

## SOYBEAN YIELD PREDICTION USING REMOTE SENSING IN SOUTHWESTERN PIAUÍ STATE, BRAZIL<sup>1</sup>

THATIANE GOMES ANDRADE<sup>2</sup>, ADERSON SOARES DE ANDRADE JUNIOR<sup>3\*</sup>, MELISSA ODA SOUZA<sup>4</sup>, JOSE WELLINGTON BATISTA LOPES<sup>5</sup>, PAULO FERNANDO DE MELO JORGE VIEIRA<sup>3</sup>

**ABSTRACT** – Recent researches have shown promising results for the use of orbital data using the Normalized Difference Vegetation Index (NDVI) to monitor and predict soybean grain yield. The objective of this work was to evaluate propositions of multiple linear regression models to predict soybean grain yield using NDVI. The research was carried out at the Celeiro Farm, in Monte Alegre do Piauí, PI, Brazil, in an area of 200 ha. Five images were collected during the soybean crop cycle: one from the Landsat 8 and four from the Sentinel 2. Regression analyses were carried out between grain yield data (predicted variable) extracted from harvest maps and spectral data (predictor variables) from NDVI of soybean crops at different developmental stages. The promising models were selected by the Akaike Information Criterion (AIC). The models were validated using Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (nRMSE), considering the mean of soybean yield of the plot. The linear regression models developed with NDVI for the V5-V6 and R2 developmental stages showed promising results for the prediction of soybean grain yield, with mean error of predictions of 153.9 kg ha<sup>-1</sup>, representing 4.2% when compared to the data from field measures.

**Keywords:** NDVI. Crop season forecast. Multiple regression.

## ESTIMATIVA DA PRODUTIVIDADE DA SOJA POR SENSORIAMENTO REMOTO NA REGIÃO SUDOESTE DO PIAUÍ

**RESUMO** – Pesquisas recentes apontam resultados promissores da integração de dados orbitais utilizando o índice de vegetação NDVI, para monitorar e estimar a produtividade da soja. O objetivo do trabalho foi avaliar a proposição de modelos de regressão linear múltipla para estimativa da produtividade de grãos com uso do índice NDVI. A pesquisa foi realizada na Fazenda Celeiro, município de Monte Alegre do Piauí, PI, sendo a área de estudo de 200 ha. Foram utilizadas cinco imagens durante o ciclo de cultivo da soja, sendo uma do Landsat 8 e quatro do Sentinel 2. Efetuou-se análises de regressão entre dados de produtividade de grãos (variável predita), extraídos dos mapas de colheita, e dados espectrais de (variáveis preditoras) oriundos dos índices de vegetação (NDVI) de diferentes estádios de desenvolvimento da soja. A seleção dos modelos promissores foi efetuada pelo critério de informação de Akaike (AIC). Para validação dos modelos utilizou-se a raiz quadrada do erro quadrado médio (RMSE) e a raiz quadrada do erro quadrado médio normalizado (nRMSE) pela média da produtividade da soja no talhão. O modelo de regressão linear gerado com o índice de vegetação NDVI nos estádios de desenvolvimento V5-V6 e R2, mostrou-se promissor na predição da produtividade de grãos de soja, com erro médio de estimativa da ordem de 153,9 kg ha<sup>-1</sup>, o que representa 4,2% em relação aos dados medidos em campo.

**Palavras-chave:** NDVI. Previsão de safra. Regressão múltipla.

\*Corresponding author

<sup>1</sup>Received for publication in 02/22/2021; accepted in 08/16/2021.

<sup>2</sup>Postgraduate Program in Agrarian Sciences, Universidade Federal do Piauí, Bom Jesus, PI, Brazil; [thatiane.eng.agronoma@gmail.com](mailto:thatiane.eng.agronoma@gmail.com) – ORCID: 0000-0003-0500-0862.

<sup>3</sup>Embrapa Meio-Norte, Teresina, PI, Brazil; [aderson.andrade@embrapa.br](mailto:aderson.andrade@embrapa.br) – ORCID: 0000-0002-0619-1851, [paulofermando.vieira@embrapa.br](mailto:paulofermando.vieira@embrapa.br) – ORCID: 0000-0002-5133-3871.

<sup>4</sup>Center of Agrarian Sciences, Universidade Estadual do Piauí, Teresina, PI, Brazil; [melissasouza@cca.uespi.br](mailto:melissasouza@cca.uespi.br) – ORCID: 0000-0003-4137-4298.

<sup>5</sup>Universidade Federal do Piauí, Bom Jesus, PI, Brazil; [wellingtonjwl@gmail.com](mailto:wellingtonjwl@gmail.com) – ORCID: 0000-0003-1340-9983.

## INTRODUCTION

Monitoring dynamic systems over large areas in annual crops, such as soybean, is challenging. The aid of technologies, such as remote sensing, using through aerial images has been beneficial in the last years; the generation and use of technologies has significantly increasing soybean crop yield in Brazil (BALBINOT JUNIOR et al., 2018).

Monitoring the growth and development of soybean crops in the field requires information that assists in the decision making during the planting, management, and cultural practices and in the prediction of yields focused on the marketing of the production. The use of remote sensing has significantly contributed to the agricultural sector, providing more accurate data; geotechnologies are promising means, in cost-benefit terms, for fast collection of reliable information in large areas (ATZBERGER et al. 2013; FORMAGGIO; SANCHES, 2017).

Remote sensing is a technique that investigates the interaction between electromagnetic radiation and different types of targets. Scientific advances have enabled the use of satellite images for vegetation studies, providing information on distribution of different vegetation types, phenological state, stress conditions, and nutrient deficiencies, thus assisting in the decision making during the crop management (MOREIRA, 2003; GUEDES; SILVA, 2018).

The products from orbital data through remote sensing techniques include vegetation indexes (ROUSE et al., 1973). The use of vegetation indexes assists in studies about spectral and temporal characteristics of the soybean crop during its phenological cycle, and factors that affect its growth, development, and yield, such as climate, relief, and soil physical and chemical attributes (TRINDADE et al., 2019). The Normalized Difference Vegetation Index (NDVI) is among the most used indexes for monitoring vegetation, which has been widely used in scientific researches (MERCANTE et al., 2010; GUSSO et al., 2013; TRINDADE et al., 2019; SARMIENTO et al., 2020).

Vegetation indexes enables field identifications of nutrient deficiencies (TRINDADE et al., 2019) and water stress (SCHIRMBECK et al., 2019); diagnoses of several biophysical parameters, including leaf area index (BERGER et al., 2019), soil cover percentage (MACEDO et al., 2020), biomass (RODIGHERI et al., 2020), and

photosynthetic activity (RODRIGUES et al., 2013); and predictions of grain yield (MELO et al., 2008; BERTOLIN et al., 2017; TRINDADE et al., 2019).

