Predicting soybean grain yield using aerial drone images


ABSTRACT: This study aimed to evaluate the ability of vegetation indices (VIs) obtained from unmanned aerial vehicle (UAV) images to estimate soybean grain yield under soil and climate conditions in the Teresina microregion, Piauí state (PI), Brazil. Soybean cv. BRS-8980 was evaluated in stage R5 and submitted to two water regimes (WR) (100 and 50% of crop evapotranspiration - ETc) and two N levels (with and without N supplementation). A randomized block design in a split-plot scheme was used, in which the plots were the water regimes and the subplots N levels, with five replicates. Each plot contained twenty 4.5 m-long rows, spaced 0.5 m apart, with a total area of 45 and 6 m² study area for grain yield evaluations. Twenty VIs obtained from multispectral aerial images were evaluated and correlated with grain yield measurements in the field. Pearson's correlation, linear regression, and spatial autocorrelation (Global and Local Moran’s I) were used to analyze the performance of the VIs in predicting grain yield. The R², RMSE and nRMSE indices were used to validate the linear regression models. The prediction model based on EVI-2 exhibited high spatial randomness for all the treatments, and smaller prediction errors of 149.68 and 173.96 kg ha⁻¹ (without and with N supplementation, respectively).

Key words: Glycine max L., remotely piloted aircraft, vegetation indices, autocorrelation, Moran’s I

RESUMO: O estudo objetivou avaliar a capacidade de índices de vegetação (IV) obtidos de imagens aéreas por veículo aéreo não tripulado em estimar a produtividade de grãos de soja, nas condições de solo e clima da microrregião de Teresina, Piauí, Brasil. Avaliou-se a cultivar de soja BRS-8980, em estádio R5, submetida a dois regimes hídricos (RH) (100 e 50% da evapotranspiração da cultura - ETc) e dois níveis de N (com e sem suplementação de N). O delineamento experimental foi o de blocos ao acaso, em parcelas subdivididas, sendo as parcelas os regimes hídricos e as subparcelas os níveis de nitrogênio, com cinco repetições. Cada parcela continha 20 linhas de 4,5 m de comprimento, espaçadas de 0,5 m entre linhas, com área de 45 m² e uma área útil para avaliação da produtividade de grãos de 6 m². Avaliaram-se 20 IV obtidos de imagens aéreas de câmera multispectral, os quais foram correlacionados com medidas de produtividade de grãos em campo. Adotaram-se análises de correlação de Pearson, de regressão linear e de autocorrelação espacial (índice I de Moran global e local) para geração e análise de desempenho dos IV na predição da produtividade de grãos. Para validação dos modelos de regressão linear, usou-se os índices R², RMSE e nRMSE. O modelo de predição baseado no EVI-2 apresenta elevada aleatoriedad espacial, para todos os tratamentos avaliados, e menores erros de predição iguais a 149,68 kg ha⁻¹ (sem suplementação de N) e 173,96 kg ha⁻¹ (com suplementação de N).

Palavras-chave: Glycine max L., aeronave remotamente pilotada, índices de vegetação, autocorrelação, Moran’s I

HIGHLIGHTS:
- Prediction models based on VIs using the red, NIR and red edge bands better explained the variability in soybean grain yield.
- The variability of soybean grain yield was due more to water regimes than nitrogen supplementation.
- The soybean grain yield prediction model generated with the EVI-2 VI showed greater accuracy and spatial randomness.

1 Research developed at Embrapa Meio-Norte, Teresina, PI, Brazil
2 Embrapa Meio-Norte, Teresina, PI, Brazil
3 Universidade Federal do Piauí/Programa de Pós-Graduação em Agronomia, Teresina, PI, Brazil
4 Embrapa Agroenergia, Brasília, DF, Brazil
Predicting soybean grain yield using aerial drone images

**Introduction**

It is important to develop crop yield prediction models because they enable crop conditions to be monitored in the field and previous adoption of management practices aimed at achieving high yields (Silva Junior et al., 2018).

Remote sensing is a promising technique that allows the accurate, low-cost prediction of the biophysical and biochemical parameters of plants. The high yield and spatial accuracy in predicting these characteristics using unmanned aerial vehicles (UAV) images can help assess genotype behavior and management practices, detect biotic and abiotic stress, and predict crop yield, contributing to decision-making by producers (Barbedo, 2019).

