ABSTRACT

Precision agricultural technologies, such as the use of spatial variability of soil properties, have been extensively studied for soybean cultivation. The objective of this study was to analyze the spatial variability of soil properties cultivated with soybean and to correlate the healthy vegetation (HV) spectral index with the bands B8A (classifying vegetation - 865 nm), B11 (measuring the moisture content of soil and vegetation - 1610 nm), B02 blue (useful for soil and vegetation discrimination - 490 nm). A sampling grid was installed for data collection in an area of 2,126.02 ha, with 270 regular points and 98 random points, totaling 368 points. For the soil, the contents of P (resin), K\(^+\), Ca\(^{2+}\), Mg\(^{2+}\), H\(^+\), Al\(^{3+}\), pH values, sum of bases (SB), cation exchange capacity (CEC), and base saturation were determined at a depth of 0.0 to 0.20 m. Most of the soil properties had exponential and spherical dependence. Clay percentages and Ca, Mg, and P contents had positive spatial correlation with the healthy vegetation spectral index (HV) while no spatial correlation was observed for pH, B, K, silt, sand, S, H+Al, Al, SB, and CEC. The sensor image used in this study in relation to HV showed good application for observing the spatial variability of the soil properties and soybean yield.

Key words: Glycine max, spatial variability, direct seeding system.

Correlations between soil properties and spectral index (healthy vegetation) in soybean crops

Correlaciones entre propiedades del suelo e índice espectral (vegetación saludable) en cultivos de soya

Cassiano Garcia Roque\(^1\), Job Teixeira de Oliveira\(^*\), Fabio Henrique Rojo Baio\(^1\), Othon Lauar Godinho\(^1\), and Fernando França da Cunha\(^2\)

RESUMEN

Las tecnologías de agricultura de precisión, como el uso de la variabilidad espacial en las propiedades del suelo, han sido ampliamente estudiadas para el cultivo de soya. El objetivo de este estudio fue analizar la variabilidad espacial de las propiedades del suelo cultivado con soya y correlacionarla con el índice espectral de vegetación saludable (VS), con las bandas B8A (clasificación de la vegetación (865 nm)), B11 (medición del contenido de humedad del suelo y la vegetación (1610 nm)), B02 azul (útil para la discriminación del suelo y la vegetación (490 nm)). La malla de muestreo se instaló para la recolección de datos en un área de 2,126.02 ha, con 270 puntos regulares y 98 puntos aleatorios, totalizando 368 puntos. Se determinaron en suelo los contenidos de P (resina), K\(^+\), Ca\(^{2+}\), Mg\(^{2+}\), H\(^+\), Al\(^{3+}\), valores de pH, suma de bases (SB), capacidad de intercambio catiónico (CIC), y saturación de bases en una muestra del suelo de 0.0 a 0.20 m de profundidad. La mayoría de las propiedades del suelo tuvieron dependencia exponencial y esférica. El porcentaje de arcilla y los contenidos de Ca, Mg y P mostraron una correlación espacial positiva con el índice espectral de vegetación saludable (VS) mientras que no se observó correlación espacial con los valores de pH, B, K, limo, arena, S, H+Al, Al, SB y CIC. La imagen del sensor utilizada en este estudio en relación con la VS mostró una buena aplicación para observar la variabilidad espacial de las propiedades del suelo y la producción de soya.

Palabras clave: Glycine max, variabilidad espacial, sistema de siembra directa.

Introduction

In Brazil, the growing expansion of soybean cultivation highlights its importance and relevance on the national and global scene as one of the largest soybean producers in the world. The 2023/2024 national harvest reached 160 million t, 3.6% above the volume harvested in the previous season and an absolute production record until then (Conab, 2023).

Vegetation indices can be used as productivity indicators for a crop for different types of vegetation studies and vegetation monitoring. For example, the indices can be used to monitor vegetation activity and health, in addition to monitoring drought conditions and plant senescence (Semeraro et al., 2019). In fact, these indices are capable of characterizing variations in the phenology and photosynthetic
potential of crops that are useful for identifying cultivation and the growth cycle (Malladi & Sowlati, 2017).