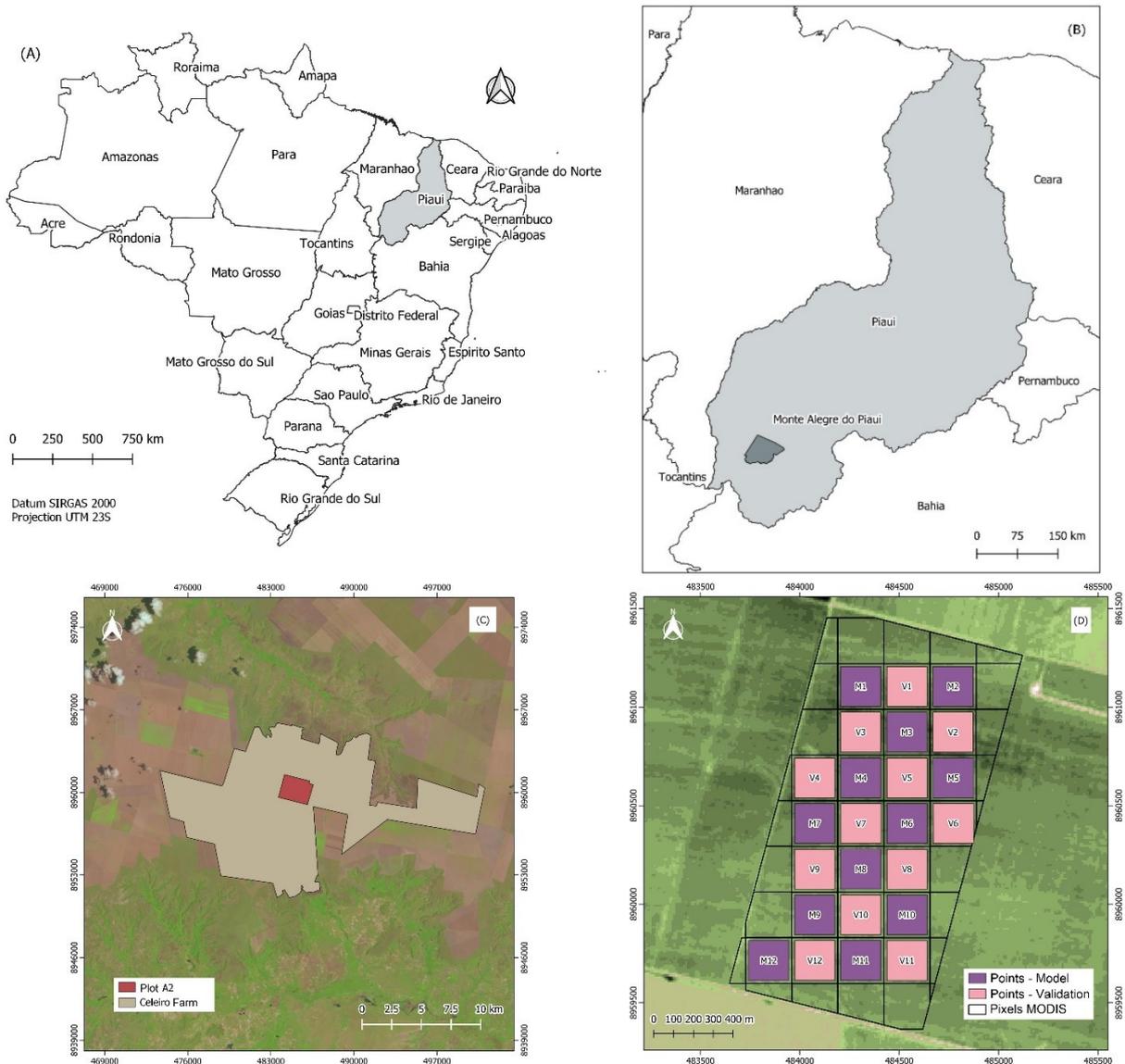
Studies on prediction models for crop yields using vegetation indexes have been conducted since the decade of 1990, with proposition of models based on only one image, which showed the highest correlation between biomass and yield. Recently, researches have focused on the monitoring during the whole crop cycle (MELO et al., 2008; MERCANTE et al., 2010; GUSSO et al., 2013; SARMIENTO et al., 2020), searching for more accurate information to propose models that predict the soybean yield. Thus, the objective of this work was to evaluate the fit of multiple linear regression models for predicting soybean grain yield based on NDVI in a reference Farm in the Cerrado biome in the southwestern Piauí State, Brazil.

## MATERIAL AND METHODS

### Characterization of the study area

The study was conducted at the Celeiro Farm, in Monte Alegre do Piauí, State of Piauí, Brazil (09° 24'01"S, 45°07'21"W) considering a plot A2 of 402 ha in an experimental site with effective area of 200 ha (Figure 1). The region presents an Aw, tropical savanna climate, according to the Köppen climate classification, with a rainy season from October to April and mean annual rainfall depth of 1,200 mm (ANDRADE JUNIOR et al., 2004). Climatic data (air temperature, relative air humidity, global solar radiation, wind speed, and rainfall) were recorded daily by an automatic agrometeorological station distant 500 m from the experimental plot. The predominant soil of the region was classified as a Typic Hapludox (Latossolo Amarelo distrófico típico; SANTOS, 2018) of loamy-clay-sandy texture.

Soybean crops were grown in the study area was in the 2018/2019 agricultural year from November to March. Soybean seeds were sowed on November 27, 2018, with spacing of 0.45 m between rows for a density of 14 plants per linear meter (311,000 plants per hectare). The soil was prepared under minimum cultivation system, with sowing on the straw of the previous crop (maize). The BMX 8579 IPRO (Bonus) soybean cultivar was used, which belongs to the maturation group 7.9, presents indeterminate growth habit, and is resistant to lodging (CELEIRO SEMENTES, 2020).



**Figure 1.** Location of the study area. A: Position of the State of Piauí in the map of Brazil; B: Position of Monte Alegre do Piauí in the map of Piauí; C: Celeiro Farm and the plot A2; D: Plot A2 highlighting the sampling points used for generation (M) and validation (V) of regression models.

**Soybean grain yield**

All areas of the plot were harvested when the soybean crop reached the final maturation stage (R9), using harvesters equipped with GNSS to obtain the harvest maps of the plot. The harvest maps of the plot were developed in vector shapes by a program that manages the harvesters.

