Universidade Federal Rural do Semi-Árido
Pró-Reitoria de Pesquisa e Pós-Graduação
https://periodicos.ufersa.edu.br/index.php/caatinga
ISSN 1983-2125 (online)

http://dx.doi.org/10.1590/1983-21252026v3915515rc

Scientific Article

Soybean yield prediction and management zone delineation using vegetation indices

Predição da produtividade da soja e delimitação de zonas de manejo utilizando índices de vegetação

Sálvio N. S. Arcoverde[1]*, Jorge W. Cortez[1], Lourenço Q. Scalon[1], Rodrigo Y. P. Marubayashi[2], Fernando M. P. Mendieta[3]

[1]Universidade Federal da Grande Dourados, Dourados, MS, Brazil. [2]Department of Statistics, Universidade Estadual de Londrina, Londrina, PR, Brazil. [3]Instituto Federal de Mato Grosso do Sul, Ponta Porã, MS, Brazil.

ABSTRACT – Vegetation indices enable the estimation of soybean yield, contributing to crop monitoring and decision-making in precision agriculture systems. The objective of this study was to evaluate the spatial and temporal variability of NDVI and NDRE vegetation indices across different soybean growing seasons, aiming at grain yield estimation, management zone delineation, and the development of specific predictive models. The study was conducted in a 122.65 ha commercial field located in Ponta Porã, Mato Grosso do Sul, Brazil, during the 2019/2020 to 2024/2025 growing seasons. Yield data obtained from a combine harvester yield monitor and Sentinel-2 satellite imagery were used to calculate NDVI and NDRE. The indices were correlated with yield using simple and multiple linear regression models, which were evaluated based on the coefficient of determination (R^2), root mean square error (RMSE), and normalized RMSE (nRMSE). Both indices showed significant correlations with yield; however, NDRE exhibited superior predictive performance. Multiple linear regression models improved soybean yield prediction accuracy by integrating vegetation indices. NDRE showed a strong relationship with yield across management zones, with better performance in Zone 2 (R^2 = 0.76; nRMSE = 7.7%) than in Zone 1 (R^2 = 0.67; nRMSE = 7.8%), indicating high predictive accuracy in both zones. The results demonstrate the potential of NDRE for soybean yield estimation, management zone delineation, and the development of predictive models applicable to precision agriculture.

Keywords: *Glycine max* (L.) Merrill. Remote sensing. Sentinel-2.

RESUMO - Índices de vegetação permitem estimar a produtividade da soja, contribuindo para o monitoramento das lavouras e a tomada de decisão em agricultura de precisão. Objetivou-se avaliar a variabilidade espacial e temporal dos índices de vegetação NDVI e NDRE em diferentes safras de soja, visando à estimativa da produtividade de grãos, à delimitação de zonas de manejo e ao desenvolvimento de modelos preditivos específicos. O estudo foi conduzido em um talhão comercial de 122,65 ha localizado no município de Ponta Porã, MS, durante as safras 2019/2020 a 2024/2025. Foram utilizados dados de produtividade obtidos por monitor de colheita e imagens Sentinel-2 para o cálculo dos índices NDVI e NDRE. Os índices foram correlacionados com a produtividade por meio de regressões lineares simples e múltiplas, sendo os modelos avaliados pelos coeficientes de determinação (R^2), raiz do erro quadrático médio (RMSE) e RMSE normalizado (nRMSE). Ambos os índices apresentaram correlações significativas com a produtividade, porém o NDRE apresentou melhor desempenho preditivo. Os modelos de regressão múltipla aumentaram a capacidade preditiva de produtividade da soja ao integrar índices de vegetação. O NDRE apresentou elevada relação com a produtividade nas zonas de manejo, com melhor desempenho na Classe 2 (R^2 = 0,76; nRMSE = 7,7%) comparado à Classe 1 (R^2 = 0,67; nRMSE = 7,8%), indicando alta precisão preditiva em ambas as classes. Os resultados demonstram o potencial do NDRE para estimativa da produtividade da soja, delimitação de zonas de manejo e desenvolvimento de modelos preditivos aplicáveis à agricultura de precisão.

Palavras-chave: *Glycine max* (L.) Merrill. Sensoriamento remoto. Sentinel-2.

Conflict of interest: The authors declare no conflict of interest related to the publication of this manuscript.

This work is licensed under a Creative Commons Attribution-CC-BY https://creativecommons.org/licenses/by/4.0/

Received for publication in: June 22, 2026.
Accepted in: September 25, 2026.

Editor in Chief: Aurélio Paes Barros Júnior
Section Editor: Antonio Policarpo Souza Carneiro

Data Availability: The data that support the findings of this study can be made available, upon reasonable request, from the corresponding author.

***Corresponding author**:
<salvionapoleao@gmail.com>

INTRODUCTION

Soybean is of major importance to Brazilian agribusiness and is one of the main agricultural commodities produced in the country. In recent decades, a substantial increase in the cultivated area and national soybean production has been observed, consolidating Brazil among the world's leading producers of this crop (BATISTELLA et al., 2023). Soybean is grown mainly in the Central-West and Southern regions of Brazil and plays a prominent role in both national agriculture and the Brazilian trade balance, representing the country's main agribusiness export commodity (SLAIFFER et al., 2025). In the 2023/2024 growing season, national production reached 147.3 million tonnes, with an average yield exceeding 3,200 kg ha^{-1} (CONAB, 2024). In this context, crop monitoring and early yield estimation have become important tools for agricultural planning and for assessing crop conditions throughout the growing seasons (GUSSO; ARVOR; DUCATI, 2017).

Remote sensing has emerged as a promising tool for agricultural monitoring, allowing information related to crop development to be obtained over large areas (ANDRADE et al., 2022; BATISTELLA et al., 2023). The use of orbital and suborbital sensors enables the generation of vegetation indices used to

monitor the spectral behavior of soybean throughout its production cycle (ANDRADE et al., 2022). Among the most widely used indices, the Normalized Difference Vegetation Index (NDVI) stands out, as it has been extensively applied in studies on crop monitoring and agricultural yield estimation (GUSSO; ARVOR; DUCATI, 2017; BATISTELLA et al., 2023). In addition, indices derived from the red-edge region have been used in spectral analyses related to crop development (BATISTELLA et al., 2023).

Recent studies have reported promising results regarding the use of vegetation indices for soybean yield estimation. Andrade et al. (2022) observed the potential of NDVI for predicting grain yield using orbital imagery acquired throughout the crop cycle. Similarly, Andrade Júnior et al. (2022) demonstrated the effectiveness of vegetation indices derived from aerial imagery for estimating soybean yield. Gusso, Arvor and Ducati (2017) highlighted the applicability of vegetation-index time series for forecasting soybean production before harvest, while Batistella et al. (2023) showed the good performance of models based on spectral indices for estimating crop yield.