Monitoring and mapping crops on a large scale are essential for assisting in management and decision-making for various crops, thus improving production efficiency more technologically (Chen et al., 2023; Zhao et al., 2023). However, remote sensing is an effective strategy, allowing better precision in agricultural monitoring. A more sophisticated imaging technique called hyperspectral imaging uses data collected in a wide spectral range with the aim of reconstructing a spatial representation of the plant under analysis through image processing procedures generating highly specialized healthy vegetation (HV) maps (Otome et al., 2024).

The spatial analysis of soil chemical properties can indicate management alternatives, not only to reduce the effects of soil variability on crop (Rodrigues, Roque et al., 2023) and plant production but also to increase the possibility of estimating responses of crops under certain management practices. Such techniques provide methods to quantify this spatial autocorrelation to incorporate it in the estimation of values in unobserved locations (Oliveira, Oliveira, Rojas Plazas, Andrade et al., 2023).

Geostatistics have been shown to be the most effective method for analyzing spatial distribution features and patterns of variation of soil properties (Cortes et al., 2023). The semivariogram can be used to describe the spatial variability of soil properties (Oliveira, Oliveira, Rojas Plazas & Roque, 2023). The use of geostatistical techniques facilitates the interpretation of the behavior of soil properties for better decision-making in management. When analyzing a lot of information simultaneously, multivariate statistics become the best tool. Therefore, some studies have applied multivariate techniques to evaluate soil variables and they have found satisfactory results (Oliveira, Oliveira, Rojas Plazas, Andrade et al., 2023; Rodrigues, Castro et al., 2023).

The objective of this study was to analyze the spatial variability of soil attributes and correlate it with the healthy vegetation (HV) spectral vegetation index in a field cultivated with soybean.

**Materials and methods**

This study of the spatial variability of the reflectance index in soybean cultivation was carried out at Fazenda Emiliana in the municipality of Balsas (MA), Brazil, in 2019 (Fig. 1). The region’s climate is tropical, type Aw according to Köppen and Geiger, with an annual mean temperature of 26.4°C and a mean annual rainfall of 1190 mm.

We used an orbital image from the sentinel-2 satellite with a spatial resolution of 10 m, captured on January 8, 2019 with a healthy vegetation (HV) image, already processed and acquired by the earth observing system (EOS), adopting the WGS 84 datum and referencing the reproductive growth period of the R5 crop. Visual georeferencing was carried out using a polygon of the study region.

The soil was classified as a typical dystrophic red latosol with a very clayey to moderate dystrophic texture (Santos et al., 2018). The soil had been cultivated with a succession of soybean/corn crops sown in the summer and off-season. The Monsoy soybean variety was sown on November 20, 2018. We sowed the crop using a 0.50 m space between rows with a density of 14 seeds m⁻¹ resulting in a density of 280,000 plants ha⁻¹. We covered the seeds with millet straw in a direct sowing system. Common cultivation practices such as phytosanitary and chemical treatments were carried out homogeneously throughout the experimental area.

We collected data from three plots totaling 2,126.02 ha where we determined soil and plant sample collection points using a regular mesh distributed randomly over the plots with 270 regular points and 98 random points totaling 368 points (Fig. 2).

The soil samples were collected on April 10, 2019 using a caneco auger at the 0.0-0.20 m depth with 4 simple subsamples collected to obtain a composite sample to determine soil properties. Analysis for soil chemical properties included the following: pH, phosphorus content (P) (mg dm⁻³), potassium content (K) (mmolc dm⁻³), calcium (Ca) (mmolc dm⁻³), magnesium (Mg) (mmolc dm⁻³), aluminum (Al) (mmolc dm⁻³), potential acidity (H+A) (mmolc dm⁻³), boron (B) (mg dm⁻³) and sulfur (S) (mg dm⁻³) according to Teixeira et al. (2017).