The use of vegetation indices (VIs) is common in remote sensing studies. VIs are reliable algorithms for evaluating plant cover, vigor and growth dynamics, and nutritional status, among other applications (Xue & Su, 2017), making them a promising tool for crop yield prediction (Zhao et al., 2020).

Several studies have reported a correlation between VIs and yield in crops such as corn, wheat, cotton, and rice (Dong et al., 2015; Baio et al., 2018; Hassan et al., 2019; Zhao et al., 2020). Holzman et al. (2014), Christenson et al. (2016), Silva Junior et al. (2018), Maimaitijiang et al. (2020), and Schwalbert et al. (2020) conducted studies with soybean and concluded that VIs produced satisfactory and promising results for predicting soybean yield. However, studies in this area are still incipient in the Mid-North region of Brazil.

The present study aimed to evaluate the ability of vegetation indices (VIs) obtained from unmanned aerial vehicle (UAV) images to estimate soybean grain yield under the soil and climate conditions of the Teresina microregion, Piauí state (PI), Brazil.

**Material and Methods**

The experiment was conducted at Embrapa Mid-North in Teresina, Piauí state, Brazil (05° 02' 13'' S, 42° 47' 49'' W, and altitude of 72 m), from July to November 2019 (Figure 1). The climate in the region is classified as Aw, with a wet summer and dry winter (Medeiros et al., 2020). Average annual temperature and rainfall are 27.4 °C and 1,325 mm, respectively, with the latter concentrated between January and May (INMET, 2020). During the experimental period, the average maximum and minimum temperatures, and cumulative rainfall were 29.6°C, 27.8 °C and 18.6 mm, respectively. At the time of the flight (11:00 a.m.-12:00 p.m.), average maximum and minimum temperature, wind speed, and global solar radiation were 30.7°C, 29.3 °C, 2.8 m s⁻¹, and 406.8 W m⁻², respectively (INMET, 2020).

The soil at the experimental site is classified as eutrophic Ultisol (Melo et al., 2014) (Table 1). Fertilization management was based on the soil chemical properties. Nitrogen (urea), phosphorus (simple superphosphate), and potassium (potassium chloride) were applied to the soil and micronutrients to the leaves (Silva, 2021). A 1,000 kg N ha⁻¹ dose was applied to meet soybean N requirements throughout the crop season in order to achieve maximum productive potential, with an expected grain yield of 6 Mg ha⁻¹. For each Mg of soybean produced, 80 kg ha⁻¹ of nitrogen is needed and nitrogen fertilization efficiency is estimated at 50% (Hungria et al., 2001).

The BRS-8980 soybean cultivar (Embrapa) with a determinate growth habit, GM 8.9, and a 125-136-day growing season was used. A randomized block design in a split-plot scheme was adopted, in which the plots were the following water regimes: deficit WR (50% replacement of crop evapotranspiration-ETc) (I₀) and full WR (100% replacement of ETc) (I₁), and the subplots were the levels of nitrogen fertilization: without nitrogen (N₀) and with nitrogen (1000 kg ha⁻¹) (N₁), resulting in four treatments (I₀N₀, I₀N₁, I₁N₀, and I₁N₁), with five replicates. Each plot contained twenty 4-5 m-long rows, spaced 0.5 m apart, with a total area of 45 m² and 6 m² study area for grain yield evaluations.

Sowing was performed manually on 23/07/2019, distributing 20 seeds per meter in the furrow. The seeds were inoculated with *Bradyrhizobium japonicum* (SEMIA 5079 and 5080), at a ratio of 100 mL of inoculant to 7 kg of seeds. After...
germination, the seedlings were thinned, leaving 10 to 12 plants per linear m⁻¹. Harvesting was performed at 113 days after sowing (DAS) (13/11/2019) at the beginning of the R8 phase, cutting the plants in the four 3 m-long center rows spaced 0.5 m apart (6 m² per plot).