In addition to yield estimation, vegetation indices have also been used to monitor soybean development throughout its phenological stages. Farias et al. (2023) observed a significant relationship between NDVI, biomass and nutrient uptake in soybean, demonstrating the potential of this index for monitoring plant growth and within-field variability. Damian et al. (2020) found significant correlations between NDVI values obtained from satellite imagery and yield data from annual crops, showing that this index can be used to represent the spatial variability present in agricultural fields and to support the delineation of management zones. These authors also emphasized that NDVI can complement information derived from yield maps, particularly when such data are unavailable or affected by operational limitations.

Among the various spectral attributes available from orbital imagery, bands located in the red-edge region have received increasing attention because of their high sensitivity to variations in chlorophyll content, vegetative vigor and photosynthetic activity in agricultural crops (PEREIRA et al., 2026). Thus, vegetation indices derived from these bands, such as the Normalized Difference Red Edge Index (NDRE), have the potential to complement the information provided by traditional indices, such as NDVI, particularly at more advanced stages of crop development, when spectral saturation of NDVI may occur. Recent studies have shown that incorporating information from Sentinel-2 red-edge bands can improve the capacity to monitor crop development and contribute to agricultural yield estimation at different spatial scales (CHRISTENSON et al., 2016; HUNT et al., 2019; MARSHALL et al., 2022; PEREIRA et al., 2026).

Although vegetation indices have been widely used for soybean monitoring and yield estimation, most studies have focused on individual growing seasons and have evaluated the relationship between spectral indices and yield without integrating temporal variability, management-zone delineation, and zone-specific prediction models. In this context, an important contribution of the present study is the integrated assessment of NDVI and NDRE across six consecutive soybean growing seasons, combining their temporal and spatial variability with yield-monitor data to identify the index and acquisition periods with greater predictive potential. In addition, the study advances beyond field-level prediction by using the spatial variability of yield and vegetation indices to delineate management zones and subsequently develop specific yield prediction models for each production environment. This approach allows the predictive performance of the vegetation indices to be evaluated not only at the field scale but also within spatially differentiated management zones, providing a more site-specific framework for soybean yield estimation in precision agriculture.

Therefore, the objective of this study was to evaluate the spatial and temporal variability of NDVI and NDRE across six consecutive soybean growing seasons, assessing their potential for grain yield prediction and comparing their performance across different crop development periods. Additionally, the study aimed to use the spatial variability of yield and vegetation indices to delineate management zones and develop zone-specific prediction models, thereby evaluating the potential of vegetation indices for site-specific soybean yield estimation.

MATERIAL AND METHODS

Study area, climate and soil

The study was conducted on a commercial farm in the municipality of Ponta Porã, state of Mato Grosso do Sul, Brazil (22°15′53″ S, 55°53′38″ W; mean altitude of 600 m), considering a field of 122.65 ha (Figure 1). According to the Köppen classification, the regional climate is Cwa, corresponding to a humid subtropical climate, mesothermal, with hot and rainy summers (ALVARES et al., 2013). The soil in the study area was classified as *Latossolo Vermelho distrófico* (Oxisol) according to the Brazilian Soil Classification System (SiBCS) (SANTOS et al., 2025).

Soybean was grown in the study area during the 2019/2020, 2020/2021, 2021/2022, 2022/2023, 2023/2024 and 2024/2025 growing seasons. Monthly accumulated rainfall data from September to March of each growing season indicated that the 2019/2020 season had the highest accumulated rainfall (967.6 mm), followed by 2023/2024 (711.0 mm) and 2021/2022 (612.0 mm), whereas 2020/2021 had the lowest accumulated rainfall (388.6 mm), according to the Weather and Climate Monitoring Center of Mato Grosso do Sul (CEMTEC, 2026).

Considering the edaphoclimatic conditions of the southern region of Mato Grosso do Sul, Brazil, the soybean crops were cultivated under a no-tillage system, following the management practices recommended for soybean production in the region, including fertilization, seed treatment and inoculation, as well as weed, pest, and disease control. The following cultivars were used: M 6410 IPRO in the 2019/2020 and 2020/2021 growing seasons, BMX FIBRA IPRO in the 2021/2022 and 2022/2023 growing seasons, M 6100 XTEND in the 2023/2024 growing season, and DM 65IX67 I2X in the 2024/2025 growing season. Sowing was performed between early October and early November, depending on the growing season and crop conditions, using a 0.45-m row spacing and an approximate plant population of 250,000 plants ha^{-1}.

Sowing and harvesting dates varied among growing seasons according to climatic conditions and cultivar cycle, and the days after sowing (DAS) corresponding to each growing season are presented in Table 1. Crop protection was carried out according to the technical recommendations for soybean production in the region, aiming to ensure adequate crop establishment, development, and yield.

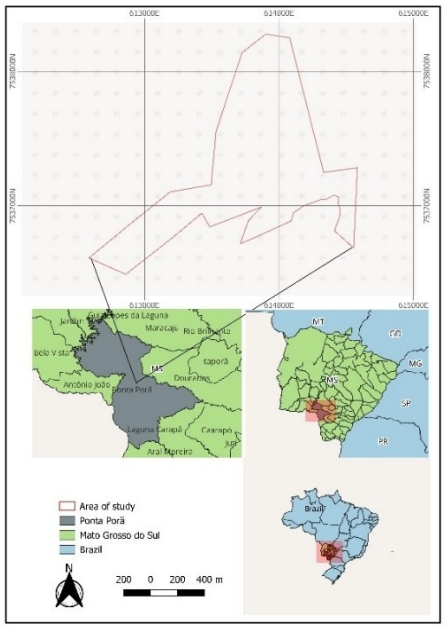


Figure 1. Location of the study area.

Table 1. Correspondence between days after sowing and soybean phenological stages on the image acquisition dates, considering the bands used to calculate the vegetation indices at a 10-m spatial resolution.