Statistical analysis included an exploratory analysis in which a variable under study was characterized; consequently, its behavior and data distribution were identified and evaluated. The following descriptive statistics were generated: mean, median, minimum, maximum, standard deviation, and coefficient of variation. To classify the coefficient of variation (CV), the following classes and magnitudes were adopted: low (CV ≤10%), medium (10% < CV ≤ 20%), high (20% < CV ≤ 30%) and very high (CV > 30%) variability (Oliveira et al., 2022).
FIGURE 1. Location of the study area in Brazil.

FIGURE 2. Sampling grid at Fazenda Emiliana located in the municipality of Balsas (MA), Brazil.
To evaluate the degree of relationship between the variables involved in the modeling process, a Pearson correlation analysis was carried out to calculate simple linear regressions for combinations, two by two, between all the soil properties studied. Therefore, linear correlations with higher values, that is, higher significance, were selected for regression modeling. A fundamental step that preceded geostatistical analysis was carried out as a careful exploratory analysis of the data.

The analysis of the spatial dependence of soil chemical attributes was carried out by calculating the semivariogram based on the stationarity assumptions of the intrinsic hypothesis. Adjustments to the simple semivariograms, depending on their models, were made primarily by initially selecting the following: a) the lowest sum of squared deviations (SSD), b) the highest coefficient of determination \( r^2 \), and c) the highest evaluator of the degree of spatial dependence (DSD). The final decision of the model that represented the adjustment was made by cross validation as well as to define the size of the neighborhood that provided the best kriging mesh (Oliveira et al., 2022).

For each soil property, the nugget effect \( (C_0) \) and the range \( (A_0) \) and threshold \( (C_0 + C) \) were related. The analysis of the dependence evaluator (DSD) was carried out according to Cambardella et al. (1994), modified by the GS+ Software according to Equation 1:

\[
DSD = \left( \frac{C}{(C_0+C)} \right) \times 100 \%
\]  
(1)

The proposed interpretation for the DSD was in accordance with Dalchiavon et al. (2012):

- DSD < 20% = very low dependence spatial variable (VL),
- 20% ≤ DSD < 40% = low dependency (LO),
- 40% ≤ DSD < 60% = medium dependence (ME),
- 60% ≤ DSD < 80% = high dependence (HI), and
- 80% ≤ DSD < 100% = very high dependency (VH).

In order to analyze the maps to define which had the best visual relationship with productivity, the relative deviation coefficient (RDC; Eq. 2) was proposed. It should be noted here that the RDC calculates the average difference in modulus of the interpolated values on a thematic map when compared with a map assumed as a reference. The justification is that the objective was to estimate departures from thematic maps using more friendly interpolators other than kriging that is considered the best interpolator but which has difficulties in its implementation according to Equation 2.

\[
RDC = \frac{\sum_{i=1}^{n} (|P_{ijk} - P_{ipad}|)}{n} \times 100
\]  
(2)

where: \( n \) = number of interpolated points; \( P_{ipad} \) = reference search at point \( i \); \( P_{ijk} \) = point test for sampling method.

**Results and discussion**

Analyzing the properties presented in Table 1, we found that H+Al (8.67%), silt (9.66%), clay (6.83%), yield (PROD) (6.81%), and pH (2.84%) showed low CV. In relation to the other attributes, the analyzes showed the following average CVs: Ca (15.20%), K (17.66%), Mg (19.93), sand (13.38%) and SB (15.99%) and very high CVs as follows: P (68.81%), B (79.47%), S (36.04%) and Al (51.68%). Analyzing the soil properties, we observed that the pH showed low variability in accordance with results obtained by Oliveira et al. (2020), whose study of the spatial variability of soil properties obtained results of 4.9% and 5.3% for depths of 0.00 to 0.10 m and 0.10 to 0.20 m.

The soil property of P content (68.81%) showed very high variability in line with Lima et al. (2013), who also found very high variability of 32.1% and 48.0% for phosphorus at both soil depths. The residual presence of fertilizers used in previous harvests in the study area may explain the change in phosphorus concentrations in the soil through the history of analyzes of the area (Oliveira et al., 2021).