Irrigation management was performed based on crop evapotranspiration (ETo) replacement. The reference evapotranspiration (ETo) was estimated by the Penman-Monteith method (Allen et al., 1998) using global solar radiation data (MJ m⁻²), air temperature (°C), relative air humidity (%), and wind speed (m s⁻¹) obtained from an automatic agro-meteorological station located 500 m from the experimental area. The soybean crop coefficients proposed by the FAO (Allen et al., 1998) were used to estimate ETo. Irrigation was applied using a conventional fixed sprinkler system, with sprinklers spaced 12 m apart. Irrigation depth was controlled by installing two blocks of 12 collectors each, one in each water regime, spaced 3 m between the lateral lines of sprinklers, in the center of the experimental plot.

From sowing to stage V3, full irrigation was used for both treatments, followed by the application of different water regimes from stages V4 to R5, with 50 and 100% ETo replacement. After stage R5, full irrigation was resumed for all treatments. The cumulative irrigation depth during the soybean season in each water regime are shown in Table 2.

Soil moisture content was monitored using Campbell CS616 probes, with three probes in each water regime, two 0-0.30 m and one 0.30-0.60 m deep, with continuous moisture data acquisition and recording by a CR1000 datalogger. From the onset of irrigation until the application of differentiated water regimes (OT), average moisture content in the 0-0.30 m layer ranged between 13.0 and 14.2% in the treatments with 50 and 100% ETo, while in the 0.30-0.60 layer, variation was 8.6% (50% ETo) and 10.3% (100% ETo) (Figure 2).

Average moisture content in the 0-0.30 m layer under the different water regimes (OT-ETo) was 9.9 and 14.7% in the treatments with 50 and 100% ETo respectively, and 7.7% (50% ETo) and 12.7% (100% ETo) in the 0.30-0.60 m layer. From the beginning of irrigation until the flight date (62 DAS), average moisture content in the 0-0.30 m layer was 10.0 and 13.2% in the treatments with 50 and 100% ETo, respectively, and 7.6% (50% ETo) and 12.3% (100% ETo) in the 0.30-0.60 m layer (Figure 2). These data indicate that in the full irrigation WR, soil moisture was maintained at higher-than-critical levels to allow adequate soybean development and production. Under deficit irrigation WR, values below the critical moisture content and above the permanent wilting point were recorded during the study period, which is a limiting factor for soybean grain development and yield.

### Table 2. Cumulative irrigation depth (mm) in each water regime during the soybean crop season

<table>
<thead>
<tr>
<th>WR (% ETo)</th>
<th>S-OT</th>
<th>OT-FD</th>
<th>OT-ETo</th>
<th>ET-ETo</th>
<th>Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>125.3</td>
<td>84.0</td>
<td>90.6</td>
<td>56.1</td>
<td>356.0</td>
</tr>
<tr>
<td>100</td>
<td>127.1</td>
<td>159.4</td>
<td>167.7</td>
<td>57.7</td>
<td>511.9</td>
</tr>
</tbody>
</table>

WR - Water regimes (% ETo); S - Sowing date; OT - Onset of differentiated WR treatment (40 DAS); FD - Flight date (62 DAS); ET - End of differentiated WR treatment (72 DAS); EIT - End of irrigation treatment (93 DAS)

An XFLY X800 hexacopter (Xfly Brasil, Bauru, SP) UAV was used to obtain aerial images. A flight was performed on 24/09/2019, at 62 DAS, between 11:00 a.m. and 12:00 p.m. The flight was planned using Pix4D Capture® software (www.pix4d.com). The flight plan was created to ensure that the images were captured with an 80% lateral and frontal overlap, maintaining the flight line at 30 m above ground level, with GSD (ground sample distance) of ≈ 1.5 cm pixel⁻¹.

The multispectral images were acquired by a RedEdge Micasense sensor, capable of capturing five spectral bands: Red (668 nm), Green (560 nm), Blue (475 nm), NIR (840 nm), and RedEdge (717 nm). The images generated were georeferenced and corrected using GPS and the solar radiation sensor installed on the top of the aircraft. A radiometric calibration standard was also used to correct the images, which were saved in 16-bit tiff format. The orthomosaic images were produced with Pix4D Mapper® software.