The harvest maps were rasterized to enable the extraction of yield values of each sampling point, using the rasterization tool of the QGIS program. Then, the mean grain yield values were extracted for each sampling point. A vector layer consisted of 12 squares with dimensions of 200 × 200 m was used, which were defined based on the image of the MODIS sensor projected on the experimental plot (Figure 1D). The mean grain yield values for

validation of prediction models were extracted from validation points (Figure 1D).

**Normalized Difference Vegetation Index (NDVI)**

Products from the sensors of the satellites Sentinel 2 and Landsat 8 were used to generate the NDVI. The images presented high to medium temporal resolution (5 days for the Sentinel 2, 15 days for the Landsat 8), which provided data in short to medium time intervals for the research. Twenty images covering the whole soybean crop cycle were obtained from the site of the United States Geological Survey (USGS) (<https://earthexplorer.usgs.gov/>) and of European Space Agency (ESA) (<https://apps.sentinel-hub.com/eo-browser/>). However, only five images were used,

which presented absence of clouds on the experimental plot, one of the Landsat 8 and four of the Sentinel 2 (Table 1), since the study was conducted in a rainy period, with high cloudiness. The QGIS program was used to redesign the images for the projection UTM-WGS84-Fuso23S.

Images of both satellites with level 2 of processing, radiometric and geometric corrections, and surface reflectance values were used. The image of the Landsat 8 was subjected to atmospheric correction using the LEDAPS and LaSRC

algorithms, which correct effects of temporal, spatial and spectral dispersion and absorption of atmospheric gases, aerosols, and water vapor, and is needed for a reliable characterization of the land surface (<https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-2-science-products>). The images of the Sentinel 2 were corrected using the product level L1, with corrections processed through the Sen2cor Sentinel 2 Toolbox algorithm (<https://earth.esa.int/web/sentinel/toolboxes/sentinel-2>).

**Table 1.** Data of the satellite images used in the study.

Dates	Satellites	Spatial resolution	Bands	Phenological stage	DAS
Dec 16	Sentinel 2	10 m	RED4 / NIR8	V2	19
Dec 26	Sentinel 2	10 m	RED4 / NIR8	V5 and V6	29
Jan 12	Landsat 8	30 m	RED4 / NIR5	R2	46
Jan 30	Sentinel 2	10 m	RED4 / NIR8	R4	64
Feb 24	Sentinel 2	10 m	RED4 / NIR8	R5	89

<sup>a</sup>NIR: near infrared; <sup>b</sup>RED: Red; DAS: Days after sowing.

The NDVI was acquired through digital values from the products provided by the satellites. The difference between the detected reflectance in the visible and near infrared bands, divided by the sum of these values was calculated in the program QGIS 2.8.2, as defined by Rouse et al. (1973), according to Equation 1:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

where  $\rho_{NIR}$  is the near infrared reflectance (Landsat 8 band 5 and Sentinel 2 band 8); and  $\rho_{RED}$  is the red reflectance (Landsat 8 band 4 and Sentinel 2 band 4).

### Models for predicting grain yield and model validation

Regressions between grain yield data (predicted variable) extracted from harvest maps and spectral data (predictor variables) from NDVI were analyzed. The mean NDVI and yield grain values were extracted from each sampling point (200 × 200 m), using the zonal statistics plugin of the QGIS 2.8.2 program. The modeling was carried out for all images obtained, and for images from the beginning, medium, and end of the soybean crop cycle. Twelve models were tested to evaluate the correlation between the predicted and predictor variables, and stepwise regression technique was applied to assess the importance of predictor variables and which of them should be maintained in the model. The stepwise regression technique consists in starting the

adjustment of a reduced model, and iteratively build a sequence of regression models including or excluding variables, one by one. The inclusion or exclusion criterion of variables was based on the Akaike Information Criterion (AIC) (AKAIKE, 1974).

The performance of the models was evaluated based on the coefficient of determination ( $R^2$ ), prediction standard error, and on the AIC. The AIC establishes a correlation between the information or distance of Kullback-Leibler (KULLBACK; LEIBLER, 1951), which is a measure of relative discrepancy between two models, the true and approximate model, and the maximum log-likelihood function for selecting the models. The AIC is defined by Equation 2:

$$AIC = -2LL + 2p \quad (2)$$

where  $LL$  is the Neperian logarithm of the maximum likelihood function, calculated for the prediction values of parameters; and  $p$  is the number of parameters of the model. The models with lower AIC values present better fit. The Shapiro-Wilk test was used to assess the normality and the Breusch-Pagan test were used to assess the heteroscedasticity of the data before the regression analysis. The analyses were carried out using the R statistical program (R CORE TEAM, 2019).

The linear regression models obtained were validated using the NDVI and grain yield values of validation sampling points from the harvest map, different of points were used for the modeling (Figure 1D). The statistical indexes: root mean square error (RMSE) and normalized root mean

square error square (nRMSE) were used as indicators of degree of fitting of the models, considering the mean soybean yield values (Equations 3 and 4).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - Y_i')^2}{n}} \quad (3)$$

$$RMSE = \frac{RMSE}{\bar{Y}} \quad (4)$$

where  $n$  is the number of observations;  $Y_i$  is the grain yield values found;  $Y_i'$  is the grain yield values predicted by the models; and  $\bar{Y}$  is the mean grain yield value found.

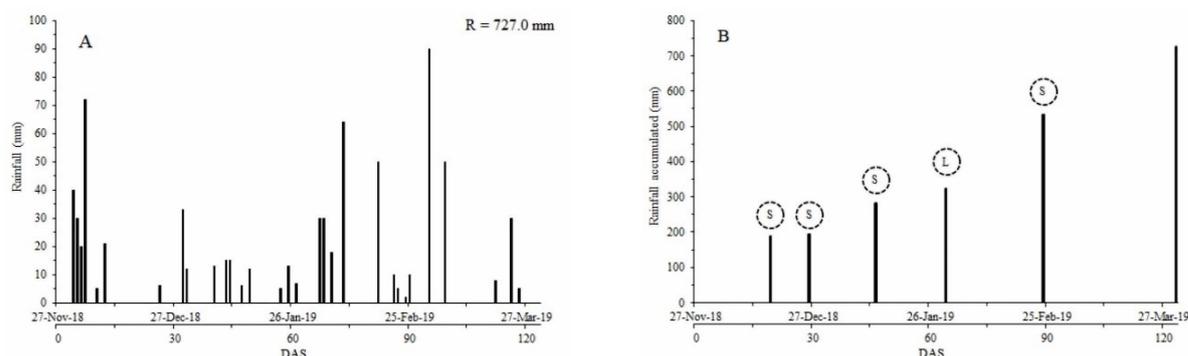
## RESULTS AND DISCUSSION

### Rainfall depths

The rainfall depths recorded during the whole soybean crop cycle was 727.0 mm, lower than the historical mean of the region (1,200 mm)

(ANDRADE JUNIOR et al., 2004). However, the rainfall distribution over the crop cycle was relatively adequate, without occurrence of long dry periods.

