVA DE ÁREA PLANTADA NAS ÁREAS DE ALGODÃO EM SAPEZAL-MT

### *ESTIMATE OF PLANTED AREA IN COTTON AREAS IN SAPEZAL-MT*

Wisley Vítor de Sousa Gonçalves<sup>1</sup>

## RESUMO

A estimativa de área plantada e produtividade de forma precoce e precisa se torna uma importante ferramenta para economia brasileira, pois fornece tempo e informações importantes para implementação e planejamento de políticas públicas na logística de escoamento ou de armazenamento da safra desse cultivar. O uso de sensoriamento remoto, como imagens de satélite (*Landsat 8 e 9*), possui uma boa resolução temporal, assim tem sido utilizado para monitoramento da lavoura, como do algodão. O objetivo deste trabalho foi estimar a área plantada de algodão no município de Sapezal-MT durante a safra de 2017 a 2023, através de imagens de satélite *Landsat 8 e 9*, utilizando NDVI para estudo da cobertura do solo e os conceitos de *Lógica Fuzzy* para estimativa das áreas plantadas. Contudo, foi possível obter uma estimativa da área plantada bem próximo a do IBGE, sendo menor erro percentual 0,82% e maior erro percentual 17,12% acima do que foi estimado pelo IBGE. Desse modo, com base na *Lógica Fuzzy* se mostrou altamente correlacionada com o valor da área oficial, com boa precisão e nível de erro relativo aceitável.

**Palavras-chave:** Sensoriamento Remoto; Fuzzy; Clusters; Índices de Vegetação; Modelos Preditivos.

## ABSTRACT

The early and accurate estimation of planted area and productivity becomes an important tool for the Brazilian economy, as it provides timely and crucial information for the implementation and planning of public policies regarding the logistics of crop transportation or storage. The use of remote sensing, such as satellite imagery (*Landsat 8 and 9*), offers good temporal resolution and has thus been employed for crop monitoring, including cotton. The objective of this study was to estimate the planted area of cotton in the municipality of Sapezal-MT during the 2017 to 2023 crop seasons, using *Landsat 8 and 9* satellite images and employing NDVI for studying ground coverage, as well as Fuzzy Logic concepts for estimating planted areas. It was possible to obtain a planted area estimate very close to that of IBGE, with a minimum percent error of 0.82% and a maximum percent error of 17.12% above what was estimated by IBGE. Therefore, based on

---

<sup>1</sup> Bacharel em Engenharia Agrônômica pelo Instituto Federal do Triângulo Mineiro Campus Uberaba. - Contato: [wisley.goncalves@gmail.com.br](mailto:wisley.goncalves@gmail.com.br)



Fuzzy Logic, it showed a high correlation with the official area value, with good precision and an acceptable level of relative error.

**Keywords:** Remote Sensing; Fuzzy; Clusters; Vegetation Indices; Predictive Models.

## 1. INTRODUÇÃO

A cadeia do algodão vem ganhando cada vez mais destaque no agronegócio brasileiro. No ano de 2018, o Brasil registrou uma produção de aproximadamente 5 milhões de toneladas da cultura. Isso o coloca como o quarto produtor mundial, ficando atrás da China, Índia e dos Estados Unidos. As exportações de algodão em pluma geraram divisas externas, em 2019, de R\$ 10,6 bilhões (FAOSTAT, 2021). O Produto Interno Bruto (PIB) do agronegócio da cadeia do algodão, calculado pelo Centro de Estudos Avançados em Economia Aplicada, foi de R\$ 16,1 bilhões em 2017 (CEPEA, 2020). Esse é um cenário sinalizado pela Associação Brasileira dos Produtores de Algodão (Abrapa) como promissor, já que, além do País estar entre os principais produtores, é também um dos maiores consumidores desse produto (ABRAPA, 2020).

A estimativa de área plantada de um cultivar é de fundamental importância para a previsão de sua safra. Essa estimativa, no Brasil, realizada por métodos subjetivos de coleta de informações, por meio de questionários aplicados junto aos setores agrícolas do país e, por isso, possuem um custo elevado, execução demorada e estão sujeitos a imprecisões (ANTUNES et al., 2012; CONAB, 2019).

A previsão de safras é uma atividade estratégica no estabelecimento de instrumentos de políticas e negócios agrícolas, tanto para o setor público quanto para o privado, com a finalidade de fornecer informações aos formuladores dessas políticas. As previsões obtidas com antecedência suficiente à época de colheita são importantes para o planejamento do transporte, processamento e armazenamento da produção, bem como para a tomada de decisões relacionadas às exportações e importações agrícolas e ao abastecimento do mercado interno (IPPOLITI-RAMILO et al., 2003; ABRAHAM et al., 2019).

Os sistemas de previsão de safras convencionais se baseiam no produto da área ocupada do cultivar pelo rendimento médio nessa área, para a obtenção da produção final do cultivar. Os



valores de cada uma dessas variáveis podem ser obtidos por meio de censos agropecuários ou pela amostragem da população agrícola. O levantamento censitário de produção agrícola tem como objetivo fornecer uma classificação detalhada da estrutura agrícola de um país, em que o valor de cada variável para a área total do levantamento é obtido a partir dos valores das variáveis em todas as unidades de produção. No levantamento amostral, o valor de cada variável para a área total do levantamento é obtido a partir de inferências feitas sobre uma amostra da área ocupada (ASSAD et al., 2007).

Dentre os principais produtos de sensoriamento remoto utilizados para o monitoramento agrícola, está o índice espectral de vegetação (IV), que consiste em um número que representa a biomassa ou vigor do vegetal, sendo obtido pela combinação de duas ou mais bandas espectrais (MESHESHA et al., 2020). Praticamente todos os índices utilizados em sensoriamento remoto utilizam canais refletidos do visível e o do infravermelho próximo, que produzem um contraste capaz de identificar áreas vegetadas e a condição da cobertura vegetal em superfície.

Para as estimativas de safras associam-se as variáveis agronômicas, como área foliar, vigor e desenvolvimento, que podem ser utilizadas como indicadores da colheita (AL-GAADI et al., 2016). Tais características são intrinsecamente correlacionadas com as reflectâncias do dossel vegetal nos canais do visível e do infravermelho próximo, o que torna os índices espectrais de vegetação (IVs) ferramentas adequadas para avaliar o crescimento das culturas, seu acúmulo de biomassa e prever a produção.

Assim, a estimativa da produtividade das culturas pode ser obtida com relações empíricas entre a biomassa vegetal e índices espectrais de vegetação. Diversos trabalhos levantaram essas relações entre o *Normalized Difference Vegetation Index* (NDVI) que através de dados de reflectância do espectro eletromagnético, funciona como um indicador da quantidade e condição da vegetação, representando o vigor da cultura e a produção registrada em campo (MOSLEH et al., 2016; MESHESHA, 2020; SHAMMI; MENG, 2021).

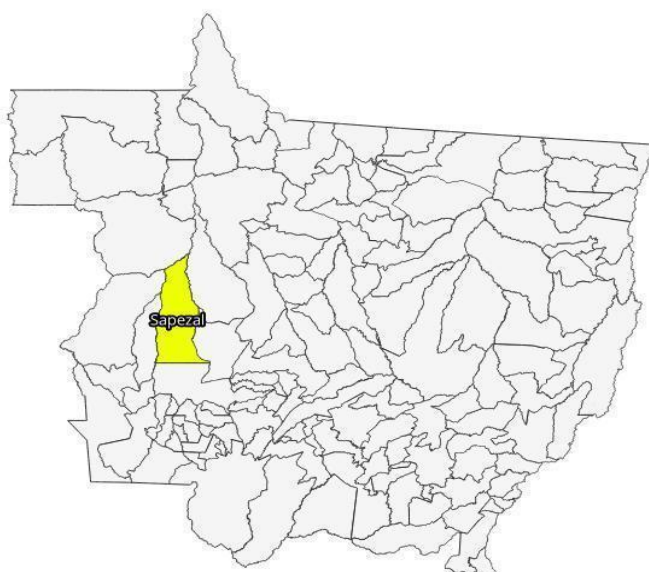
Dessa forma, o objetivo deste trabalho foi estimar a área plantada de algodão no município de Sapezal-MT durante a safra de 2017 a 2023, através de imagens de *satélite Landsat 8 e 9*, utilizando NDVI para estudo da cobertura do solo e os conceitos de *Lógica Fuzzy* para estimativa das áreas plantadas.