Dalchiavon et al. (2011) suggest that the P content in soil may be related to direct sowing by not disturbing the soil, reducing contact between colloids and the phosphate ion, softening adsorption reactions, especially due to the fact that organic P, originating from the decomposition of the remaining roots along the soil profile, constitutes an important reserve of labile P for plants in the deeper layers of the soil.

Potassium (K) (17.66%) showed medium variability, different from that obtained by Dalchiavon et al. (2011), where the authors found very high variability for K (38.5%). The authors describe the high variability for K due to the residual presence of fertilizers previously used in the predecessor crop (corn).

In Table 1, both Ca (15%) and Mg (19.93%) showed average variability, disagreeing with Alves et al. (2014), who obtain magnesium CV values in a direct planting system of around 35% and for conventional tillage of 49%. The
non-conformity of the Ca and Mg levels can be explained by the failure to carry out liming in the last 3 years and that was only carried out in 2013. Bottega et al. (2013), evaluating the variability of Al, obtain very high CV values like those found in this research that can also be explained by the lack of liming.

The asymmetry and kurtosis coefficients were different from zero for all tested variables. The asymmetry was negative for properties Ca, Mg, silt, clay, PROD, Al, and SB. However, kurtosis showed negative values only for PROD. Similar results were found by Cambardella et al. (1994). Skewness represents the degree of deviation of a curve in the horizontal direction that can be positive with a greater concentration of high values or negative with a predominance of low values.

The study of Pearson’s linear correlations was presented through a network of correlations where the green lines represent positive correlations, and the red lines represent negative correlations. The thickness of the line represents the degree of correlation between the variables. In other words, the thicker the line, the greater the correlation (Fig. 3). There were positive and significant correlations of PROD with Ca ($r = 0.78$), Mg ($r = 0.87$), P ($r = 0.63$), Clay ($r = 0.81$), and HV ($r = 0.89$). Oliveira et al. (2020) also obtain positive direct relationships between bean productivity and pH, Mg, and soil base saturation. In Figure 3, positive correlations were also observed between Ca and Mg ($r = 0.94$), indicating significant correlations at 5% and demonstrating strong correlation.

<table>
<thead>
<tr>
<th>Property</th>
<th>Average</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>CV (%)</th>
<th>Asymmetry</th>
<th>Kurtosis</th>
<th>Pr&lt;W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca</td>
<td>2.70</td>
<td>2.57</td>
<td>2.54</td>
<td>2.72</td>
<td>15.20</td>
<td>-0.23</td>
<td>6.00</td>
<td>0.00</td>
</tr>
<tr>
<td>H+Al</td>
<td>2.60</td>
<td>2.65</td>
<td>1.90</td>
<td>3.80</td>
<td>8.67</td>
<td>1.07</td>
<td>7.57</td>
<td>0.00</td>
</tr>
<tr>
<td>K</td>
<td>0.09</td>
<td>0.08</td>
<td>0.06</td>
<td>0.17</td>
<td>17.66</td>
<td>3.29</td>
<td>4.09</td>
<td>0.00</td>
</tr>
<tr>
<td>Mg</td>
<td>1.21</td>
<td>1.12</td>
<td>1.16</td>
<td>1.21</td>
<td>19.93</td>
<td>-0.01</td>
<td>5.86</td>
<td>0.30</td>
</tr>
<tr>
<td>P</td>
<td>6.05</td>
<td>5.11</td>
<td>5.44</td>
<td>8.67</td>
<td>68.81</td>
<td>0.23</td>
<td>31.79</td>
<td>0.13</td>
</tr>
<tr>
<td>B</td>
<td>0.29</td>
<td>0.26</td>
<td>0.11</td>
<td>1.55</td>
<td>49.47</td>
<td>0.34</td>
<td>46.88</td>
<td>0.03</td>
</tr>
<tr>
<td>Silt</td>
<td>106.66</td>
<td>103.98</td>
<td>75.00</td>
<td>175.00</td>
<td>9.66</td>
<td>-0.12</td>
<td>10.49</td>
<td>0.00</td>
</tr>
<tr>
<td>Clay</td>
<td>580.55</td>
<td>574.30</td>
<td>529.88</td>
<td>632.19</td>
<td>6.83</td>
<td>-0.51</td>
<td>3.23</td>
<td>0.00</td>
</tr>
<tr>
<td>S</td>
<td>7.42</td>
<td>7.01</td>
<td>4.00</td>
<td>32.00</td>
<td>36.04</td>
<td>0.31</td>
<td>23.51</td>
<td>0.09</td>
</tr>
<tr>
<td>Sand</td>
<td>312.72</td>
<td>322.02</td>
<td>207.62</td>
<td>357.62</td>
<td>13.38</td>
<td>0.21</td>
<td>4.03</td>
<td>0.29</td>
</tr>
<tr>
<td>PROD</td>
<td>2996.25</td>
<td>3115.13</td>
<td>2662.6</td>
<td>3129.14</td>
<td>6.91</td>
<td>-1.97</td>
<td>-1.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Al</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.60</td>
<td>51.68</td>
<td>-0.24</td>
<td>207.17</td>
<td>0.00</td>
</tr>
<tr>
<td>pH</td>
<td>5.59</td>
<td>5.52</td>
<td>5.20</td>
<td>6.10</td>
<td>2.84</td>
<td>1.08</td>
<td>0.84</td>
<td>0.00</td>
</tr>
</tbody>
</table>