There were 31 rainy days during the crop cycle, which last 123 days (November 27, 2018 to March 30, 2019), resulting, on average, in one rainfall event every four days (Figure 2A). There was a 13-day period without rainfall, between December 10 and 22, 2018, 13 to 25 days after sowing (DAS), in the initial development stage of plants, which caused production losses. There were no periods without rainfall longer than 7 days in the essential stages for grain filling. However, there was another dry spell at the end of physiological maturation, March 5 to March 20, which did not compromise the soybean development and yield (SILVA, 2011). The accumulated rainfall up to the day of acquisition of images, due to the passing of satellites, was 188.0 mm, at the passage of the Sentinel 2 satellite (December 16, 2018; 19 DAS), and 534.0 mm, at the passage of the Landsat 8 satellite (February 24, 2019; 89 DAS) (Figure 2B).



**Figure 2.** Daily rainfall depths over the soybean cycle (A) and accumulated rainfall depths (B) up to the passing of the Landsat (L) and Sentinel (S) satellites (B). Celeiro Farm, Plot A2, 2018-2019 crop season.

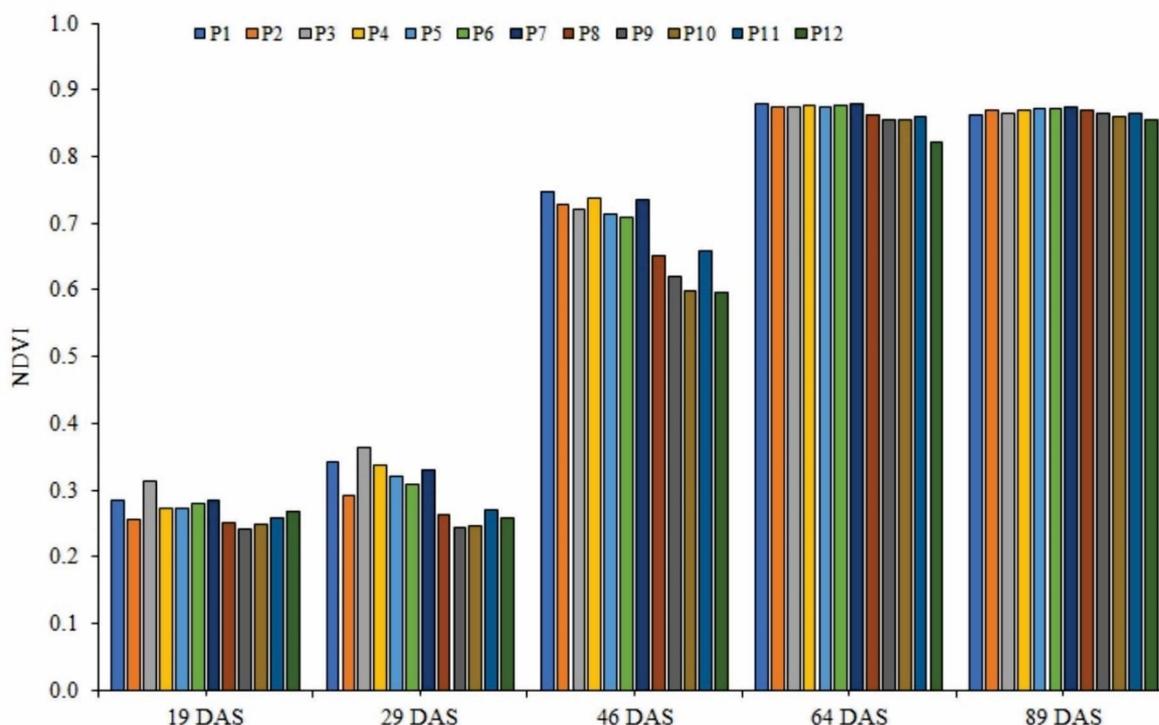
Water availability is important for soybean crops, mainly from the flowering to the grain filling stage (SILVA, 2011). Water deficiency during the flowering (R1-R3) results in lower number of pods, since soybean is sensitive to water deficiency (CAMARGO; BRUNINI; MIRANDA, 1986). The flowering and grain filling stages present the highest water consumption; thus, they are the most critical stages for water deficits. The rainfall distribution was adequate during the soybean crop cycle, including the more critical stages (Figure 2), ensuring a good crop development and grain yield.

### Normalized Difference Vegetation Index (NDVI)

NDVI values vary over the soybean cycle (Figure 3), denoting that NDVI increases as the plants develop, mainly, in terms of leaf area and shoot dry biomass weight (ALLEN; PEREIRA,

2009). The NDVI values started to change at the initial stages of the soybean crop cycle. In the V2 stage (19 DAS), the values of NDVI ranged from 0.2 to 0.36 between sampling points. This dynamic was found up to the V5 stage (29 DAS).

In the R2 stage (46 DAS), NDVI values started to show differences between sampling points, P12 presented the lowest (0.59) and P1 presented the highest (0.74) NDVI value. NDVI values showed small differences between sampling points again at the final stages of the crop cycle (64 DAS and 89 DAS). These results are consistent with those of Crusiol et al. (2013), who evaluated the temporal spectral profile of soybean crops (cultivar BRS 284) at the main stages of the growth cycle and found the highest increases in NDVI values between the V2 and R2 stages, and smaller increases after the R2 stage, denoting a stabilization of NDVI from the R2 to the R6 stage.



**Figure 3.** Normalized Difference Vegetation Index (NDVI) in the sampling points during the soybean crop cycle: at 19 days after sowing (DAS) (December 16, 2018); 29 DAS (December 26, 2018); 46 DAS (January 12, 2019); 64 DAS (January 30, 2019), and 89 DAS (February 24, 2019). P1 to P12 = sampling points.