The orthomosaic underwent a supervised classification process (maximum likelihood method), allowing the rasterization of the orthomosaic into two classes (soil and leaves). This enabled the removal of pixels classified as mosaic soil, ensuring that the VI's were estimated only with pixels classified as leaves. The process was performed using the Semi-Automatic Classification (SCP) plug-in of QGIS v. 3.16 (QGIS, 2021). The vector layer containing the pixels classified as leaves was subdivided into two equal parts, thus obtaining two distinct datasets for the training and validation phases of the VIs. The “Split Polygon” plug-in of QGIS v. 3.16 (QGIS, 2021) was used.
The spectral response of the soybean to the treatments was quantified using 20 VIs estimated from the multispectral image bands (R, G, B, red edge, and NIR) (Table 3). The multispectral indices were estimated with QGIS v. 3.16 (QGIS, 2021). The VI values of each subplot were extracted with the zonal statistics plugin of QGIS v. 3.16 (QGIS, 2021). To that end, the vector layer of the subplots was used, containing only the areas classified as leaves, subdivided into two equal parts (modeling and validation).

The strategy adopted for statistical analysis of the data consisted of the following steps: a) Pearson’s correlation analysis between grain yield data, measured in the plots, and the 20 VIs evaluated, also extracted at the experimental plot level, aimed at pre-selecting the most promising indices (r ≥ 0.8); b) creation of grain yield prediction models using the most promising VIs via linear regression analysis, applying the coefficient of determination (R²) (Eq. 1) and standard error of estimate (SEE) as the selection criteria (Eq. 2); c) validation of grain yield determination (R²) (Eq. 1) and standard error of estimate (SEE) via linear regression analysis, applying the coefficient of grain yield prediction models using the most promising indices (r ≥ 0.8); b) creation of the most promising prediction models, from steps a, b and c, through Moran’s global index (Eq. 5) and Moran’s local index (Eq. 6) (Anselin, 1995), using the residuals of grain yield prediction models (RMR) (Eq. 7) to assess the degree of model randomness, as recommended by Maimaitijiang et al. (2020). For the estimates of R², SEE, RMSE, and nRMSE, the Excel Real Statistics Resource Pack (version 7.6) supplement was used (Zaiontz, 2020).

\[
R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \quad (1)
\]

\[
\text{SEE} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}} \quad (2)
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-1}} \quad (3)
\]