Dates	Satellites	Spatial resolution	Bands	Phenological stage	DAS
Oct 31, 2019	Sentinel 2	10 m	R/ RE/ NIR	V4–V5	46
Nov 20, 2019	Sentinel 2	10 m	R/ RE/ NIR	R2–R4	66
Nov 29, 2020	Sentinel 2	10 m	R/ RE/ NIR	V7–V8	41
Dec 09, 2020	Sentinel 2	10 m	R/ RE/ NIR	V8–V9	51
Nov 14, 2021	Sentinel 2	10 m	R/ RE/ NIR	V8–V9	39
Nov 24, 2021	Sentinel 2	10 m	R/ RE/ NIR	V10–V11	49
Nov 24, 2022	Sentinel 2	10 m	R/ RE/ NIR	V10–V11	42
Nov 29, 2022	Sentinel 2	10 m	R/ RE/ NIR	V11–V12/ BF (R1)	47
Nov 04, 2023	Sentinel 2	10 m	R/ RE/ NIR	V11–V12 / R1	36
Nov 29, 2023	Sentinel 2	10 m	R/ RE/ NIR	R1–R2	61
Oct 24, 2024	Sentinel 2	10 m	R/ RE/ NIR	R3	29
Nov 13, 2024	Sentinel 2	10 m	R/ RE/ NIR	R4	48

DAS: days after sowing. BF: beginning of flowering (R1); R: Red; RE: Red-edge; NIR: Near-infrared.

Yield and vegetation indices

The area corresponding to the selected field was harvested when soybean reached the final maturity stage (R9), using a combine harvester equipped with a mass flow sensor, monitor and GNSS receiver, with the capacity to collect data every 1 s of operation (SCHWAMBACH et al., 2021). Yield data were extracted using the manufacturer's software and subsequently processed and filtered in QGIS version 3.28.14 (QGIS, 2024), aiming to remove inconsistencies and outliers from the combine-harvester dataset (NUNES et al., 2025). The filtering procedure considered records with positioning errors and/or null geographic coordinates, grain mass moisture equal to zero, or data collected with an effective header working width lower than 95% of the nominal working width (SCHWAMBACH et al., 2021).

In addition, the raw yield-monitor data were screened to remove records associated with transient harvesting

conditions and operational anomalies. The mass-flow and grain-moisture sensors were calibrated according to the combine harvester manufacturer's recommended procedures prior to data acquisition, ensuring appropriate measurement of grain flow and moisture content during harvesting. Records obtained during headland maneuvers and at the beginning and end of harvesting passes were excluded to minimize errors associated with the grain-flow filling and emptying intervals of the combine harvester. Data showing abnormal grain-flow values, operational speeds inconsistent with normal harvesting conditions, and extreme yield values were also removed. Outliers were identified based on the distribution of the yield data and their spatial consistency (MENEGATTI; MOLIN, 2004). These procedures were adopted to reduce errors associated with combine movement, sensor response time, grain-flow instability, and spatial positioning.

After data cleaning, the yield data were interpolated using the inverse distance weighting (IDW) method. All yield data were estimated at a spatial resolution of 10 m using QGIS version 3.28.14 (QGIS, 2024). After interpolation, the data were reclassified according to the proposed legends and grouped into five classes (NUNES et al., 2025).

Vegetation indices were obtained using multispectral images from the MultiSpectral Instrument (MSI) sensor onboard the Sentinel-2A satellite (PEREIRA et al., 2026). Images corresponding to the soybean development period, between October and March of the respective growing seasons, were selected. The product used was Sentinel-2 Level-2A, which provides atmospherically corrected surface reflectance data and a spatial resolution of 10 m for the visible and near-infrared bands (PEREIRA et al., 2026). For each growing season, two images were selected (Table 1), prioritizing dates immediately prior to vegetation-index saturation and with cloud cover lower than 30% over the study area. The spectral response of the crop was quantified using the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Red Edge Index (NDRE), calculated from the red (R), red-edge and near-infrared (NIR) bands. Because the red-edge band has an original spatial resolution of 20 m, it was resampled to a 10-m pixel size, matching the spatial resolution of the red and near-infrared bands, to ensure spatial compatibility among the bands used to calculate the vegetation indices.

NDVI was calculated using Equation 1, according to Slaiffer et al. (2025). NDVI values can range from -1 to 1, with values close to 1 indicating the presence of photosynthetically active vegetation, whereas lower values or values close to zero indicate exposed soil (SLAIFFER et al., 2025).

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \qquad (1)$$

where NIR is the reflectance in the wavelength range corresponding to the near-infrared region (0.76 to 0.90 µm), and R is the reflectance in the wavelength range corresponding to the red region (0.63 to 0.69 µm).

NDRE was calculated using Equation 2, according to Slaiffer et al. (2025).

$$\mathrm{NDRE} = \frac{\mathrm{NIR} - \mathrm{RedEdge}}{\mathrm{NIR} + \mathrm{RedEdge}} \qquad (2)$$

where RedEdge is the reflectance in the red-edge band (0.715 µm).

Geospatial analyses and processing of the harvest data and vegetation indices were performed in QGIS Firenze, version 3.28.14 (QGIS, 2024). The yield maps, originally in vector format, were converted to raster format, enabling the extraction of yield values at each sampling point. Vegetation-index (VI) values were obtained using the zonal statistics tool, considering the spatial resolution of the processed Sentinel-2 image pixels (10 m), so that each sampling unit was associated with the mean spectral value corresponding to the pixel area.

Yield and vegetation-index data obtained from the spatially distributed sampling units within the experimental field were exported in .XLS format for subsequent statistical analyses and adjustment of the predictive models. A total of 12.265 valid observations were used in the statistical analyses. Modelling was performed using all sampling units, whereas model validation was conducted using a subset corresponding to 50% of the data, defined based on the median of the sampling dataset.

Statistical analyses, modelling and validation

For the statistical analyses, each spatially distributed sampling unit was considered one observation. Yield and vegetation-index data were spatially paired at the same sampling locations, with yield values obtained from the 10-m IDW-interpolated yield surface and NDVI and NDRE values extracted from the corresponding Sentinel-2 image pixels using zonal statistics. Therefore, the Pearson correlations and regression analyses were performed using spatially paired yield and vegetation-index observations. The number of valid observations (n) used in each analysis corresponded to the number of spatially paired sampling units remaining after data filtering.

Initially, descriptive statistical analysis was performed for grain yield and vegetation-index data. Subsequently, the samples/dates that showed positive and significant Pearson correlations ($p \leq 0.05$) were subjected to linear regression analysis between grain yield data, as the dependent variable, and the spectral data from the NDVI and NDRE indices, as independent variables, aiming to obtain equations for estimating soybean yield in the experimental field. The regression models were selected based on the significance of the regression coefficients, the regression F-test and the coefficient of determination (R^2). Models with R^2 values greater than 0.50 were considered adequate.