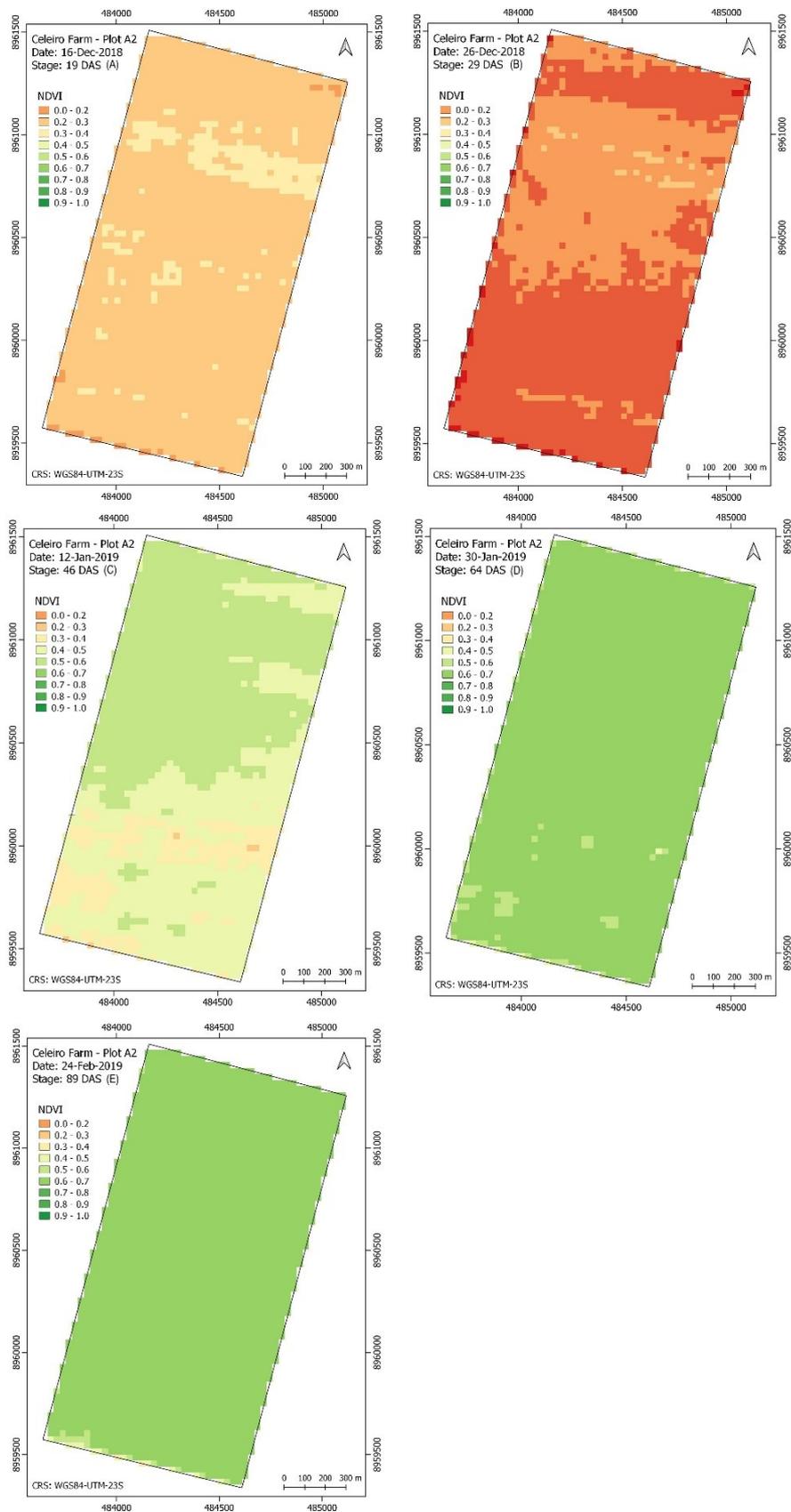
In the stage V2, at 19 DAS (December 16, 2018), there were predominance of NDVI pixels values classified as  $< 0.3$  (Figure 4A), with a percentage of 88.9% (Figure 5A), since the crop was at the beginning of vegetative development. A slight increase in NDVI pixel values was found at 29 DAS (stage V5) to the class of 0.2 to 0.4, due to the crop advanced vegetative development (Figure 4B), with percentages varying from 40.35% to 58.37% (Figure 5B).

A higher variation in NDVI values was found at 46 DAS (stage R2), with predominance of classes 0.4 to 0.6, in the areas P4 to P12, and 0.6 to 0.8 in the areas P1, P2, and P3 (Figure 4C), with percentages of 38.82% to 49.38% (Figure 5C). However, after the establishment of the crop, there was a higher predominance of NDVI values in the classes 0.8 to 1.0, at the final stages of the soybean crop cycle, with percentage of 98.15%.

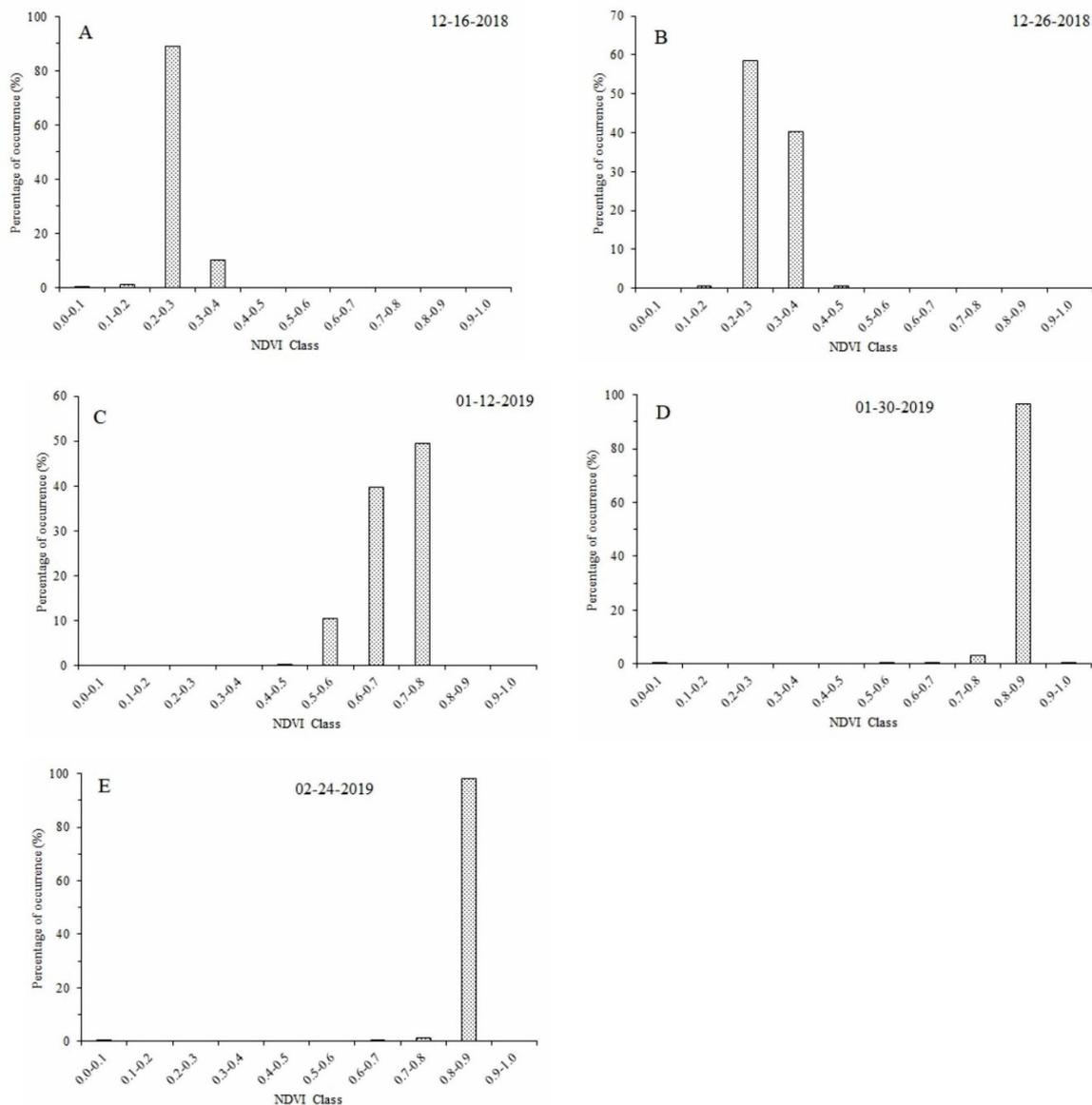
According to Allen and Pereira (2009), variations in mean NDVI values are related to increases in biomass, leaf area index, and soil cover fraction. This was confirmed in the present study; higher increases in NDVI were found in these