Table 3. Multispectral vegetation indices evaluated in the study

<table>
<thead>
<tr>
<th>Bands/Indices</th>
<th>Acronym</th>
<th>Equation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll index red</td>
<td>CI-RED</td>
<td>(\frac{R_n}{R_r} - 1)</td>
<td>Gitelson et al. (2005)</td>
</tr>
<tr>
<td>Chlorophyll vegetation index</td>
<td>CVI</td>
<td>(\frac{R_n R_g}{R_g^2})</td>
<td>Vincini et al. (2008)</td>
</tr>
<tr>
<td>Enhanced vegetation index</td>
<td>EVI</td>
<td>(2.5(R_n - R_b))</td>
<td>Gitelson et al. (2005)</td>
</tr>
<tr>
<td>Two-band enhanced vegetation index</td>
<td>EVI-2</td>
<td>(\frac{2.5(R_n - R_b)}{R_n + 2.4R_b + 1})</td>
<td>Jiang et al. (2008)</td>
</tr>
<tr>
<td>Green chlorophyll index</td>
<td>GCI</td>
<td>(\frac{R_n}{R_g} - 1)</td>
<td>Gitelson et al. (2005)</td>
</tr>
<tr>
<td>Green normalized difference vegetation</td>
<td>GNDVI</td>
<td>(\frac{R_n - R_g}{R_g + R_g})</td>
<td>Hunt &amp; Daughtry (2018)</td>
</tr>
<tr>
<td>Green ratio vegetation index</td>
<td>GRVI</td>
<td>(\frac{R_n}{R_g})</td>
<td>Sripada et al. (2006)</td>
</tr>
<tr>
<td>Modified excess green</td>
<td>MEXG</td>
<td>(1.262R_g - 0.884R_r - 0.311R_b)</td>
<td>Burgos-Artizu et al. (2011)</td>
</tr>
<tr>
<td>Modified normalized green red difference</td>
<td>MNGRD</td>
<td>(\frac{R_g^2 - R_r^2}{R_g^2 + R_r^2})</td>
<td>Bendig et al. (2015)</td>
</tr>
<tr>
<td>Normalized difference Red-Edge</td>
<td>NDRE</td>
<td>(\frac{R_n - R_{RE}}{R_n + R_{RE}})</td>
<td>Wang et al. (2014)</td>
</tr>
<tr>
<td>Normalized difference vegetation index</td>
<td>NDVI</td>
<td>(\frac{R_n + R_{RE}}{R_n - R_r})</td>
<td>Gitelson et al. (2005)</td>
</tr>
<tr>
<td>Normalized green red difference</td>
<td>NGRD</td>
<td>(\frac{R_n - R_g}{R_g + R_g})</td>
<td>Hamuda et al. (2016)</td>
</tr>
<tr>
<td>Optimized Soil Adjusted Vegetation Index</td>
<td>OSAVI</td>
<td>(\frac{R_n + R_r + 0.16}{R_n + R_r + R_g})</td>
<td>Roundeaux et al. (1996)</td>
</tr>
<tr>
<td>Pigment-specific normalized difference index</td>
<td>PSND</td>
<td>(\frac{R_n}{R_n - R_b})</td>
<td>Blackburn (1998)</td>
</tr>
<tr>
<td>Renormalized Difference Vegetation Index</td>
<td>RDVI</td>
<td>(\frac{(R_n - R_r)^0.5}{R_{RE}})</td>
<td>Roujean &amp; Breon (1995)</td>
</tr>
<tr>
<td>Red-Edge chlorophyll index</td>
<td>RECI</td>
<td>(\frac{R_n}{R_{RE}} - 1)</td>
<td>Gitelson et al. (2005)</td>
</tr>
<tr>
<td>Ratio vegetation index</td>
<td>RVI</td>
<td>(\frac{R_n}{R_r})</td>
<td>Gitelson et al. (2005)</td>
</tr>
<tr>
<td>Soil Adjusted Vegetation Index</td>
<td>SAVI</td>
<td>(\frac{1.5(R_n - R_r)}{(R_n + R_r + 0.5)})</td>
<td>Zhong et al. (2019)</td>
</tr>
<tr>
<td>Simplified canopy chlorophyll content index</td>
<td>SCCCI</td>
<td>(\frac{NDRE}{R_{RE}})</td>
<td>Raper &amp; Varco (2015)</td>
</tr>
<tr>
<td>Triangular greenness index</td>
<td>TGI</td>
<td>(R_g - (0.39R_b) - (0.61R_r))</td>
<td>Hunt et al. (2011)</td>
</tr>
</tbody>
</table>

\(R_g\) - Spectral reflectance - near infrared (840 nm); \(R_r\) - Spectral reflectance - green (560 nm); \(R_{RE}\) - Spectral reflectance - red edge (717 nm); \(R_n\) - Spectral reflectance red (668 nm) and \(R_b\) - Spectral reflectance - blue (475 nm)
RESULTS AND DISCUSSION

Pearson’s correlation shows the interaction between soybean grain yield and VIs (Table 4). There was greater positive interaction between GY and the RECI, EVI, EVI-2, OSAVI, SAVI, and GCI VIs, and low interaction between the MEXG, MNGRD, and TGI and GY VIs (Table 4).

The correlation shows the similarity between the VIs that use the same spectral bands, such as EVI, EVI-2, NDRE, and RECI, which use the red, NIR, and red-edge bands in their estimations, as well as the SAVI and OSAVI, which correct the soil effect on the vegetation signal (Nguy-Robertson et al., 2012; Silva Junior et al., 2018). The highest Pearson correlation values (r ≥ 0.8) were obtained for EVI, EVI-2, RDVI, SAVI, and OSAVI (Table 4).

Silva Junior et al. (2018) found a high correlation between SAVI and MSAVI and the GY of soybean cv. AS 3730 IPRO (Agroceres), possibly due to the use of soil effect correction to estimate these VIs (Hatfield & Prueger, 2010). SAVI, MSAVI, and MTVI are directly related to soil reflectance, crop canopy sensitivity, vegetation cover, and chlorophyll absorption (Haboudane et al., 2004).