Regression models with significant coefficients of determination were validated using the dataset reserved for model validation. At this stage, the coefficient of determination (R^2) (Equation 3), the root mean square error (RMSE) (Equation 4), expressed in kg ha^{-1}, and the normalized root mean square error (nRMSE) (Equation 5), expressed as a percentage, were evaluated. This same evaluation procedure was used by Batistella et al. (2023) and Pereira et al. (2026).

$$R^2 = \frac{(\Sigma(\hat{Y}_i - \bar{Y})^2)}{(\Sigma(Y_i - \bar{Y})^2)} \quad (3)$$

$$RMSE = \frac{\sqrt{[\Sigma(Y_i - \hat{Y}_i)^2}}{n} \quad (4)$$

$$RMSE = \frac{RMSE}{\bar{Y}} x\ 100 \quad (5)$$

where:
Y_i = observed value;
$\hat{Y}_i$ = value estimated by the model;
$\bar{Y}$ = mean of the observed values;
n = number of observations.

Additionally, multiple linear regression was performed to predict soybean yield by growing season, using NDVI and NDRE as independent variables. For each growing season, the image with the best performance in the simple linear regression models was selected, based on the complementarity approach of spectral indices, aiming to increase explanatory power and reduce estimation errors (PEREIRA et al., 2026).

Geostatistics and ordinary kriging

Geostatistical analyses of grain yield and the NDVI and NDRE vegetation indices were performed to characterize the spatial variability of the evaluated growing seasons and to support the delineation of management zones. Experimental semivariograms were calculated, and spherical, exponential and Gaussian theoretical models were fitted. The quality of the fits was assessed using the slope coefficient of the cross-validation regression, seeking values close to 1, as well as the lowest root mean square error (RMSE) values. After defining the geostatistical models, spatial interpolation was performed by ordinary kriging using the nugget effect, sill and range parameters of the selected models (SCALON et al., 2026).

Interpolation was performed on a regular grid with a spatial resolution of 10 × 10 m, adopting the WGS 84 / UTM zone 21S coordinate reference system (EPSG:32721). The interpolated maps of yield, NDVI and NDRE obtained for each growing season were used to analyse spatial variability and subsequently delineate management zones.

Delineation of management zones

Clustering was performed using a fuzzy classification algorithm in the Smart-Map plugin (PEREIRA et al., 2022), within the QGIS environment, using interpolated data processed in the management-zone module based on multivariate statistics. The number of zones was defined by evaluating a pre-established range, with the optimal value selected based on the lowest Fuzziness Performance Index (FPI) and Normalized Classification Entropy (NCE) values, without compromising the interpretability of the zones (SCALON et al., 2026).

Subsequently, linear regression models between grain yield and vegetation indices were fitted for each management zone, aiming at prediction in site-specific environments. Only statistically significant models with R^2 greater than 0.50 were considered. Validation was performed using R^2, RMSE and nRMSE, with 50% of the sample data defined by the median of the dataset, as previously described.

RESULTS AND DISCUSSION

According to the classification proposed by Pimentel-Gomes and Garcia (2002), coefficient of variation (CV) values can be considered low when below 10%, moderate when between 10 and 20%, high when between 20 and 30%, and very high when above 30%. For NDVI, a moderate CV was observed on 31 October 2019, 14 November 2021, 24 November 2022 and 29 November 2022 (Table 2). Similarly, moderate CV values were observed for NDRE on 9 December 2020 and 24 November 2021. On the remaining dates, both NDVI and NDRE showed high or very high CV values.

Mean NDVI values tended to be higher than mean NDRE values (Table 2), except on 20 November 2019, 9 December 2020, 24 November 2021 and 24 October 2024. This indicates that NDVI detects crop biomass accumulation earlier, while also reaching spectral saturation earlier, as observed, for example, on 31 October 2019. Even when mean NDVI reached 0.82, mean NDRE did not exceed 0.60.

The correlations between yield and the NDVI and NDRE indices were positive and significant on all assessment dates, with values greater than 0.73, indicating a strong association between the spectral indices and grain yield (Table 2). According to Jung et al. (2025), Pearson correlation coefficients from 0.70 to 0.89 indicate strong correlation, whereas values from 0.90 to 1.00 indicate very strong correlation.

All image acquisition dates showed significant correlations with yield; therefore, linear equations were fitted to estimate soybean yield, and the best-performing models are presented in Table 3. All dates showed high coefficients of determination, with NDVI and NDRE explaining more than 70.0% of the yield variation in the studied field, except for NDVI on 20 November 2019, which explained 58.0% of the variation in grain yield. These results are similar to those reported by Andrade et al. (2022), who used NDVI images acquired during the late vegetative stage and different reproductive stages and correlated them with yield maps, observing a maximum coefficient of determination of 88.0% using Sentinel-2 satellite imagery.

The coefficient of determination (R^2), root mean square error (RMSE), and normalized RMSE (nRMSE) were used for validation of the developed models. Overall, NDRE showed better performance in estimating soybean yield, with a mean RMSE of 176.84 kg ha^{-1} and a mean nRMSE of 5.90%, compared with NDVI, which showed a mean RMSE of 223.00 kg ha^{-1} and a mean nRMSE of 7.22%. The 2019/2020 growing season is noteworthy, in which the NDRE -based model showed an R^2 of 0.98 and RMSE of 79.83 kg ha^{-1}, whereas the NDVI-based model showed an R^2 of 0.58 and RMSE of 432.48 kg ha^{-1}. These results indicate the greater capacity of NDRE to estimate soybean yield across the evaluated growing seasons.

Table 2. Mean values, standard deviations and coefficients of variation (CV, %) of the NDVI and NDRE vegetation indices, and Pearson correlation coefficients (r) between vegetation indices and soybean yield on different assessment dates/days after sowing (DAS).