## 2. MATERIAL E MÉTODOS

A análise foi desenvolvida no município de Sapezal – MT (Figura 1), sendo uma área que vem a cada ano sendo destaque na produção de soja e algodão, bem como aparece entre as maiores em valor agrícola, estando entre as dez maiores segundo Ferreira (2023). Dessa forma, para estimativa da área plantada foi utilizada imagens de satélite; índice de vegetação para detectar os pontos de algodão; *Lógica Fuzzy* para estimar os parâmetros de estudo desse trabalho em cada safra.

**Figura 1.** Município de Sapezal – MT.



**Fonte:** IBGE (2022)

Nesse sentido foi utilizado as imagens do *Landsat 8* e *9*, acoplado com sensor *Operational Land Imager* (OLI) e *Operational Land Imager 2* (OLI-2), respectivamente, disponibilizado pela *United States Geological Survey* (USGS). Logo, as imagens do *Landsat 8* e *9* foram obtidas em uma resolução espacial de 30 metros para vermelho visível (0,64 a 0,67  $\mu\text{m}$ ) e infravermelho próximo (0,85 a 0,88  $\mu\text{m}$ ).

Contudo, o processamento das imagens consistiu na mosaicagem, ou seja, juntar várias imagens durante o período para cobrir o município de Sapezal – MT e o recorte da imagem



(*raster*) de acordo com extensão do limite administrativo do município, ambos os processos foram realizados no *Qgis* 3.28. As imagens foram captadas entre o período de abril de 2018 a agosto de 2023 (Tabela 1), além disso as imagens durante todo o período não foram afetadas pelas nuvens.

**Tabela 1** - Datas das obtenções das imagens dos satélites *Landsat 8 e 9* no município de Sapezal (MT), durante os anos de 2018 a 2023.

Safra	Coleta da Imagem	Intervalo
2017/2018	28/07/2018 a 13/08/2018	17 dias
2018/2019	22/07/2019 a 16/08/2019	25 dias
2019/2020	09/08/2020 a 18/08/2020	9 dias
2020/2021	12/08/2021 a 21/08/2021	9 dias
2021/2022	24/08/2022 a 31/08/2022	7 dias
2022/2023	25/07/2023 a 26/07/2023	1 dia

**Fonte:** Autoria Própria (2024)

No estudo da vegetação para detectar os pontos de algodão foi utilizado o NDVI, que foi calculado com script em Python no Google Colab, usando a biblioteca *Rasterio* (GILLIES, 2023) para manipular os *rasters*. Nesse sentido, através do NDVI é possível determinar a densidade de fitomassa foliar fotossinteticamente ativa por unidade de área. De acordo com Novo (1989), à medida que aumenta a quantidade de vegetação verde, aumenta a reflexão na banda do infravermelho próximo e diminui a reflexão na banda do vermelho fazendo com que o aumento da razão seja potencializado, realçando assim a vegetação. Dessa forma, Jensen (1996) descreve para o cálculo do NDVI a seguinte equação:

$$NDVI = \frac{NIR - R}{NIR + R}$$

Contudo, para avaliar distribuição do NDVI foi utilizado o *box-plot* e histograma. Além disso, foi utilizado estatística probabilística com a distribuição normal dos dados com estatística descritiva para avaliação da assimetria e curtose. Por fim, foi realizado a estatística descritiva dos dados, com avaliação da média e coeficiente de variação (CV%).

No entanto, para avaliação da curva normal e distribuição dos dados foi utilizado a



metodologia do Santos et al. (2019), que utiliza assimetria, curtose, média, mediana e moda como parâmetro de avaliação. Logo, se o coeficiente de assimetria  $> 0$ , a distribuição é assimétrica à direita. Se o coeficiente de assimetria  $< 0$ , a distribuição é assimétrica à esquerda e se o coeficiente de assimetria  $= 0$ , a distribuição é simétrica.

Para a curtose, que representa a dispersão, achatamento, da distribuição em relação a um padrão, geralmente a curva normal, dá-se que o coeficiente de curtose  $= 0$  a distribuição é mesocúrtica, se o coeficiente de curtose  $< 0$  mostra-se platocúrtica e curtose  $> 0$  leptocúrtica.

O controle executado pela lógica difusa imita um comportamento baseado em regras (inferências) ao invés de um controle restrito a modelos matemáticos complexos. Pode-se dizer que o objetivo da lógica difusa é gerar uma saída lógica a partir de um conjunto de entradas não precisas, com faltantes ou até mesmo ruídos (SHAW e SIMÕES, 1999). Dessa forma, antes da aplicação da Lógica Fuzzy na estimativa de área plantada, foi realizado teste do melhor número de cluster ( $c$ ) e a melhor distância do centroide ( $m$ ).

Logo, para definição do melhor número de cluster foi realizado o teste de *Elbow*. Segundo Cui (2020), o método de *Elbow* é adequado para valores de  $k$  relativamente pequenos. Contudo, esse método calcula a diferença quadrada de diferentes valores de  $k$ . À medida que o valor de  $k$  aumenta, a média da distorção diminui. Isso ocorre porque o número de amostras em cada categoria diminui, e as amostras ficam mais próximas do centro de gravidade. Conforme  $k$  aumenta, o ponto em que a redução mais significativa no efeito de melhoria da distorção ocorre é o valor de  $k$  correspondente ao *Elbow*. Dessa forma, foi realizado script em Python para realizar esse teste e foi utilizado o *KMeans* do pacote *scikit-learn* (Mueller, 2024), para realização do teste de *Elbow*.

Nesse sentido, para definir a melhor distância foi realizado *Lógica Fuzzy* com 5% da amostra total, selecionada de forma aleatória, além disso para simulação foi utilizado o melhor número de cluster em cada safra. Logo, a distância foi avaliada em um range de 1 a 6, os testes foram realizados em relação aos maiores graus de pertinência em cada safra e os critérios de avaliação foram: o *silhouette score* (varia de -1 a 1, assim mostrando a correlação dos dados em relação a distância, ou seja, valores próximos a -1 simbolizam que os pontos foram atribuídos a clusters incorretos; valores próximos a 0 simbolizam que os clusters estão sobrepostos, valores próximos a 1 simbolizam que pontos estão bem separados entre os clusters e bem próximos aos



outros pontos em seu próprio cluster); caso houvesse o score semelhante entre as distâncias definida no range em cada safra, foi utilizado métricas de erro (Erro Quadrado Médio - *MSE*, Raiz Quadrada do Erro - *RMSE*, Desvio Padrão - *STD*), para avaliar a melhor distância na *Lógica Fuzzy*.

Assim, esse teste também foi realizado no Python. Portanto, para *Lógica Fuzzy* foi realizado com pacote *scikit-fuzzy* (Warner, 2019); *silhouette score* e as métricas de erro foi realizado com o pacote *scikit-learn* (Mueller, 2024). Após os testes, foi realizada a *Lógica Fuzzy* com todos os valores de pixels definidos pelo NDVI, como pontos de algodão.