Linear regression performed with the VIs selected as the most significant in the correlation analysis (r ≥ 0.8) indicated that EVI-2, EVI, RDVI, and SAVI were promising for estimating GY (Figure 3). EVI-2 was more efficient in estimating the GY of the cv. BRS-8980 (R² = 0.715; SEE = 324.85 kg ha⁻¹). The worst fit for the regression models was the SAVI vegetation index (R² = 0.70; SEE = 332.83 kg ha⁻¹) (Figure 3). It is important to emphasize the ability of the linear regression models to better estimate cv. BRS-8980 GY, demonstrated by the lower SEE values obtained by the regression models.

The VIs that best estimated GY were those that used the bands in the NIR, red, and red-edge regions to express the spectral response of soybean to the treatments applied (water regimes and soil N levels). A similar trend was observed by Christenson et al. (2016), who concluded that yield prediction models using VIs that adopt the red, NIR, and red-edge spectral bands better explained grain yield variability among soybean genotypes.

Christenson et al. (2016) generated a GY prediction model using different VIs for soybean cultivars grouped into two maturity groups (G3 and G4). The authors found that the resulting model exhibited R² = 0.58 and SEE = 703.0 kg ha⁻¹, performance indices lower than those obtained with the models generated in the present study. The authors attribute this poor performance to the fact that the model was generated for the

Table 4. Pearson’s correlation between vegetation indices (VIs) and soybean grain yield (GY, kg ha⁻¹) (n = 20)

<table>
<thead>
<tr>
<th>VI*</th>
<th>EVI</th>
<th>EVI-2</th>
<th>OSAVI</th>
<th>SAVI</th>
<th>RECI</th>
<th>R²</th>
<th>SEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIR</td>
<td>0.789***</td>
<td>0.708***</td>
<td>0.820***</td>
<td>0.802***</td>
<td>0.828***</td>
<td>0.715</td>
<td>324.85</td>
</tr>
<tr>
<td>CVI</td>
<td>0.781***</td>
<td>0.751***</td>
<td>0.803***</td>
<td>0.820***</td>
<td>0.828***</td>
<td>0.707***</td>
<td>332.83</td>
</tr>
<tr>
<td>EVI</td>
<td>0.839***</td>
<td>0.870***</td>
<td>0.802***</td>
<td>0.803***</td>
<td>0.828***</td>
<td>0.715***</td>
<td>324.85</td>
</tr>
<tr>
<td>EVI-2</td>
<td>0.845***</td>
<td>0.740***</td>
<td>0.802***</td>
<td>0.828***</td>
<td>0.828***</td>
<td>0.715***</td>
<td>324.85</td>
</tr>
</tbody>
</table>

* For description of VIs see Table 3. Significance levels (t-test): ** - p ≤ 0.05; *** - p ≤ 0.01; **** - p ≤ 0.001

Validation of the linear regression models showed that among the promising VIs in predicting soybean GY, cv BRS-8980 stood out as the most efficient (higher $R^2$ associated with lower RMSE and nRMSE values). The VIs that performed best were EVI-2, RDVI, EVI and SAVI, with RMSE, and nRMSE values ranging from 303.55 kg ha$^{-1}$ and 15.5% ($R^2 = 0.727$), respectively, for the model generated with EVI-2, at 309.25 kg ha$^{-1}$; and 15.8% ($R^2 = 0.718$), for the model generated with SAVI (Figure 4).

Holzman et al. (2014) conducted a study to estimate soybean grain yield at a regional scale in the Pampas region of Argentina using the temperature vegetation dryness index (TDVI) generated from EVI and surface temperature ($T_s$). Despite the regional scope of the study, the model proved to be effective in predicting GY. Depending on the agro-climatic zone, the $R^2$ values ranged from 0.68 to 0.79, with RMSEs of 366 and 380 kg ha$^{-1}$ and nRMSEs of 12.0 to 13.0%, respectively, very close to those obtained in the present study. The authors concluded that there is a strong correlation between TDVI and soil moisture measurements, with $R^2$ values ranging from 0.61 to 0.83, as well as adequate agreement with the spatial pattern of soil moisture, which explains the good performance of the model.