Parameters	NDVI	NDRE	NDVI	NDRE
	31 October 2019 (46 DAS)		20 November 2019 (66 DAS)	
Mean	0.82 ± 0.099	0.20 ± 0.13	0.31 ± 0.18	0.51 ± 0.12
C.V. (%)	12.16	66.98	56.16	23.52
r	0.95**	0.77**	0.76**	0.99**
	29 November 2020 (41 DAS)		9 December 2020 (51 DAS)	
Mean	0.70 ± 0.15	0.40 ± 0.11	0.40 ± 0.11	0.60 ± 0.08
C.V. (%)	21.29	27.5	27.4	13.21
r	0.99**	0.96**	0.96**	0.96**
	14 November 2021 (39 DAS)		24 November 2021 (49 DAS)	
Mean	0.82 ± 0.099	0.31 ± 0.09	0.46 ± 0.11	0.48 ± 0.06
C.V. (%)	12.16	28.60	24.48	13.58
r	0.84**	0.93**	0.93**	0.91**
	24 November 2022 (42 DAS)		29 November 2022 (47 DAS)	
Mean	0.69 ± 0.13	0.37 ± 0.11	0.67 ± 0.08	0.48 ± 0.10
C.V. (%)	19.11	29.90	11.90	21.44
r	0.93**	0.85**	0.92**	0.93**
	4 November 2023 (36 DAS)		29 November 2023 (61 DAS)	
Mean	0.56 ± 0.14	0.31 ± 0.13	0.44 ± 0.18	0.30 ± 0.13
C.V. (%)	26.15	42.50	39.83	42.50
r	0.93**	0.85**	0.86**	0.86**
	24 October 2024 (29 DAS)		13 November 2024 (48 DAS)	
Mean	0.33 ± 0.19	0.49 ± 0.14	0.70 ± 0.15	0.49 ± 0.11
C.V. (%)	55.84	70.95	20.76	22.44
r	0.75**	0.73**	0.99**	0.99**

** Significant at the 1% probability level by the t-test.

Table 3. Linear regression equations and coefficients of determination (R^2) for estimating soybean yield from NDVI and NDRE indices on different image acquisition dates.

Image date	Linear equation	R^2	RMSE (kg ha^{-1})	nRMSE (%)
20 November 2019 (66 DAS)	2654.89*NDVI + 3111.62	0.58**	432.48	12.5
	5015.81*NDRE + 1377.27	0.98**	79.83	2.3
9 December 2020 (51 DAS)	5333.60*NDVI + 1786.14	0.94**	90.36	2.6
	7324.88*NDRE – 479.11	0.92**	173.19	5.5
14 November 2021 (39 DAS)	3464.49*NDVI – 650.10	0.71**	176.36	9.4
	4259.31*NDRE + 864.79	0.87**	136.66	7.3
24 November 2022 (42 DAS)	5619.63*NDVI + 521.02	0.87**	319.10	8.7
	6199.38*NDRE + 2108.37	0.73**	299.81	8.2
4 November 2023 (36 DAS)	2729.78*NDVI + 1645.99	0.87**	150.05	5.3
	2783.94*NDRE + 2311.12	0.73**	239.79	8.4
13 November 2024 (48 DAS)	5157.00*NDVI + 564.86	0.97**	169.63	4.8
	6899.46*NDRE + 792.19	0.98**	131.74	3.7

** Significant at the 1% probability level by the t-test. DAS: days after sowing; R^2: coefficient of determination; RMSE: root mean square error; nRMSE: normalized root mean square error.

RMSE was used to quantify the deviations between observed and estimated values, whereas nRMSE allowed predictive performance to be compared in percentage terms (MAIMAITIJIANG et al., 2020). According to Silva et al. (2018), nRMSE values below 10% indicate excellent performance. Thus, the models evaluated in this study showed predominantly excellent performance. Similar results were reported by Andrade et al. (2022), who obtained a mean nRMSE of 4.34% and RMSE of 158.4 kg ha^{-1} for soybean yield estimation.

The models showed high coefficients of determination, explaining at least 88% of the yield variation (Table 4). The low nRMSE values (3.0%) and RMSE values (< 100 kg ha^{-1}) indicate excellent model performance (SILVA et al., 2018). Furthermore, the results were superior to those obtained by Stepanov et al. (2020), who reported an RMSE of 130 kg ha^{-1} for soybean yield estimation, reinforcing the efficiency of the combined use of vegetation indices in predictive models (MERCANTE et al., 2010).

Table 4. Multiple linear regression equations fitted for estimating soybean yield using the NDVI and NDRE indices.

Image date	Linear equation	R^2	RMSE (kg ha^{-1})	nRMSE (%)
20 Nov. 2019 (66 DAS)	378.40*NDVI + 4663.20*NDRE + 1444.40	0.98**	104.00	3.0
9 Dec. 2020 (51 DAS)	2895.20*NDVI + 3713.40*NDRE + 527.20	0.97**	105.00	3.0
14 Nov. 2021 (39 DAS)	1653.31*NDVI + 3121.84*NDRE – 132.17	0.96**	72.33	3.9
24 Nov. 2022 (42 DAS)	5227.05*NDVI + 669.84*NDRE + 540.52	0.88**	294.44	8.0
4 Nov. 2023 (36 DAS)	6644.66*NDVI – 4427.36*NDRE + 826.68	0.94**	100.10	3.5
13 Nov. 2024 (48 DAS)	-6230.62*NDVI + 15157.71*NDRE +1119.66	0.99**	100.59	2.8

** Significant at the 1% probability level by the t-test. DAS: days after sowing; R^2: coefficient of determination; RMSE: root mean square error; nRMSE: normalized root mean square error.

Figure 2 shows the NDVI maps corresponding to the dates that provided the best results for estimating soybean yield. The map obtained on 9 December 2020 (51 DAS) (Figure 2B) showed high values and a relatively homogeneous spatial distribution, possibly because the crop was between the R1 and R2 stages, characterized by flowering and canopy closure. Under this condition, vegetation cover reduces the influence of soil on the spectral response and may favor NDVI saturation, limiting its capacity to discriminate small differences in plant vigor (CARNEIRO et al., 2020).

In contrast, the map obtained on 20 November 2019 (66 DAS) (Figure 2A) showed greater spatial variability in NDVI values, with contrasting regions of vegetative vigor. During this period, the crop was probably between the R3 and R4 stages, a phase of intense pod formation and development, when differences related to yield potential, water availability and soil attributes tend to be more strongly expressed in the canopy. Although this growing season had the highest accumulated rainfall during the crop cycle, the greater spatial variability observed suggests that the phenological stage had a stronger influence on the spectral response than total rainfall. According to Mercante et al. (2010) and Carneiro et al. (2020), the capacity of vegetation indices to discriminate spatial variability and explain soybean yield tends to increase during intermediate reproductive stages, when differences in growth and yield potential are more pronounced.

The maps obtained on 14 November 2021 (39 DAS; Figure 2C) and 24 November 2022 (42 DAS; Figure 2D) showed a relatively homogeneous spatial distribution of NDVI values, indicating reduced variability in vegetative vigor within the area. However, an increase in index values was observed at 42 DAS (Figure 2). According to Carneiro et al. (2020), NDVI values tend to increase during soybean vegetative growth as a function of leaf expansion and biomass accumulation.

Similarly, the maps obtained on 4 November 2023 (36 DAS; Figure 2E) and 13 November 2024 (48 DAS; Figure 2F) showed a homogeneous spatial distribution of NDVI values. However, a marked increase in index values was observed at 48 DAS, reflecting greater vegetative vigor of the crop, in agreement with Farias et al. (2023) and Carneiro et al. (2020).