Contudo, os grupos (*clusters*) foram definidos com grau de pertinência acima de 0.999 (99% a 100%) de pertencimento, ou seja, o total de pixels de algodão que foram processadas na *Lógica Fuzzy*, foram distribuídos aos grupos que tivesse maior grau de pertinência. Logo, a distribuição dos dados foi avaliada da mesma forma da distribuição dos dados do NDVI.

Após a *Lógica Fuzzy*, foi definida a área estimada de algodão. Portanto, foi definida de duas maneiras, sendo a primeira definida através dos graus de pertinência da *Lógica Fuzzy* e a segunda definida com a quantidade de pixels do NDVI. Dessa forma, o cálculo da área estimada com os graus de pertinência, seguiu uma adaptação da metodologia descrita por França et al. (2021), que faz a estimativa de acordo com a seguinte equação:

$$A_{\text{algodão}} = \frac{A_{\text{pixel}} * \sum_{i=1}^n \mu_{ci} * 100}{255}$$

Em que  $A_{\text{akgodão}}$  = área de algodão em hectares (ha);  $A_{\text{pixel}}$  = área do pixel do satélite *Landsat 8 e 9*, igual a 900 m<sup>2</sup> (30 x 30 m);  $\mu_{ci}$  = graus de pertinência dos pixels da imagem fração referente a classe c, no caso, algodão. Em contrapartida, a estimativa de área a partir do NDVI também foi uma adequação da metodologia de França et al (2021), que multiplicou a proporção de pontos amostrais identificados como algodão ( $P_{\text{algodão}}$ ) e do número total de pontos amostrais com algodão no município de Sapezal ( $P_M$ ), pela área total do município  $A_M$ . A equação foi descrita da seguinte forma:



$$A_{\text{algodão}} = \frac{P_{\text{algodão}}}{P_M} * A_m$$

Em que:  $A_{\text{algodão}}$  = área estimada da cultura de algodão no município, em ha;  $P_{\text{algodão}}$  = número de pontos amostrais identificados como algodão;  $P_M$  = número total de pontos amostrais no município, de 0 a 100;  $A_M$  = área total do município, em hectares.

Por fim, foi calculado o erro percentual e absoluto para ambas as estimativas, que foi a métrica utilizada para classificar o nível de precisão dos modelos preditivos. Logo, as equações foram representadas da seguinte forma:

$$\%E = \frac{A_{\text{estimada}} - A_{\text{real}}}{A_{\text{real}}} * 100$$

$$|E| = |\%E|$$

Em que:  $\%E$  = erro percentual;  $A_{\text{estimada}}$  = área plantada estimada nos clusters da *Lógica Fuzzy* e NVDI;  $A_{\text{real}}$  = área plantada estimada pelo IBGE;  $|E|$  = erro absoluto;  $|\%E|$  = módulo do erro percentual. Após a estimativa do erro percentual e absoluto, foi elaborado uma classificação (Tabela 2) para descrever nível de precisão que a estimativa alcançou em relação a estimativa do IBGE. Contudo, a partir da classificação foi aplicado teste de normalidade (*Shapiro-Wilk*) e homogeneidade da variância (*Levene*), para definir entre os testes paramétrico e não paramétrico de uma amostra. Por fim, foi realizado o teste estatístico para validar o melhor modelo em cada safra, como média de 5% do erro percentual.

**Tabela 2** - Classificação do Nível de Precisão em Relação ao Erro Percentual

Erro Percentual	Classificação
< 5%	Alta Precisão
> 5% e < 15%	Regular
> 15% e < 25%	Média
> 25%	Baixa Precisão

**Fonte:** Autoria Própria (2024)



### 3. RESULTADOS E DISCUSSÃO

Os valores de NDVI quanto mais próximo de 1, mais densa é a vegetação; o valor 0 (zero) indica superfície não vegetada (ROSENDO, 2005). Segundo Boratto e Gomide (2013), os maiores valores de NDVI correspondem aos Números Digitais (ND) mais elevados, que se relacionam às áreas de vegetação com maior vigor. Enquanto os menores valores equivalem aos ND baixos, representando as áreas de vegetação estressada, bem menos densas ou até mesmo áreas desnudas.

Dessa forma, na Tabela 3 está descrito a classificação NDVI deste estudo, assim sendo a Classe 1 podendo ser descrito como solo exposto, construções urbanas e rurais e mananciais hídricos; Classe 2 a 4 sendo descrito por regiões de Reserva Legal, Área de Preservação Permanente (APP) e outros tipos de vegetação; Classe 5 são os pontos de algodão detectados entre o período de abril a setembro em cada safra descrita neste estudo.

**Tabela 3** - Classificação NDVI durante as safras 2018 a 2023

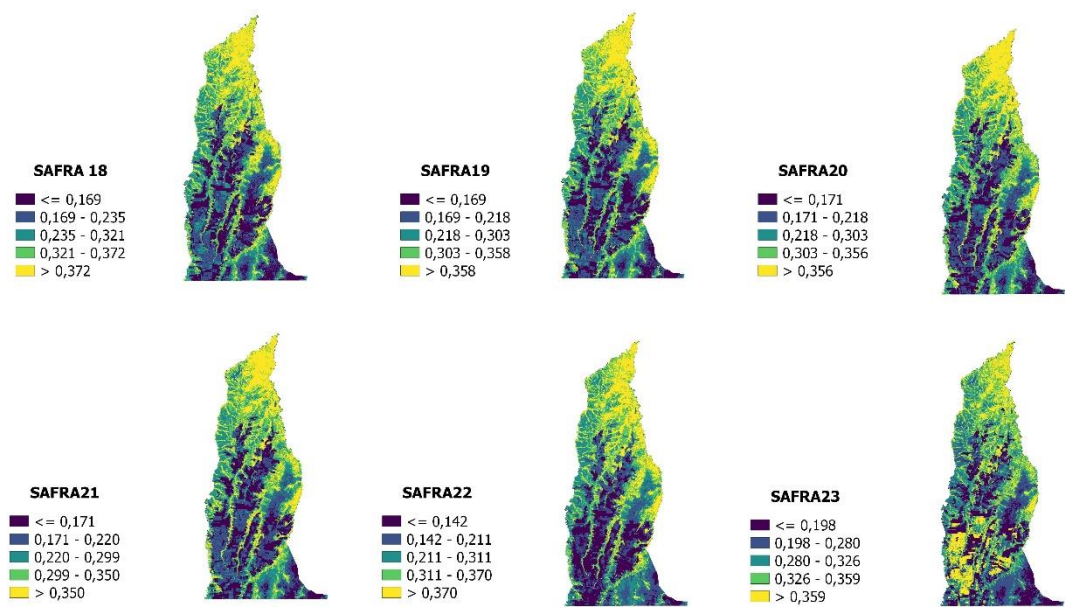
Safras	Classificação				
	1	2	3	4	5
2018	$\leq 0,169$	0,169 a 0,235	0,235 a 0,321	0,321 a 0,372	$> 0,372$
2019	$\leq 0,169$	0,169 a 0,218	0,218 a 0,303	0,303 a 0,358	$> 0,358$
2020	$\leq 0,171$	0,171 a 0,218	0,218 a 0,303	0,303 a 0,356	$> 0,356$
2021	$\leq 0,171$	0,171 a 0,220	0,220 a 0,299	0,299 a 0,350	$> 0,350$
2022	$\leq 0,142$	0,142 a 0,211	0,211 a 0,311	0,311 a 0,370	$> 0,370$
2023	$\leq 0,198$	0,198 a 0,280	0,280 a 0,326	0,326 a 0,359	$> 0,359$

**Fonte:** Autoria Própria (2024)

A Figura 2 mostra o processamento do NDVI para as safras 2018 a 2023. Contudo, pode ser verificado que os pontos com tons amarelados simbolizam presença de algodão, enquanto tons mais roxeados e azulados simbolizam presença de construções, mananciais hídricos e áreas de preservação. França et al. (2021) em estudo de estimativa de área plantada, detectou pontos de soja com uma variação de 0,199 a 0,895, no município de São Gabriel do Oeste-MS durante a safra 2018/2019.