developmental stages. Bariani et al. (2015) found that NDVI values enables the identification of phenological stages of soybean crops and its temporal monitoring; they monitored irrigated soybean crops using a temporal series of NDVI and found that NDVI values vary from 0.2 to 0.85 in vegetative stages (V2 to V5) and from 0.54 to 0.84 in reproduction stages (R1 to R5). In addition, they found decrease in NDVI values after the beginning of physiological maturation (R6), which was not found in the present study due to the indeterminate growth habit of the Bonus cultivar.

Sarmiento et al. (2020) evaluated the capacity of a spectral agrometeorological model to predict soybean grain yield in the State of Mato Grosso, Brazil, and found variations in NDVI over the phenological stages, with higher variations in vegetative stages (V1 and R5), maximum NDVI in the stage R5 (0.89), and decreases at the beginning of the maturation stage (R8). They also found that NDVI analysis can identify and monitor phenological stages and whether the crop development is within the expected biomass production.



**Figure 4.** Variation in Normalized Difference Vegetation Index (NDVI) predicted for the plot A2 at different developmental stages of soybean crops. Celeiro Farm, 2018/2019 crop season. A: December 16, 2018; B: December 26, 2018; C: January 12, 2019; D: January 30, 2019; and E: February 24, 2019.



**Figure 5.** Histogram of percentage distribution of Normalized Difference Vegetation Index (NDVI) values in each evaluation. A: December 16, 2018; B: December 26, 2018; C: January 12, 2019; D: January 30, 2019; and E: February 24, 2019.

### Models to predict soybean grain yield

The multiple linear regression equations developed to predict soybean grain yield as a function of NDVI values, obtained from images of the Sentinel 2 and Landsat 8 satellites at different developmental stages, presented high coefficients of determination ( $R^2 = 0.91$  to  $0.88$ ), denoting that 91% to 88% of the variability in soybean grain yield can be explained by multiple linear regression equations using NDVI (Table 2). Gusso et al. (2013) evaluated the fit of a model for regional scale and conditions climatic of the State of Rio Grande do Sul, Brazil, to predict soybean grain yield using the EVI vegetation index, extracted from MODIS images and grain yield data obtained from the Brazilian Institute of

Geography and Statistics (IBGE); they found that the model explained only 64% of the grain yield variability, medium fit, since they did not use grain yield data from field measures.

Mercante et al. (2010) proposed linear regression models for the western Paraná State, Brazil, to predict soybean grain yield using the NDVI and GVI vegetation indexes from images of the Landsat 5/TM satellite; they found the mean index values (NDVI and GVI) of all images were more correlated to soybean grain yield than to each evaluation date, separately. The use of multiple regressions with these two vegetation indexes in all evaluation dates provided a better correlation with grain yield.

**Table 2.** Multiple linear regression equations for predicting soybean grain yield based on the Normalized Difference Vegetation Index (NDVI), at different stages of the crop cycle.

Reg.	Intercept	NDVI-1	NDVI-2	NDVI-3	NDVI-4	NDVI-5	R <sup>2</sup>	PSE	AIC
1	10013.01 <sup>ns</sup>	2511.45 <sup>ns</sup>	-9470.67 <sup>ns</sup>	11817.33*	6959.04 <sup>ns</sup>	-21154.59 <sup>ns</sup>	0.91	179.07	128.19
2	-4628.38 <sup>ns</sup>	2311.45 <sup>ns</sup>	-8016.53 <sup>ns</sup>	10689.25*	3147.09 <sup>ns</sup>	-	0.88	184.73	128.79
3	-2472.14 <sup>ns</sup>	2109.49 <sup>ns</sup>	-8230.50 <sup>ns</sup>	11690.95 <sup>ns</sup>	-	-	0.88	174.63	127.04
4	4599.10**	-21409.52*	16255.48**	-	-	-	0.66	278.36	137.64
5	2002.19 <sup>ns</sup>	6206.77 <sup>ns</sup>	-	-	-	-	0.09	434.42	147.59
6	1644.52 <sup>ns</sup>	-	6812.42*	-	-	-	0.40	351.17	142.48
7	-1079.53 <sup>ns</sup>	-	-	6920.73**	-	-	0.79	206.80	129.77
8	-16158.27**	-	-	-	22895.01**	-	0.74	231.70	132.50
9	-42664.66*	-	-	-	-	53453.67*	0.37	360.63	143.12
10	-6862.05 <sup>ns</sup>	-	-	4716.39 <sup>ns</sup>	8423.39 <sup>ns</sup>	-	0.81	207.33	130.57
11	-2641.14 <sup>ns</sup>	-	-	4703.08 <sup>ns</sup>	9900.46 <sup>ns</sup>	-6334.36 <sup>ns</sup>	0.82	218.55	132.42
12	10624.52 <sup>ns</sup>	-	-7488.29*	11128.70*	6788.65 <sup>ns</sup>	-21044.91 <sup>ns</sup>	0.91	167.08	126.38
Step.	-1979.33*	-	-6556.00*	11074.55**	-	-	0.88	165.36	125.14

Reg.: number of the regression equation; Step.: stepwise; NDVI-1: NDVI image of November 27, 2018 (sowing); NDVI-2: NDVI image of December 16, 2018 (V2); NDVI-3: NDVI image of December 26, 2018 (V5 and V6); NDVI-4: NDVI image of January 12, 2019 (R2); NDVI-5: NDVI image January 30, 2019 (R3); R<sup>2</sup>: coefficient of determination; PSE: prediction standard error (kg ha<sup>-1</sup>); AIC: Akaike Information Criterion; Significance level of coefficients of equations by the t test: \*\*\*: < 1% significance level; \*\*: 1% significance level; \*: 5% significance level; and ns: not significant.