| P, pH, K, Ca, Mg, H+Al, Al, PROD, B, and S: phosphorus extracted by resin (mg dm$^{-3}$), hydrogen potential (pH$_{CaCl_2}$), potassium (mmolc dm$^{-3}$), calcium (mmolc dm$^{-3}$), magnesium (mmolc dm$^{-3}$), potential acidity (H+Al), exchangeable aluminum (mmolc dm$^{-3}$), soybean yield (PROD, kg ha$^{-1}$), boron (mg dm$^{-3}$), sulfur (mg dm$^{-3}$). CV: coefficient of variation; Pr<W: probability of the normality test.

**FIGURE 3.** Network of correlations of soybean yield, HV index, and some properties of a dystrophic red oxisol at Fazenda Emiliana in the municipality of Balsas (MA), Brazil. P, pH, K, Ca, Mg, H+Al, Al, PROD, B, S, and HV are: phosphorus extracted by resin, hydrogen potential, potassium, calcium, magnesium, potential acidity, exchangeable aluminum, soybean yield, boron, sulfur, and healthy vegetation index, respectively.
Considering the vegetation spectral index HV as the main variable for analysis, it resulted in the same positive correlations as PROD with the chemical attributes of the soil (Fig. 3), that is, HV established positive and significant correlations with Ca ($r = 0.69$), Mg ($r = 0.79$), and Clay ($r = 0.71$).

Figure 4 presents the linear correlations of soil properties with the HV of soybean crops. The simple linear correlation coefficients were found for the soil properties, and also the linear correlation was obtained between HV and PROD. Therefore, to correlate HV with soil properties, the analysis of spatial dependence is justified to better understand the pattern of occurrence of these properties in space that are very positive. In a study by Oliveira, Oliveira, Rojas Plazas, Andrade et al. (2023), the authors highlight the importance of the spatial autocorrelation of soil properties and soybean planting, presenting significance maps and correlation with crop productivity.

The geostatistical analysis (Tab. 2) showed that there was spatial dependence for the semivariogram of the Ca, Mg, and HV properties adjusted to the exponential model, while PROD and clay adjusted to the spherical model.