Figura 2 – Processamento do NVDI durante as safras 2018 a 2023



Fonte: Autoria Própria (2024)

Tabela 4 - Classificação NVDI durante as safras 2018 a 2023

	Safras					
	2018	2019	2020	2021	2022	2023
Média	0,273	0,262	0,263	0,260	0,257	0,286
DP	0,096	0,090	0,089	0,085	0,105	0,082
CV%	35,17	34,15	33,93	32,69	40,94	28,65

Fonte: Autoria Própria (2024)

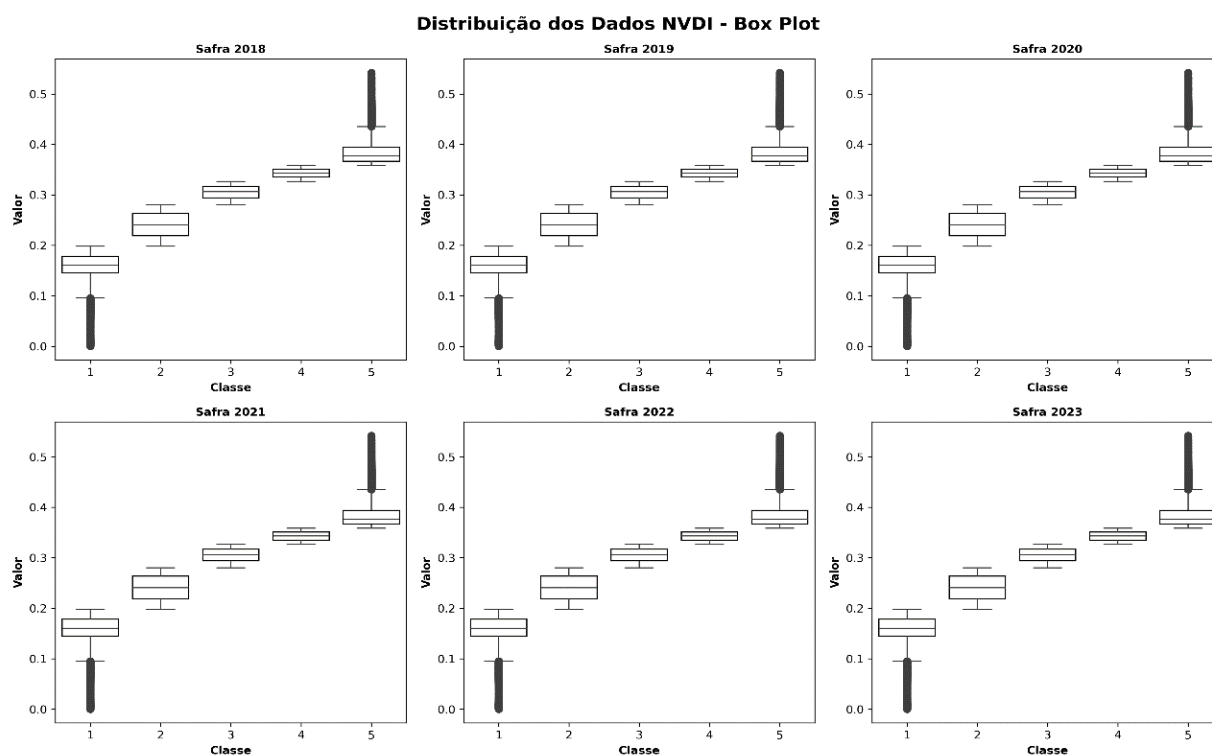
Na Tabela 4, podemos concluir que o Coeficiente de Variação (CV%) durante a safra de 2018 a 2022 (CV% de 35,17%; CV% de 34,15%; CV% de 33,93%; CV% de 32,69%; CV% de 40,94%; respectivamente) teve muito alta variabilidade, em contrapartida, na safra 2023 (CV%



de 28,65%) teve alta variabilidade.

Logo, podemos concluir que através da aplicação do NDVI e *box-plot* (Figura 3), não houve aumento significativo dos pontos de algodão entre as safras de 2017 a 2023. Além disso, houve uma baixa dispersão dos valores de NDVI, mas também concluímos que a classe que apresentou maior variação nos valores de NDVI foi a Classe 2.

**Figura 3 – Box-Plot do NVDI durante as safras 2018 a 2023**



**Fonte:** Autoria Própria (2024)

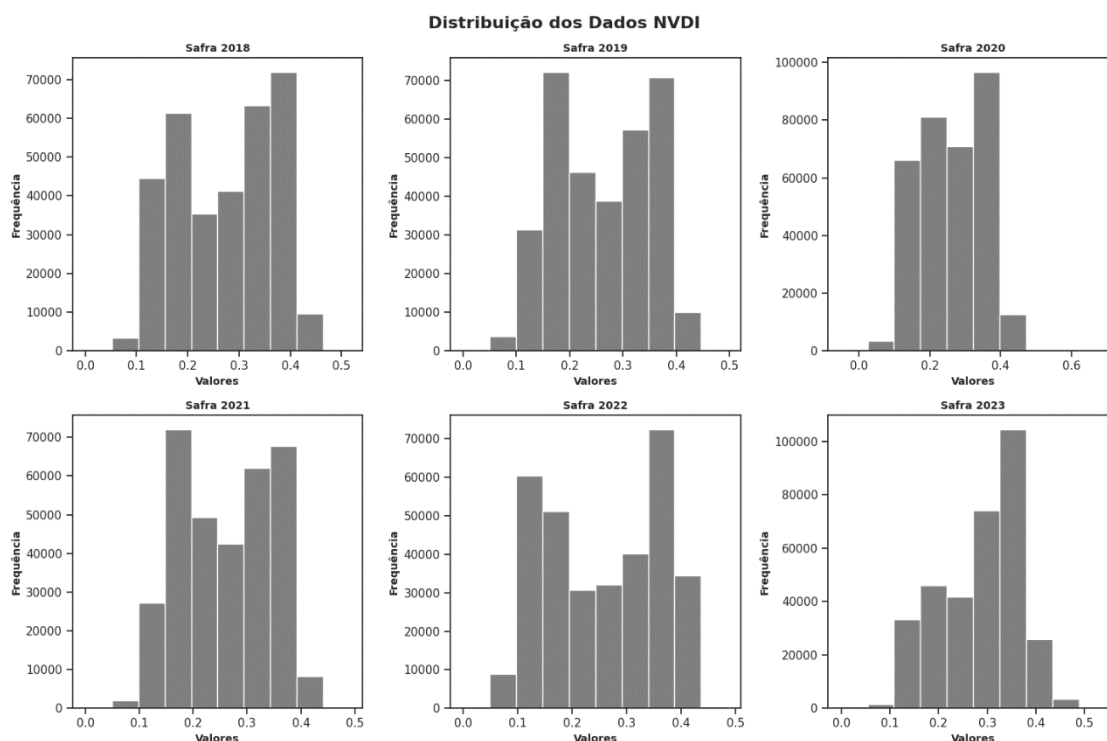
Na Figura 4, os histogramas das safras de 2018, 2020, 2022 e 2023; foram observados frequência máxima correspondentes a classe com maior valor de NDVI, assim sendo evidenciado pelas áreas com presença de algodão. Enquanto, nas safras de 2019 e 2021 houve uma frequência ligeiramente maior de valores de NDVI intermediário, dessa forma sendo evidenciado pelas classes que apresentam outros tipos de vegetação.