Figure 3 shows the NDRE maps corresponding to the dates that provided the best results for estimating soybean yield, considering the performance indicators of the fitted models (R^2, RMSE and nRMSE). The NDRE maps showed a temporal pattern similar to that observed for NDVI, with increasing index values throughout crop development and spatial differences among assessment dates. However, the spatial variability observed for NDRE was more pronounced, especially between the 2019/2020 (Figure 3A) and 2020/2021 (Figure 3B) growing seasons and between the 2023/2024 (Figure 3E) and 2024/2025 (Figure 3F) growing seasons, being even more evident when comparing the latter two years.

This behavior is related to the greater sensitivity of NDRE to variations in chlorophyll concentration and plant physiological status. At more advanced developmental stages, when soybean has a higher leaf area index and accumulated biomass, NDRE maintains a high capacity to discriminate canopy variability, whereas NDVI may present limitations resulting from spectral saturation (CARNEIRO et al., 2020).

The spatial patterns observed in the NDRE maps followed the variability in grain yield among the evaluated growing seasons (Figure 4), with this correspondence being more evident in the most recent seasons. In these situations, areas with higher index values coincided with regions of higher yield, reinforcing the potential of NDRE for soybean yield estimation.

In addition, NDRE showed greater capacity to discriminate differences in crop development at more advanced reproductive stages because it uses the red-edge spectral region, which remains sensitive to chlorophyll variations even under high plant biomass conditions. Thus, this index has potential for monitoring spatial variability and estimating soybean yield (MILLER et al., 2018).

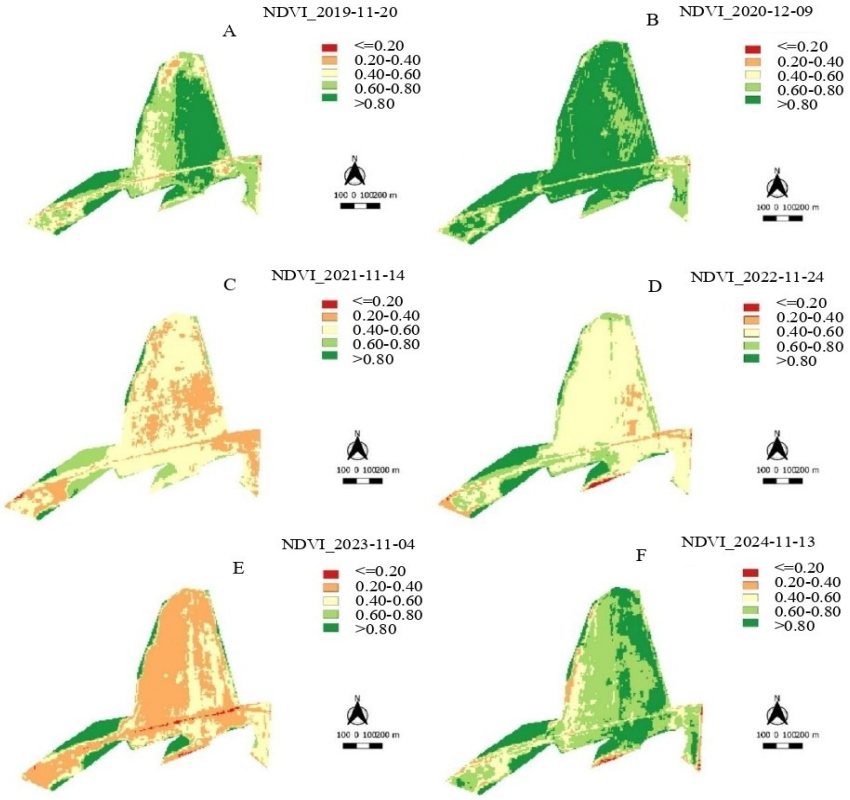


Figure 2. Normalized Difference Vegetation Index (NDVI) maps obtained on the image acquisition dates that provided the best performance for soybean yield estimation.

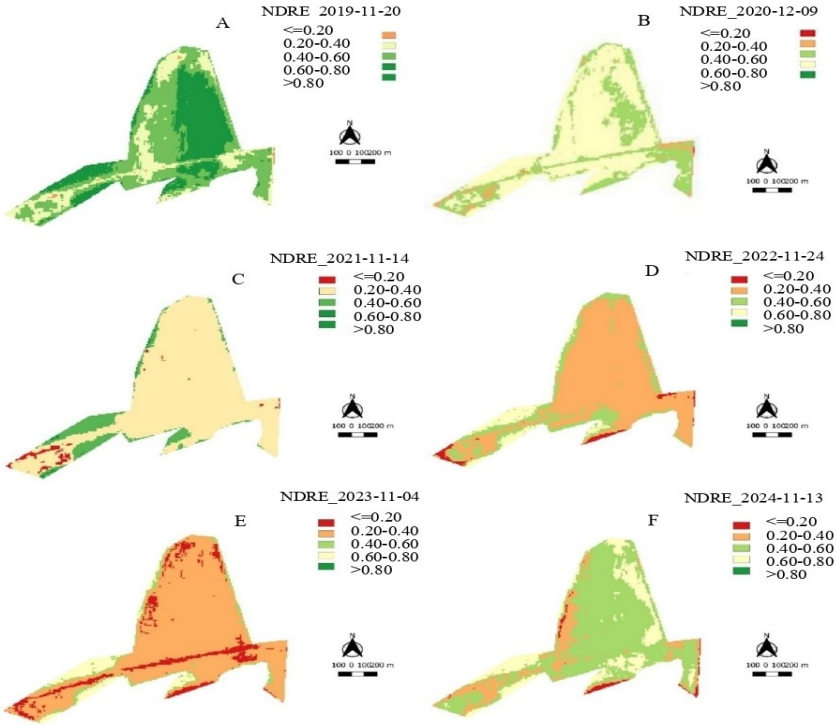


Figure 3. Normalized Difference Red Edge Index (NDRE) maps obtained on the image acquisition dates that provided the best performance for soybean yield estimation.