**TABLE 2.** Semivariogram parameters.

<table>
<thead>
<tr>
<th>Property</th>
<th>Model</th>
<th>$C_0$</th>
<th>$C_0+C$</th>
<th>$A_r$(m)</th>
<th>$R^2$</th>
<th>DSD %</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca</td>
<td>Exponential</td>
<td>0.00046</td>
<td>0.00092</td>
<td>222.3</td>
<td>0.677</td>
<td>50.10</td>
<td>ME</td>
</tr>
<tr>
<td>Mg</td>
<td>Exponential</td>
<td>0.00004</td>
<td>0.00007</td>
<td>1417.0</td>
<td>0.815</td>
<td>50.10</td>
<td>ME</td>
</tr>
<tr>
<td>PROD</td>
<td>Spherical</td>
<td>12750.00000</td>
<td>42970.00000</td>
<td>1273.0</td>
<td>0.982</td>
<td>70.30</td>
<td>HI</td>
</tr>
<tr>
<td>Clay</td>
<td>Spherical</td>
<td>127.100000</td>
<td>254.300000</td>
<td>1124.0</td>
<td>0.872</td>
<td>50.00</td>
<td>ME</td>
</tr>
<tr>
<td>HV</td>
<td>Exponential</td>
<td>0.09300</td>
<td>0.66300</td>
<td>67.0</td>
<td>0.473</td>
<td>86.00</td>
<td>VH</td>
</tr>
</tbody>
</table>

P, Ca, Mg, PROD; HV: phosphorus, calcium, magnesium, soybean yield, healthy vegetation index, respectively. ME: medium dependence; HI: high dependence; and VH: very high dependence; For each property, the nugget effect ($C_0$) and the range ($A_r$) and threshold ($C_0+C$) and degree of spatial dependence (DSD) were related.

**FIGURE 4.** Regression equations of the HV – healthy vegetation spectral index of soybeans with all properties that demonstrated a positive relationship. A) Ca (mmolc dm$^{-3}$), B) Mg (mmolc dm$^{-3}$), C) Clay (g), and D) yield (PROD, kg ha$^{-1}$).
In this way, the variables Ca, Mg, PROD, Clay, and HV showed spatial dependence, showing that the mesh used was sufficient for the study of spatial variability. According to Cortes et al. (2023), one of the ways to evaluate the performance of semivariograms is their analysis based on their respective spatial determination coefficients ($R^2$), so that the behavior according to the decreasing relationship of the best fits was: 1) PROD (0.98), 2) clay (0.87), 3) Mg (0.82), 4) Ca (0.677), and 5) HV (0.47).

Range is a parameter provided by geostatistics that represents the spatial behavior of the studied variable, indicating the distance at which the variable has spatial continuity (Oliveira, Oliveira, Rojas Plazas, Andrade et al., 2023). This parameter becomes important for precision agriculture due to the use of geostatistical packages to feed agricultural machinery for the application of inputs at variable rates. Thus, in decreasing order of range: 1) Mg (1414.00 m), 2) PROD (1273.00 m), 3) clay (1124.00 m), 4) Ca (222.30 m), and 5) HV (67 m).

To analyze the degree of spatial dependence of the variables, Dalchiavon et al. (2012) was used, where semivariograms with a nugget effect ($C_0/C_0+C_1$) were considered to have strong spatial dependence. Thus, Ca, Mg, clay, and P had a medium degree of spatial dependence and PROD and HV had high and very high dependence.

FIGURE 5. Kriging maps of the properties production, Ca, Mg, PROD, and Clay of attributes evaluated on the Emiliana farm in the municipality of Balsas (MA), Brazil.
A good application for observing the spatial variability of the studied soil properties, since it presented an excellent linear correlation with the productivity of the crop.

Among the mathematical models of geostatistics, the ones that best explained the results were the spherical and exponential models.

Conflict of interest statement
The authors declare that there is no conflict of interests regarding the publication of this article.

Author’s contributions
CGR: conceptualization, methodology, validation, formal analysis, investigation, writing - original draft; JTO: designed the experiment, analyzed the data, wrote, and edited the manuscript, final approval; FHRB: resources, writing, review and editing, final approval; OLG: validation, resources, writing - review and editing, final approval; FFC: validation, writing - review and editing, final approval. All authors reviewed the final version of the manuscript.

Literature cited