As análises da estatística probabilística realizada durante a safra de 2018 a 2023 (Tabela



5), permitiram constatar assimetria à esquerda da curva normal de ambos os resultados de NDVI. Logo, as diferenças entre a média e mediana foram negativas, indicando uma assimetria negativa, assim sendo representado à esquerda.

**Figura 4 – Histograma do NVDI durante as safras 2018 a 2023**



**Fonte:** Autoria Própria (2024)

**Tabela 5 – Assimetria e Curtose dos dados de NVDI durante as safras 2018 a 2023**

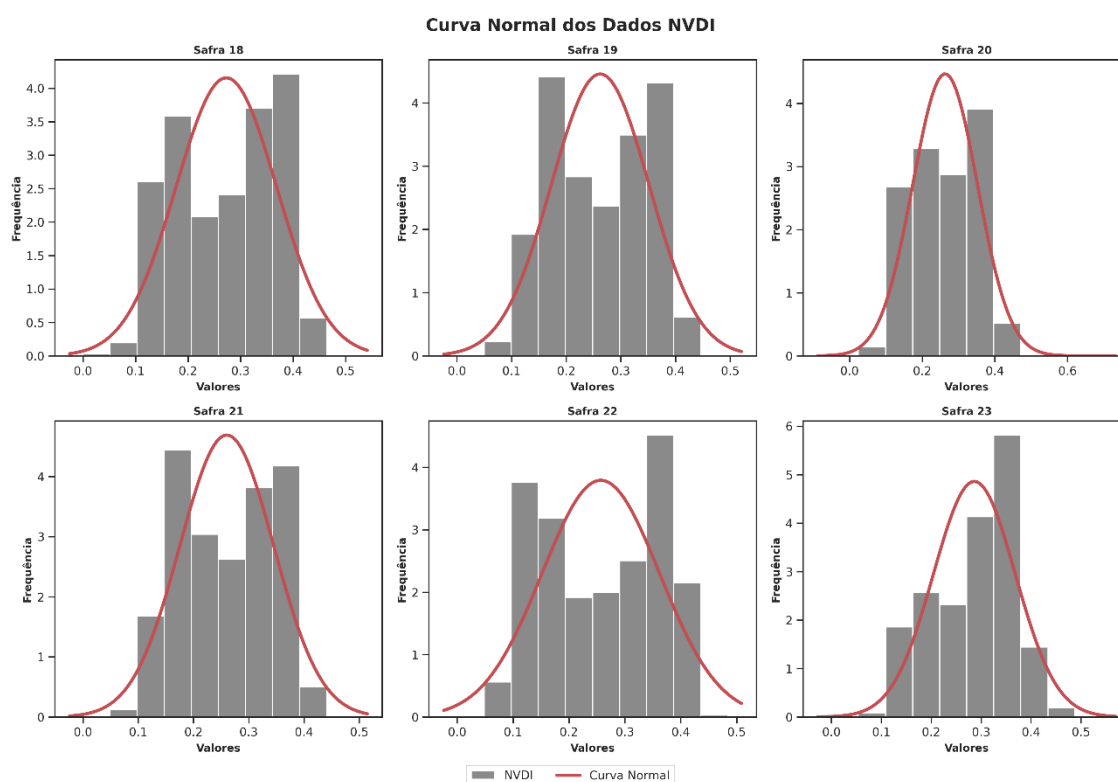
	<b>Safras</b>					
	2018	2019	2020	2021	2022	2023
<b>Média</b>	0,273	0,262	0,263	0,260	0,257	0,286
<b>Mediana</b>	0,289	0,267	0,271	0,267	0,265	0,306
<b>Moda</b>	0,333	0,333	0,333	0,333	0,333	0,333
<b>Curtose</b>	-1,304	-1,308	-1,256	-1,282	-1,443	-0,724
<b>Assimetria</b>	-0,156	-0,049	-0,048	-0,056	-0,092	-0,429

**Fonte:** Autoria Própria (2024)



A curtose calculada demonstrou que a distribuição dos dados apresentou uma distribuição platicúrtica baixa para todas as safras que foram calculadas no NDVI, indicando pouca dispersão em relação à curva normal, conforme a Figura 5. No entanto, Barros et al. (2020) realizou a geração de um mapa temático da situação da cobertura vegetal do município de Juazeiro do Norte-CE a partir do NDVI. Logo, para NDVI constatou assimetria à esquerda e negativa (-1,0678) e a curtose foi leptocúrtica (5,2487).

**Figura 5** – Curva normal dos dados do NVDI durante as safras 2018 a 2023



**Fonte:** Autoria Própria (2024)

O mapeamento dos cultivos agrícolas é realizado por meio de sensoriamento remoto. Ele tem por objetivo contribuir com a estimativa de área e de produtividade, oferecendo informações precisas sobre a distribuição geográfica em cada estado. Na estimativa de área, o resultado do mapeamento auxilia na análise da informação declarada, como um dado passível de verificação



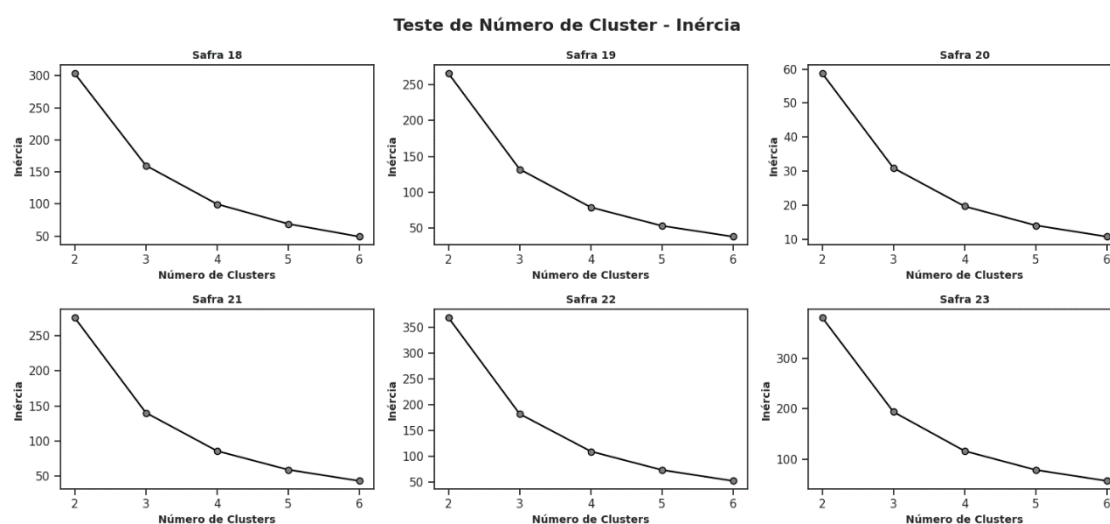
em campo. Na estimativa de produtividade, o conhecimento da localização das áreas de cultivo possibilita o monitoramento das áreas produtivas através de parâmetros agrometeorológicos, oferecendo indicativos sobre a previsão de rendimento das lavouras (CONAB).

Contudo, o levantamento dos pontos de algodão pode estar suscetível a extrapolação derivada do sensoriamento remoto. Mas também, durante a coleta das imagens definidas neste estudo, pode ter pixels de outras vegetações com o mesmo comprimento de onda do algodão.