The regression equations with NDVI values of all images (regression 1), and images of intermediate and final stages of the soybean crop (regression 12) presented the highest R<sup>2</sup> values (0.91), lowest prediction standard errors, 179.1 to 167.1 kg ha<sup>-1</sup>, and lowest AIC (128.19 and 126.38), respectively (Table 2). These results were consistent with those obtained by Mercante et al. (2010), who found lower R<sup>2</sup> values probably due to the use grain yield statistics of the IBGE to develop the equations data, whereas more accurate grain yield data, from harvest maps of the plot, was used in the present study.

The regression equation obtained by the stepwise procedure denoted better grain yield prediction for the use of joint NDVI images of the stages V5-V6 and R2 (Table 2). Mercante et al. (2010) found better fit of regression equation to predict soybean grain yield for NDVI images of initial and full development vegetative stages, which is similar to the present study. Trindade et al. (2019) found best soybean grain yield prediction models using NDVI values of the stage R2. The use of NDVI images of reproduction stages of soybean crops is more capable of predicting the crop production potential, since there is a high correlation between soybean grain yield and the proper nutrient and water supply up to this stage (SARMIENTO et al., 2020).

#### Validation of models for predicting soybean grain yield

The validation of multiple linear regression models is presented in Figure 6. All models

presented good fit for grain yield (GY) predictions, with RMSE values varying from 153.9 to 160.9 kg ha<sup>-1</sup> (Figure 6). The lowest RMSE values were found for prediction models obtained with the regression equation 12 and stepwise, which used NDVI images of intermediate (V5-V6) and flowering (R2) stages, confirming the results obtained by Mercante et al. (2010). Sarmiento et al. (2020) found similar values, with RMSE of 131.54 kg ha<sup>-1</sup> and R<sup>2</sup> = 0.72.

The nRMSE of the validated models showed prediction errors from 4.24% to 4.45%, when compared to grain yield values measured in the field, denoting that the models are promising for the prediction of soybean grain yield. Sarmiento et al. (2020) found a similar nRMSE value, 3.97% for a spectral agrometeorological model for soybean grain yield prediction in Mato Grosso State, Brazil. In the State of Rio Grande do Sul, Gusso et al. (2013) found errors in soybean grain yield prediction from -9.36% to 14.45%, higher than those found in the present study; however, they used for validation of a model for prediction of grain yield in regional scale using data of the IBGE.

Silva et al. (2018) evaluated models for identification of the resistance of soybean cultivars to water stress and established the following nRMSE ranges to assess the quality of validation of the models: excellent, when nRMSE is lower than 10%; good, between 10% and 20%; reasonable, between 20% and 30%, and bad when higher than 30%. Considering this criterion, the models were promising, mainly the model generated by the stepwise procedure, which fits into the excellent category and can be applied for the prediction of

soybean grain yield in the study region. However, these promising models should be evaluated in other

farms of the region to confirm the results presented and regionalization of the models.

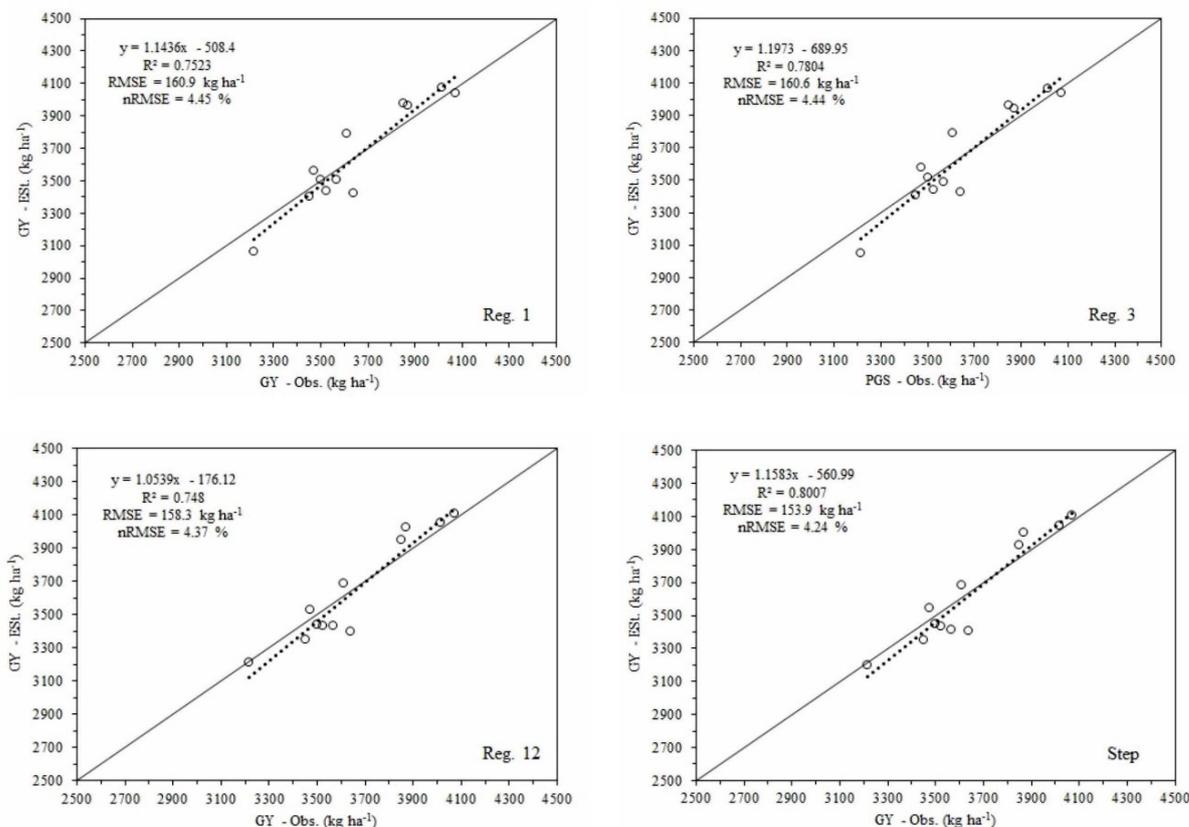


Figure 6. Validation of regression models with the best fit for predicting soybean grain yield.