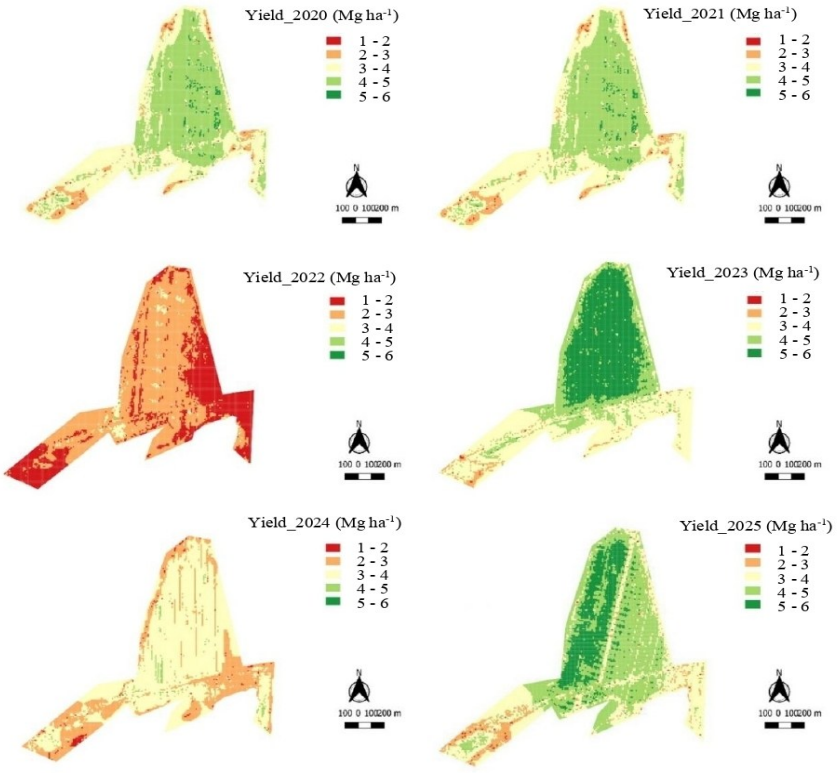


Figure 4. Soybean grain yield maps obtained from yield-monitor data in the evaluated growing seasons.

Management zones were delineated using grain yield maps and the NDVI and NDRE vegetation indices obtained on the dates that provided the best estimates of soybean yield. The definition of the number of classes was based on the joint analysis of the Fuzziness Performance Index (FPI) and Normalized Classification Entropy (NCE), with the lowest values observed for the two-class configuration (FPI = 0.068 and NCE = 0.085, respectively). Thus, the delineation of two management zones was selected (Figure 5), since this configuration provided the best clustering quality and spatial interpretation.

The resulting management zones were represented as polygons across the field (Figure 5), showing a clear spatial division, with Zone 1 (Z1) occupying the outer portions of the field, with a larger area in the western portion, and Zone 2 (Z2) showing a broad distribution throughout the area. In terms of extent, Z1 covered approximately 24.898 ha, whereas Z2 occupied approximately 97.752 ha. The recurrence of similar spatial patterns observed in the yield, NDVI and NDRE maps across the different growing seasons suggests the existence of a relatively stable spatial structure in the study area.

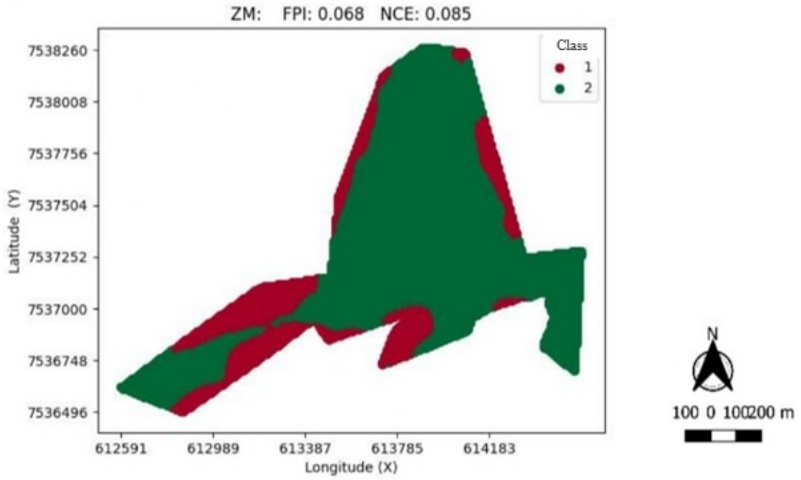


Figure 5. Management zones generated from yield maps and from NDVI and NDRE maps obtained on the dates that provided the best yield estimates.

The mean soybean grain yield values for the two delineated management zones indicate a consistent yield gradient between zones (Table 5), with a progressive increase in mean values from the lower-yield-potential zone (Zone 2) to the higher-yield-potential zone (Zone 1).

It was observed that the zones classified with higher mean NDRE values generally showed higher mean yields. Zone 1 had the highest mean NDRE value (0.46) and also the highest mean yield (4,348.27 kg ha^{-1}), followed by Zone 2, with a mean NDRE of 0.38 and a mean yield of 4,005.10 kg ha^{-1}.

Table 5. Mean values and coefficients of variation (CV, %) of yield and NDRE in the delineated management zones.

Parameters	Zone 1		Zone 2	
	Yield (kg ha^{-1})	NDRE		
Mean	4348.27 ± 662.33	0.46 ± 0.08	4005.1 ± 630.06	0.38 ± 0.08
C.V. (%)	15.23	18.66	15.64	21.22

CV: coefficient of variation; NDRE: Normalized Difference Red Edge Index.

The NDRE values obtained in this study are similar to those reported by Miller et al. (2018), who observed values ranging from approximately 0.30 to 0.60 in soybean fields under different developmental conditions, highlighting the high sensitivity of this index for identifying variations in vegetative vigor and canopy chlorophyll content.

Regarding yield, the mean values observed in the management zones can be considered high when compared with the national average yield of Brazilian soybean, demonstrating that even in high-yielding areas, it was possible to identify distinct production environments. The mean difference in yield between Zones 1 and 2 was approximately 340 kg ha^{-1}, and this difference indicates the existence of spatial variability in the yield potential of the area, suggesting the possibility of adopting differentiated management strategies according to the specific needs of each zone.

All the regression models fitted to estimate soybean yield as a function of NDRE in the two delineated management zones (Table 6) showed high and significant coefficients of determination, demonstrating the capacity of NDRE to explain yield variability within each productive environment.

Table 6. Regression models fitted to estimate soybean yield as a function of NDRE in the different delineated management zones.

Zone	Linear equation	R^2	RMSE (kg ha^{-1})	nRMSE (%)
1	413.86 + 14019.51*NDRE - 11743.91*(NDRE$^{\wedge 2}$)	0.67**	295.67	7.8
2	1294.80 + 6817.98*NDRE	0.76**	332.07	7.7

Zone 2 showed the best performance, with a coefficient of determination of 0.76, indicating that 76% of the yield variation was explained by NDRE. In addition, the model showed an RMSE of 332.07 kg ha^{-1} and an nRMSE of 7.7%, being classified as excellent according to the criteria proposed by Silva et al. (2018). In Zone 1, the model explained 67% of the yield variability, with an RMSE of 295.67 kg ha^{-1} and an nRMSE of 7.8%, indicating high accuracy of the estimates.