Segundo Macedo et al. (2023), na ausência de dados *in situ*, uma forma possível de solucionar o problema consiste na comparação visual e estatística com informações de satélites de alta resolução espacial, que permitem observar a superfície terrestre com o detalhamento adequado. Apesar de permitirem ampla análise da cobertura espacial da área de interesse, os processos de mapeamento a partir de sensoriamento remoto podem conter erros sobre as áreas calculadas, por exemplo na contagem de pixels, e estão potencialmente sujeitos a vies.

Em contrapartida, no teste da *Lógica Fuzzy*, pode constatar que o melhor número de *cluster* foi 6, conforme a Figura 6, pois manteve a menor inércia. Segundo Serek et al (2020), o valor da inércia é um dos parâmetros mais importantes do *K-means*, cuja definição é o quão distantes os pontos de dados estão de seus centróides. Isso pode se manifestar como um desvio padrão em relação à média. Portanto, a conclusão disso é que quanto menor for a inércia, melhor será nosso modelo *K-means* em termos de sua divisão de *clusters*.

**Figura 6** – Teste de número de *clusters* na *Lógica Fuzzy* durante as safras 2018 a 2023

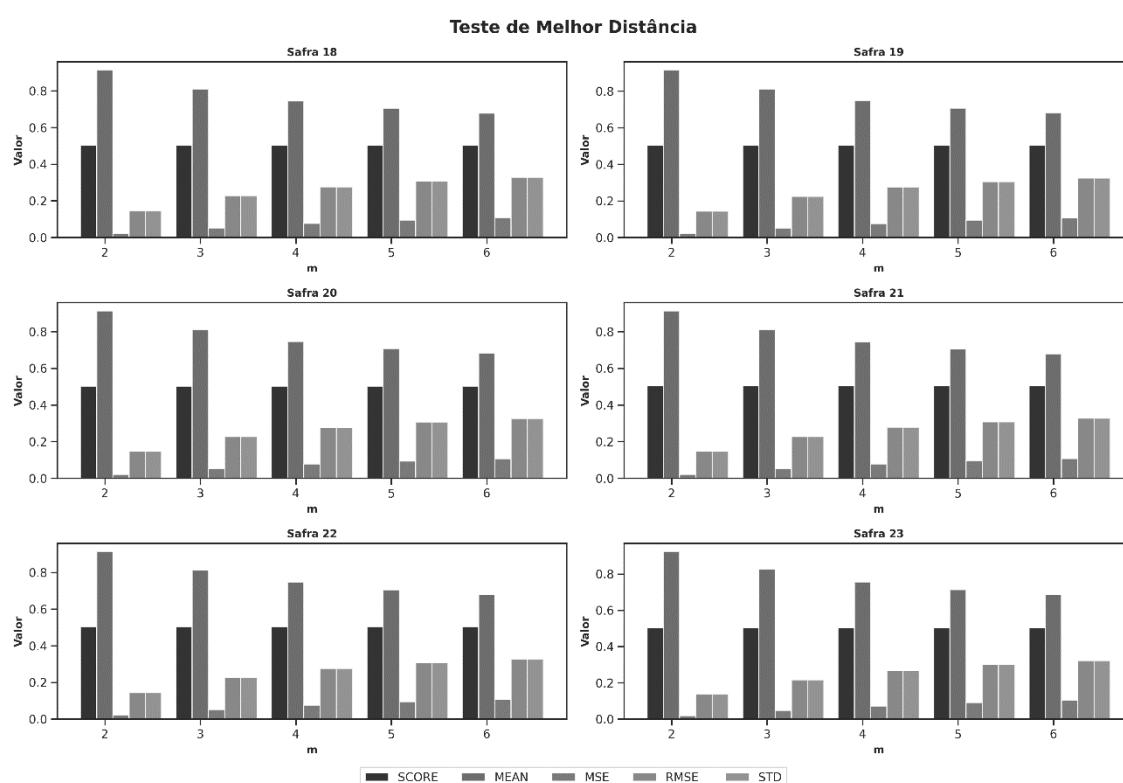




**Fonte:** Autoria Própria (2024)

Enquanto, a melhor distância ( $m$ ) foi 2 para ambas as safras no teste de melhor distância da *Lógica Fuzzy*. Portanto, *silhouette score* (*score*) foi semelhante em todas as distâncias testadas nas diferentes safras, conforme a Figura 7. Assim, para ambas as safras a média foi maior e as métricas de erro foi menor, quando distância ( $m$ ) foi 2. Belorkar & Wong (2016), em sua metodologia de estudo de expressão genética, conclui que *silhouette score* varia de -1 a 1, assim uma *silhouette score* mais alta indica uma melhor clusterização.

**Figura 7** – Teste de distância ( $m$ ) na *Lógica Fuzzy* durante as safras 2018 a 2023



**Fonte:** Autoria Própria (2024)

Após o processamento da *Lógica Fuzzy*, o coeficiente de variação (CV%) foi baixo para ambos *cluster* e safras, conforme a Tabela 6. As análises da estatística probabilística realizada também permitiram constatar assimetria à esquerda da curva normal de ambos os resultados da



*Lógica Fuzzy*. Neste sentido, a curtose calculada demonstrou que a distribuição dos dados apresentou uma distribuição platicúrtica baixa para todas as safras que foram calculadas em cada *clusters*, indicando pouca dispersão em relação à curva normal, conforme a Tabela 7.

**Tabela 6** – Assimetria na *Lógica Fuzzy* durante as safras 2018 a 2023

<i>Clusters</i>	<b>Safras</b>					
	2018	2019	2020	2021	2022	2023
<b>C1</b>	-0,624	-0,641	-0,595	-0,733	-0,687	-0,635
<b>C2</b>	-0,745	-0,483	-0,395	-0,377	-0,540	-0,626
<b>C3</b>	-0,552	-0,494	-0,789	-0,528	-0,677	-0,701
<b>C4</b>	-0,693	-0,637	-0,560	-0,580	-0,613	-0,715
<b>C5</b>	-0,768	-0,656	-0,665	-0,546	-0,228	-0,719
<b>C6</b>	-0,625	-0,639	-0,598	-0,615	-0,685	-0,559

**Fonte:** Autoria Própria (2024)

**Tabela 7** – Curtose na *Lógica Fuzzy* durante as safras 2018 a 2023

<i>Clusters</i>	<b>Safras</b>					
	2018	2019	2020	2021	2022	2023
<b>C1</b>	-1,013	-0,855	-0,939	-0,626	-0,725	-0,915
<b>C2</b>	-0,625	-1,175	-1,105	-1,332	-1,014	-0,875
<b>C3</b>	-0,977	-1,053	-0,622	-0,959	-0,877	-0,751
<b>C4</b>	-0,757	-0,742	-0,891	-1,013	-0,952	-0,620
<b>C5</b>	-0,613	-0,910	-0,808	-1,033	-1,415	-0,683
<b>C6</b>	-0,893	-0,864	-0,983	-0,823	-0,717	-0,988

**Fonte:** Autoria Própria (2024)

Portanto, as variações da *Lógica Fuzzy* (CV%) apresentaram baixa variações devido os graus de pertinência estarem entre 0,999 e 1 (99% a 100%) para ambas as safras e *clusters*. O CV constituiu-se numa estimativa do erro experimental em relação à média geral do ensaio. Considera-se que, quanto menor a estimativa do CV maior será a precisão do experimento e vice-versa, e, quanto maior a precisão, maior a qualidade experimental e menores diferenças entre estimativas de médias serão significativas (FILHO CARGNELUTTI & STORCK, 2007). A estimativa definida pela IGBE durante os anos 2017 a 2022, estão na Tabela 8.



**Tabela 8** – Área de algodão durante os anos de 2017 a 2022, em ha

Safras					
2018	2019	2020	2021	2022	2023
132.298	168.198	198.843	204.059	192.138	208.549

**Fonte:** IBGE

Na Tabela 9, está representado a estimativa de área para ambos *clusters* e safras. Logo, podemos concluir que a melhor estimativa para as safras de 2018 a 2023 foram os *clusters* c4, c3, c5, c6, c6, c3, respectivamente.