Maimaitijiang et al. (2020) evaluated the joint use of RGB, multispectral, and thermal sensors embedded in RPAs to estimate soybean grain yield using artificial neural networks. The authors concluded that the most accurate GY estimate ($R^2 = 0.72$; RMSE = 478.9 kg ha$^{-1}$; nRMSE = 15.9%) was obtained with the combined use of all the sensors evaluated. However, it is noteworthy that they also obtained adequate accuracy ($R^2 = 0.52$; RMSE = 630.5 kg ha$^{-1}$; nRMSE = 20.9%) using only multispectral sensors, confirming the results obtained in the present study.

Figure 5 shows the spatial variability of the GY generated with the EVI-2, RDVI, EVI, and SAVI VIs, with better performance in estimating GY in response to the water regimes evaluated (WR-50% ETc and WR-100% ETc). The maps show considerable similarity to each other for the same water regime. The maps of GY at WR = 50% ETc show areas with a greater predominance of GY below 1500-2000 kg ha$^{-1}$, while under WR = 100% ETc, the areas with greater predominance of GY are above 2000-2500 kg ha$^{-1}$ (Figure 5).

The level of soil water availability for plants directly affects soybean grain yield. Gava et al. (2015) found that imposing a water deficit of 70% ETc throughout the growing season reduced the growth and quality of grains per plant, lowering grain yield. According to Vivan et al. (2013), grain yield increases with the application of the ideal irrigation depth required by the crop. The irrigation depth with full ETc replacement was 511.9 mm (Table 2). Studies indicate that, depending on the soil conditions, climate, sowing date and cultivar, soybean requires an irrigation depth of 450 to 850 mm during the growing season to achieve maximum grain yield (Allen et al., 1998).

Distribution of the prediction errors of the soybean GY estimation models showed spatial clustering patterns at different levels as a function of the water regimes and N levels applied to the soil (Table 5 and Figure 6).
values range from -1 to 1, showing negative to positive spatial autocorrelation, respectively, with values close to zero and significant, indicating spatial randomness, where greater randomness indicates better regression model performance (Maimaitijiang et al., 2020).

The soybean GY prediction models based on EVI-2 provided the smallest average prediction errors (RMSE and nRMSE), ranging from 149.68 kg ha$^{-1}$ (nRMSE = 8.31%) (N0) to 173.96 kg ha$^{-1}$ (nRMSE = 10.24%) (N1) (Table 5). The better performance and spatial randomness of these models in predicting GY (Figure 6) occurs because they use the bands in the NIR, red, and red-edge regions to express the spectral response of soybean to the spatial variability of the water regimes and soil N levels (Christenson et al., 2016).
Continuation of Figure 5

Figure 5. Yield maps of soybean cv. BRS-8980, generated with the promising vegetation indices (VIs), in response to the water regimes evaluated (50 and 100% ETc)

Table 5. Spatial autocorrelation metrics (Global Moran’s I, RMSE and nRMSE) obtained with the regression models for estimating the GY of cv. BRS-8980 with promising vegetation indices (VIs) in response to the N levels evaluated

<table>
<thead>
<tr>
<th>IV*</th>
<th>Metrics</th>
<th>NO</th>
<th>N1</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI</td>
<td>Moran’s I</td>
<td>Q1</td>
<td>84.86</td>
<td>79.81</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>Med.</td>
<td>147.46</td>
<td>177.39</td>
</tr>
<tr>
<td></td>
<td>nRMSE</td>
<td>Med.</td>
<td>3.88</td>
<td>4.87</td>
</tr>
<tr>
<td></td>
<td>Q3</td>
<td>10.04</td>
<td>12.75</td>
<td>11.91</td>
</tr>
<tr>
<td></td>
<td>Moran’s I</td>
<td>Q1</td>
<td>-0.01449***</td>
<td>-0.01449*</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>Med.</td>
<td>149.68</td>
<td>173.96</td>
</tr>
<tr>
<td></td>
<td>nRMSE</td>
<td>Med.</td>
<td>4.23</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td>Q3</td>
<td>10.19</td>
<td>13.39</td>
<td>11.81</td>
</tr>
</tbody>
</table>

* For description of IV see Table 3. Moran’s I: Global Moran’s spatial correlation index (Eq. 5); RMSE - Root mean square error (kg ha⁻¹) (Eq. 3); nRMSE - Normalized root mean square error (kg ha⁻¹) (Eq. 4); Q1 - Lower quartile; Med - Mean; Q3 - Upper quartile; Pseudo p-values: *** - p ≤ 0.001; ** - p ≤ 0.01; * - p ≤ 0.05; o - p ≤ 0.1; ns - Not significant