## CONCLUSION

The linear regression models developed using the Normalized Difference Vegetation Index (NDVI), extracted from images of the Landsat 8 and Sentinel 2 satellites for soybean plants at V5-V6 and R2 developmental stages was promising for predicting soybean grain yield, with mean error of predictions of  $153.9 \text{ kg ha}^{-1}$ , corresponding to 4.2% when compared to the data from field measures.

## ACKNOWLEDGEMENTS

The authors thank the Farm Celeiro for providing the soybean production data, and logistic support for the conduction of this study.

## REFERENCES

ALLEN, R. G.; PEREIRA, L. S. Estimating crop coefficients from fraction of ground cover and height. *Irrigation Science*, 28: 17-34, 2009.

ANDRADE JUNIOR, A. S. et al. *Atlas Climatológico do Estado do Piauí*. Teresina, PI: Embrapa Meio-Norte. 2004. 151 p.

ATZBERGER, C. et al. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. *Remote Sensing*, 5: 949-981, 2013.

AKAIKE, H. A. New Look at Statistical Model Identification. *Transactions on Automatic Control*, 19: 717-723, 1974.

BALBINOT JUNIOR, A. A. et al. *Crescimento de plantas de soja em função da redução da densidade de semeadura e sua relação com a produtividade*. Londrina, PR: Embrapa Soja, 2018. 18 p. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1099933/1/BOLETIMPB18Alvadi.pdf>> Acesso em: 05 mar. 2019.

BARIANI, C. et al. Monitoramento da fenologia da

- soja irrigada usando perfis de série temporal de NDVI. In: INOVAGRI INTERNATIONAL MEETING, 3., 2015, Fortaleza. **Anais...** Fortaleza: INOVAGRI, 2015. p. 1610-1618.
- BERGER, R. et al. Índices de vegetação para a estimativa do índice de área foliar em plantios clonais de *Eucalyptus saligna* Smith. **Ciência Florestal**, 29: 885-889, 2019.
- BERTOLIN, N. O. et al. Predição da produtividade de milho irrigado com auxílio de imagens de satélite. **Revista Brasileira de Agricultura Irrigada**, 11: 1627-1638, 2017.
- CAMARGO, M. B. P.; BRUNINI, O.; MIRANDA, M. A. C. Modelo agrometeorológico para estimativa da produtividade para a cultura da soja no Estado de São Paulo. **Bragantia**, 45: 279-292, 1986.
- CELEIRO SEMENTES. **All Rights Reserved**. Brasília, 2020. Disponível em: <<http://celeirosementes.com.br/celeiro/portfolio-item/bmx-bonus-ipro-2/>>. Acesso em: 10 jul. 2020.
- CRUSIOL, L. G. T. et al. NDVI de estádios de desenvolvimento da soja BRS 284 em condições de campo. In: JORNADA ACADÊMICA DA EMBRAPA SOJA, 8., 2013, Londrina. **Anais...** Londrina: Embrapa Soja, 2013. p. 87-91.
- FORMAGGIO, A. R.; SANCHES, I. DEL A. **Sensoriamento remoto em agricultura**. São Paulo, SP: Oficina de Textos, 2017. 288 p.
- GUEDES, J.; SILVA, S. M. P. Sensoriamento remoto no estudo da vegetação: princípios físicos, sensores e métodos. **Acta Geográfica**, 12: 127-144, 2018.
- GUSSO, A. et al. Spectral model for soybean yield estimate using MODIS/EVI data. **International Journal of Geosciences**, 4: 1233-1241, 2013.
- KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **Annals of Mathematical Statistics**, 22: 79-86, 1951.
- MACEDO, C. E. S. et al. Detecção da variação de cobertura vegetal a partir da estimativa do índice de vegetação por diferença normalizada (NDVI) na unidade de conservação do Parque Estadual das Carnaúbas - CE com utilização de técnicas do sensoriamento remoto. **Anuário do Instituto de Geociências**, 43: 477-485, 2020.
- MELO, R. W. et al. An agrometeorological spectral model to estimate soybean yield, applied to southern Brazil. **International Journal of Remote Sensing**, 29: 4013-4028, 2008.
- 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.
- MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. Viçosa, MG: Ed. Universidade de Viçosa, 2003. 307 p.
- R CORE TEAM (2019). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Áustria. Disponível em: <<https://www.R-project.org/>>. Acesso em: 20 mar. 2019.
- RODIGHERI, G. et al. Net primary productivity and dry matter in soybean cultivation utilizing datas of NDVI multi-sensors. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, 42: 431-436, 2020.
- RODRIGUES, C. P. et al. NDVI e fluxo de CO<sub>2</sub> em lavoura de soja no Rio Grande do Sul. **Revista Brasileira de Meteorologia**, 28: 95-104, 2013.
- ROUSE, J. W. et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA special publication**, 351: 309, 1973.
- SANTOS, H. G. et al. **Sistema Brasileiro de Classificação de Solos**. 5 ed. Brasília, DF: Embrapa, 2018. 356 p.
- SARMIENTO, C. M. et al. A spectral agrometeorological model for estimating soybean grain productivity in Mato Grosso, Brazil. **Engenharia Agrícola**, 40: 405-412, 2020.
- SCHIRMBECK, L. W. et al. Condições hídricas de lavoura de soja usando sensoriamento remoto terrestre. **Agrometeoros**, 27: 173-181, 2019.
- SILVA, D. F. **Uso de modelos agrometeorológicos de estimativa de produtividade e de risco climático para a soja no vale do médio Paranapanema - SP**. 2011. 80 f. Dissertação (Mestrado em Agricultura Tropical e Subtropical: Área de concentração: Gestão de Recursos Agroambientais) - Instituto Agrônomo de Campinas, Campinas, 2011.
- SILVA, V. D. P. R. et al. Calibration and validation of the AquaCrop model for the soybean crop grown under different levels of irrigation in the Matopiba region, Brazil. **Ciência Rural**, 48: 1-8, 2018.

TRINDADE, F. S. et al. A. Relação espectro-temporal de índices de vegetação com atributos do solo e produtividade da soja. **Revista de Ciências Agrárias**, 62: 1-11, 2019.