**Tabela 9** – Área de algodão estimada com a *Lógica Fuzzy* entre 2018 e 2023, em ha

Clusters	Safras					
	2018	2019	2020	2021	2022	2023
<b>C1</b>	220.235	272.117	168.000	99.882	102.000	58.941
<b>C2</b>	283.058	34.588	30.000	33.176	153.882	105.176
<b>C3</b>	98.823	172.235	112.235	229.058	268.941	181.058
<b>C4</b>	154.941	225.176	255.882	161.647	236.117	28.941
<b>C5</b>	28.235	211.058	200.470	268.235	38.117	272.823
<b>C6</b>	235.764	117.176	217.058	207,176	207.529	454.941

**Fonte:** Autoria Própria (2024)

Em contrapartida, na Tabela 10 estão representadas as estimativas de áreas com NDVI, logo podemos concluir que não houve uma boa representatividade com relação a estimativa do IBGE.

**Tabela 10** – Área de algodão estimada com NDVI entre 2018 e 2023, em ha

Safras					
2018	2019	2020	2021	2022	2023



271.901	272.210	272.164	272.091	271.893	271.814
---------	---------	---------	---------	---------	---------

**Fonte:** Autoria Própria (2024)

Conforme a metodologia descrita neste estudo, a classificação com relação ao nível de erro percentual e absoluto respectivamente, para melhores clusters durante a safra de 2018 a 2023 foi de média precisão (17,12%; c4), alta precisão (2,40%; c3), alta precisão (0,82%; c5), alta precisão (1,53%; c6), precisão regular (8,01%; c6), precisão regular (13,18%; c3), segundo a Tabela 11.

**Tabela 11** – Erro percentual e absoluto com os melhores *clusters* entre a safra de 2018 a 2023

	Safras/Cluster					
	2018/c4	2019/c3	2020/c5	2021/c6	2022/c6	2023/c3
<b>Erro Percentual (%)</b>	17,12	2,40	0,82	1,53	8,01	-13,18
<b>Erro Absoluto</b>	17,12	2,40	0,82	1,53	8,01	13,18

**Fonte:** Autoria Própria (2024)

Enquanto, para NVDI a classificação de erro percentual e absoluto foi de baixa precisão para ambas as safras (Tabela 12)

**Tabela 12** – Erro percentual e absoluto da estimativa NVDI entre a safra de 2018 a 2023

	Safras					
	2018	2019	2020	2021	2022	2023
<b>Erro Percentual (%)</b>	105,52	61,84	36,87	33,34	41,51	30,34
<b>Erro Absoluto</b>	105,52	61,84	36,87	33,34	41,51	30,34

**Fonte:** Autoria Própria (2024)

França et al (2022) em estimativa de produção de soja com *Lógica Fuzzy* e imagens de satélite no município de São Gabriel-MS, obteve uma diferença de 2,1% maior que a estimativa do IBGE.

Neste sentido, Nunes et al. (2019) aplicou uma metodologia de classificação automática, baseada em *lógica fuzzy*, utilizando imagens AVHRR-NOAA para estimar a área plantada da cultura da soja no nível de sub-píxel. Logo, os resultados obtidos com a pesquisa foram bastante



satisfatórios, visto que os valores obtidos ficaram muito próximos dos valores obtidos pelo IBGE para a safra de soja 2013/2014, em que a produtividade permaneceu constante 58 sacas/ha, mas a área cultivada com soja, de 120.000ha, ficou um pouco abaixo da safra 2018/2019, que foi de 122.511,8ha.

Contudo, o teste de normalidade para estimativa de área plantada com *clusters* apresentaram  $p\text{-valor} > 0,05$ , logo constando normalidade dos dados. Enquanto, o teste de homogeneidade apresentou heterocedasticidade regular (não homogeneidade dos dados), com  $p\text{-valor} < 0,05$  para ambos os *clusters*, assim o teste estatístico utilizado foi o paramétrico (*Teste t*) (Tabela 13).

**Tabela 13** – Teste de Homogeneidade e Normalidade os clusters (*Lógica Fuzzy*)

	Safras					
	2018	2019	2020	2021	2022	2023
<b>Shapiro -Wilk (<math>p\text{-valor}</math>)</b>	0,8306	0,8104	0,7253	0,8220	0,8398	0,4512
<b>Shapiro -Wilk (estatístico)</b>	0,9614	0,9588	0,9481	0,9603	0,9626	0,9122
<b>Levene (<math>p\text{-valor}</math>)</b>	0,0035	0,0140	0,0179	0,0067	0,0030	0,0165
<b>Levene (estatístico)</b>	14,48	8,83	8,01	11,57	15,18	8,26

**Fonte:** Autoria Própria (2024)

Em contrapartida, para a estimativa de área plantada com NDVI o teste utilizado foi o não paramétrico (*Teste de Wilcoxon* de uma amostra), pois o teste de normalidade apresentou  $p\text{-valor} < 0,05$  ( $p\text{-valor} = 0,041$ ), enquanto que o de homogeneidade apresentou homogeneidade dos dados, pois  $p\text{-valor} > 0,05$  ( $p\text{-valor} = 0,062$ ).

Estimativa de Área	
NDVI	<i>Lógica Fuzzy</i>



		2018	2019	2020	2021	2022	2023
<i>Teste T (p-valor)</i>	*	0,011	0,029	0,098	0,067	0,031	0,015
<i>Teste T (estatístico)</i>	*	3,926	3,020	2,034	2,332	2,986	3,619
<i>Teste Wilcoxon (p-valor)</i>	0,031	*	*	*	*	*	*
<i>Teste Wilcoxon (estatístico)</i>	0	*	*	*	*	*	*

Fonte: Autoria Própria (2024)

Por fim, com relação a área estimada de algodão através da *Lógica Fuzzy*, podemos concluir que para ambos os anos a média estimada pelo IBGE não foi diferente da média alcançada por esses estudos, assim os resultados podem ser considerados satisfatório. Em contrapartida, na estimativa de área com NDVI a mediana do estudo não foi diferente da mediana estimada pelo IBGE, porém os resultados não foram satisfatórios conforme a Tabela 14.

4. CONCLUSÃO

Os resultados obtidos nesse estudo permitiram constatar uma diferença significativa na estimativa de área com a *Lógica Fuzzy* e NDVI. Além disso, também constatou uma representação da área plantada com *Lógica Fuzzy* satisfatória, sendo menor erro percentual 0,82% e o maior erro 17,12% acima do que foi estimado pelo IBGE.

Contudo, este estudo atendeu a todos os objetivos propostos, que foi aplicação de *Lógica Fuzzy* com sensoriamento remoto (imagens satélites) para estimar área plantada no algodão.

9. REFERÊNCIAS

ABRAHAM, E. R. et al. Estimativa da produção da soja brasileira utilizando redes neurais artificiais. Rev. **Agrarian**, v.12, n.44, p.261-271, 2019.



ABRAPA – ASSOCIAÇÃO BRASILEIRA DOS PRODUTORES DE ALGODÃO. **Algodão no Brasil**. Brasília, 2020. Disponível em: <https://www.abrapa.com.br/Paginas/Dados/Algod%C3%A3o%20no%20Brasil.aspx>. Acesso em: 6 de jul. de 2020.

AL-GAADI, K. A.; HASSABALLA, A. A.; TOLA, E.; KAYAD, A. G.; MADUGUNDU, R.; ALBLEWI, B.; ASSIRI, F. *Prediction of Potato Crop Yield Using Precision Agriculture Techniques*. **PLOS ONE**, v. 11, n. 9, p. e0162219, 9 set. 2016.