It was also observed that the model fitted for Zone 2 showed a linear behavior, whereas in Zone 1 the best fit was obtained using a quadratic model. This result suggests that, in environments with higher yield potential, the yield response to increasing NDRE values may not occur linearly.

Overall, the results indicate that the prior delineation of management zones contributed to obtaining more specific models for each production environment, reducing the internal heterogeneity of the data and increasing the predictive capacity of NDRE. Similar results were reported by Bazzi et al. (2013), who highlighted that stratifying the area into management zones allows the spatial variability of the production system to be more adequately represented, favoring the development of more precise agronomic models. Likewise, Maimaitijiang et al. (2020) observed that indices derived from the red-edge region have high potential for soybean yield estimation, especially when associated with the characterization of within-field spatial variability.

CONCLUSIONS

The NDVI and NDRE vegetation indices were significantly correlated with soybean yield across the evaluated growing seasons, with NDRE showing superior predictive performance under the conditions of this study.

The use of multiple regression models improved soybean yield prediction under the conditions evaluated, resulting in better model fit and greater estimation accuracy compared with simple regression models.

Management-zone delineation was consistent with the spatial variability of grain yield and enabled the development of NDRE-based predictive models specific to each production environment under the conditions evaluated.

REFERENCES

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, 22: 711-728, 2013.

ANDRADE JÚNIOR, A. S. et al. Predicting soybean grain yield using aerial drone images. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 26: 466-476, 2022.

ANDRADE, T. G. et al. Soybean yield prediction using remote sensing in Southwestern Piauí State, Brazil. **Revista Caatinga**, 35: 105-116, 2022.

BATISTELLA, D. et al. Comparative analysis of orbital sensors in soybean yield estimation by the random forest algorithm. **Ciência e Agrotecnologia**, 47: e002423, 2023.

BAZZI, C. L. et al. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, 33: 952-964, 2013.

CARNEIRO, F. M. et al. Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. **Precision Agriculture**, 21: 979-1007, 2020.

CEMTEC - Centro de Monitoramento do Tempo e do Clima de Mato Grosso do Sul. **Banco de dados**. 2026. Available at: https://www.cemtec.ms.gov.br/bancodedados/. Access on: Abr. 17, 2026.

CHRISTENSON, B. S. et al. Predicting soybean relative maturity and seed yield using canopy reflectance**. Crop Science**, 56: 625-643, 2016.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de grãos, safra 2023/2024, décimo segundo levantamento, setembro de 2024**. Available at: https://www.gov.br/conab/pt-br/atuacao/informacoes-agropecuarias/safras/safra-de-graos/boletim-da-safra-de-graos/12o-levantamento-safra-2023-2024. Access on: Set. 23, 2024.

DAMIAN, J. M. et al. Applying the NDVI from satellite images in delimiting management zones for annual crops. **Scientia Agricola**, 77: e20180055, 2020.

FARIAS, G. D. et al. Normalized Difference Vegetation Index (NDVI) for soybean biomass and nutrient uptake estimation in response to production systems and fertilization strategies. **Frontiers in Sustainable Food Systems**, 6: 959681, 2023.

GUSSO, A.; ARVOR, D.; DUCATI, J. R. Modelo para previsão da produção de soja baseado em condições físicas predominantes. **Pesquisa Agropecuária Brasileira**, 52: 95-103, 2017.

HUNT, M. L. et al. High resolution wheat yield mapping using Sentinel-2. **Remote Sensing of Environment**, 233: 111410, 2019.

JUNG, E. A. et al.. Operational and energy performance of a powershift transmission tractor during tillage at different speeds. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 29: e287659, 2025.

MAIMAITIJIANG, M. et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote Sensing of Environment**, 237: 111599, 2020.

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**, 8: 126-134, 2004.

MARSHALL, M. et al. Field-level crop yield estimation with PRISMA and Sentinel-2. **ISPRS Journal of Photogrammetry and Remote Sensing**, 187: 191-210, 2022.

MERCANTE, E. et al. Modelos de regressão lineares para estimativa de produtividade da soja no oeste do Paraná, utilizando dados espectrais. **Engenharia Agrícola**, 30: 504-517, 2010.

MILLER, J. J. et al. Characterizing soybean vigor and productivity using multiple crop canopy sensor readings. **Field Crops Research**, 216: 22-31, 2018.

NUNES, H. D. et al. Variabilidade espacial do desempenho operacional de colhedora de soja. **Scientia Agraria**, 21: e99911, 2025.

PIMENTEL-GOMES, F.; GARCIA, C. H. **Estatística aplicada a experimentos agronômicos e florestais**. Piracicaba, SP: FEALQ, 2002. 309 p.

PEREIRA, E. C. et al. Soybean yield estimation in the Brazilian Midwest using Sentinel-2 imagery. **Big Earth Data**, 10: 1-25, 2026.

PEREIRA, G. W. et al. Smart-Map: an open-source QGIS plugin for digital mapping using machine learning techniques and ordinary kriging. **Agronomy**, 12: 1350, 2022.

QGIS.org. QGIS **Geographic Information System**. QGIS Association. Available at: https://qgis.org/pt_BR/site/index.html. Access on: Feb. 4, 2024.

SANTOS, H. G. et al. **Sistema Brasileiro de Classificação de Solos**. 6. ed. rev. e ampl. Brasília, DF: Embrapa, 2025. 393 p.

SCALON, L. Q. et al. Soil physical attributes for discriminating soybean yield-based management zones. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 30: e302513, 2026.

SCHWAMBACH, D. A. et al. Variability of soybean and corn yield and soil texture in the generation of management zones. **Energia na Agricultura**, 36: 335-347, 2021.

SILVA, V. P. R. et al. Calibração e validação do modelo AquaCrop para a cultura da soja sob diferentes níveis de irrigação na região do Matopiba, Brasil. **Ciência Rural**, 48: 1-8, 2018.

SLAIFFER, H. et al. Índices de vegetação e sua correlação com a produtividade da cultura da soja em área irrigada e de sequeiro. **Nativa**, 13: 495-502, 2025.

STEPANOV, A. et al. Predicting soybean yield at the regional scale using remote sensing and climatic data. **Remote Sensing**, 12: 1-23, 2020.