The global Moran’s index values obtained for the GY prediction model generated with the EVI-2 VI were low and statistically significant for all N levels, ranging from -0.01449, for N0 (p ≤ 0.001) and N1 (p ≤ 0.05) separately, to -0.00719 (p ≤ 0.05), for all N levels (Table 5 and Figure 6). This indicates that the model generated a low spatial clustering pattern for GY prediction errors, that is, high model randomness (Figure 6).

However, the values of the global Moran’s index obtained for the EVI and SAVI models only indicated high randomness in the prediction of soybean GY with and without N supplementation, respectively, and should not be used, given the limitation of the models for this specific condition. For prediction models based on the RDVI, there was no spatial randomness for any of the conditions evaluated (Table 5). The models should be able to express the GY prediction randomness of different soybean cultivars in any environmental condition (Li et al., 2014).

In a study of soybean grain yield prediction based on the aerial images of RGB, multispectral and thermal sensors associated with different deep learning network methods, Maimaitijiang et al. (2020) found Moran’s I values in relation to grain yield model prediction errors ranging from 0.028 to 0.144 (p ≤ 0.001) when using the ‘inverse distance’ clustering method, and 0.045 to 0.153 (p ≤ 0.001) with the ‘contiguity edges’ method. They concluded that the fusion of all imaging strategies led to the weakest clustering pattern of prediction errors of all regression models evaluated, indicating that the joint adoption of all multimodal data can improve the adaptability of the model in space by weakening the spatial clustering of prediction errors.

Evaluation of the grain yield prediction model based on EVI-2 indicated that the lowest local Moran’s index values (≤ 0.005), with significant pseudo-p-values (p ≤ 0.05), were observed equally in the subplots with and without N supplementation, with 35 and 33 subplots, respectively. The evaluation of all subplots (N0 and N1) revealed that the lowest local Moran’s I (≤ 0.005) were observed in 78 subplots, 61 of which had pseudo-p-values less than 5% (p ≤ 0.05), indicating high spatial randomness of the soybean grain yield prediction model (Figure 6).
Maimaitijiang et al. (2020) found a better spatial distribution pattern of the residuals of the soybean grain yield prediction models, with values ranging from 50 to 400 kg ha⁻¹, with the joint use of RGB, multispectral, and thermal sensors embedded in the UAV under study to estimate grain yield using artificial neural networks. The authors attributed this superior performance to better efficacy and high randomness in spatial correlation, represented by the low Moran’s I values of these prediction models.

The spatial pattern of grain yield variability is conditioned by spatial factors (soil properties, irrigation management, and crop practices), temporal factors (soil pathogens, diseases, and crop production problems) (Peralta et al., 2016), and the cultivars studies (Haghighatallah et al., 2017). Thus, the importance of evaluating spatial autocorrelation in studies proposing prediction models for soybean GY is due to the risk of inaccurate estimates, distorted variance or, more importantly, erroneous conclusions (Peralta et al., 2016).

The superior performance of the soybean GY prediction models may also be related to the development stage of the cultivar used to generate the GY prediction models. In this study, aerial images obtained at the R5 stage were used to create the GY prediction models. Several studies have shown that the ideal stage of soybean development for predicting grain yield is between flowering and initial grain filling (stages R2 to R5) (Gao et al., 2018). In particular, the initial stage of grain filling (R5) was considered the best for predicting grain yield (Zhang et al., 2019).

Figure 6. Local Moran’s spatial autocorrelation index (I) and pseudo-p-values for the grain yield of soybean cv. BRS-8980, generated with the promising vegetation index - (EVI-2), in response to the N levels evaluated (with - N1 and without - N0 supplementation)

Conclusion

The soybean grain yield prediction model based on EVI-2 exhibits high spatial randomness, for all evaluated treatments, and smaller prediction errors of 149.68 kg ha⁻¹ (without N supplementation) and 173.96 kg ha⁻¹ (with N supplementation).

Literature Cited

Predicting soybean grain yield using aerial drone images