ANTUNES, J. F. G. et al. Estimativa de área de soja por classificação de imagens normalizada pela matriz de erros. **Pesq. Agropec. Bras.**, v.47, n.9, p.1288-1294, 2012.

ASSAD, E.D. et al. Sistema de previsão da safra de soja para o Brasil. *Pesq. Agropec. Bras.*, v.42, n.5, p.615-625, 2007.

BARROS, A. S.; FARIAS, L. M. de; MARINHO, J. L. A. Aplicação do Índice de Vegetação por Diferença Normalizada (NDVI) na Caracterização da Cobertura Vegetativa de Juazeiro Do Norte–CE. **Revista Brasileira de Geografia Física**, v. 13, n. 6, p. 2885-2895, 2020.

BELORKAR, A.; WONG, L. *GFS: fuzzy preprocessing for effective gene expression analysis*. **BMC bioinformatics**, v. 17, p. 169-184, 2016.

BORATTO, I. M. P.; GOMIDE, R. L. Aplicação dos índices de vegetação NDVI, SAVI e IAF na caracterização da cobertura vegetativa da região Norte de Minas Gerais. **Anais... XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE.

CEPEA – CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA. **PIB cadeias do agronegócio - 2017**. Piracicaba, 2020. Disponível em: <https://www.cepea.esalq.usp.br/br/pib-de-cadeias-agropecuarias.aspx/>. Acesso em: 6 jul. 2020.

CONAB – Companhia Nacional de Abastecimento. **Fechamento da safra 2018/19 aponta produção recorde de grãos de 242,1 milhões de t**. Disponível em: <https://www.conab.gov.br/ultimas-noticias/3042-fechamento-da-safra-2018-2019-aponta-producao-recorde-de-graos-estimada-em-242-1-milhoes-de-t>. Acesso em: 20 maio 2020.

CONAB – Companhia Nacional de Abastecimento. **MAPEAMENTOS AGRÍCOLAS**. Disponível em: < <https://www.conab.gov.br/info-agro/safras/mapeamentos-agricolas?start=30>>. Acesso em: 30 de abr. 2024.

CUI, M. et al. *Introduction to the k-means clustering algorithm based on the elbow method*. **Accounting, Auditing and Finance**, v. 1, n. 1, p. 5-8, 2020.

FAOSTAT – *Food and Agriculture Organization of United Nations*. **Data**. Roma, 2021.



Disponível em: <http://www.fao.org/faostat/en/#data/TP>. Acesso em: 5 fev. 2021.

FERREIRA, I. **Em 2022, Sorriso (MT) manteve a liderança na produção agrícola**. [Rio de Janeiro]: IBGE - Instituto Brasileiro de Geografia e Estatística, 14 set. 2023. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37894-em-2022-sorriso-mt-manteve-a-lideranca-na-producao-agricola>>. Acesso em: 14 de jan. 2024.

FILHO CARGNELUTTI, A.; STORCK, L. Estatísticas de avaliação da precisão experimental em ensaios de cultivares de milho. **Pesquisa agropecuária brasileira**, Brasília, v. 42, n. 1, p. 17-24, jan. 2007.

FRANÇA, J. F. et al. Previsão da Produção de Soja Utilizando Imagens de Satélite e Lógica Fuzzy. **Ensaio e Ciência C Biológicas Agrárias e da Saúde**, v. 25, n. 2, p. 232-238, 2021.

GILLIES, S. *rasterio* 1.3.9. **PyPI**, 19 out. 2023. Disponível em: <<https://pypi.org/project/rasterio/>>. Acesso em: 21 de jan. 2024.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Malha Municipal**. Disponível em: <<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html>>. Acesso em: 14 de jan. 2024.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Produção Agrícola Municipal**. Disponível em: < <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e-permanentes.html>>. Acesso em: 30 de abr. 2024.

IPPOLITI-RAMILO, G. A.; EPIPHANIO, J.C.N.; SHIMABUKURO, Y. E. *Landsat-5 Thematic Mapper data for pre-planting area evaluation in tropical countries*. **Int. J. Remote Sensing**, v.24, n.7, p.1521-1534, 2003.

*Landsat* 8. **USGS. United States Geological Survey**. Disponível em: < <https://www.usgs.gov/landsat-missions/landsat-8> >. Acesso em: 27 jan. 2024.

*Landsat* 9. **USGS. United States Geological Survey**. Disponível em: < <https://www.usgs.gov/landsat-missions/landsat-9> >. Acesso em: 14 jan. 2024.

JENSEN, J. R. *Introductory digital image processing: a remote sensing perspective*. 2ª. ed. Upper Saddle River: Prentice-Hall, 1996.

MACEDO, A. dos S. et al. Experimento de validação em produtos de sensoriamento remoto de monitoramento de área queimada. In: XX Simpósio Brasileiro de Sensoriamento Remoto. Florianópolis, SC. **Anais...** Florianópolis: INPE, 2023. ISSN: 978-65-89159-04-9.

MESHESHA, D. T., AHMED, M M., ABDI, D. Y., HAREGEWEYN, N. *Prediction of grass*



*biomass from satellite imagery in Somali regional state, eastern Ethiopia. Helyon*, v. 6, p. 5272, 2020.

MOSLEH, M. K.; HASSAN, Q. K.; CHOWDHURY, E. H. *Development of a remote sensing-based rice yield forecasting model. Spanish Journal of Agricultural Research*, v. 14, n. 3, p. 0907, 31 ago. 2016.

MUELLER, A. *scikit-learn* 1.4.2. **PyPI**, 09 abr. 2024. Disponível em: <<https://pypi.org/project/scikit-learn/>>. Acesso em: 30 de abr. 2024.

NOVO, E. L. M. **Sensoriamento remoto: princípios e aplicações**. Edgar Blucher, São Paulo, 1989.

NUNES, L. F. et al. Aplicação de lógica Fuzzy em imagens de satélites para a previsão da produção de safras de grãos. In: 10º Seminário de Iniciação Científica. **Anais...** Kroton, 2019. ISSN: 2237-8901.

PIMENTEL-GOMES, F. P. **Curso de Estatística Experimental**. Nobel: São Paulo, 1985. 467p.

ROSENDO, J. dos S. **Índices de Vegetação e Monitoramento do uso do solo e cobertura vegetal na Bacia do rio Araguari -MG - utilizando dados do sensor Modis**. 2005. 130 p. Dissertação (Mestrado em Geografia e Gestão do Território) - Programa de Pós-Graduação em Geografia, Universidade Federal de Uberlândia, Uberlândia. 2005.

SANTOS, M. A. G dos; PEREIRA FILHO, W.; KUPLICH, T. M. Correlação entre Índices de Vegetação Gerados a partir de Dados de Espectrorradiômetro. In: XX Simpósio Brasileiro de Sensoriamento Remoto. Santos, SP. **Anais...** Santos: INPE, 2019. ISSN: 978-85-17-00087-3.

SHAMMI, S. A.; MENG, Q. *Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. Ecological Indicators*, v. 121, p. 107124, 1 fev. 2021.

SHAW, I. S.; SIMÕES, M. G. **Controle e Modelagem Fuzzy**. São Paulo: Edgard Blücher, 1999.

TALASBEK, A. et al. *Personality classification experiment by applying k-means clustering. International Journal of Emerging Technologies in Learning (iJET)*, v. 15, n. 16, p. 162-177, 2020.

WARNER, J. *scikit-fuzzy* 0.4.2. **PyPI**, 14 nov. 2019. Disponível em: <<https://pypi.org/project/scikit-fuzzy/>>. Acesso em: 30 de abr. 2024