BIOENG

Brazilian Journal of Biosystems Engineering (2025), 19 1261

REGULAR ARTICLE

Spatial variability and correlation of soybean yield components and harvest losses

Anna Kelly Severino Santos [1]; **Jorge Wilson Cortez** [1*]

[1] Faculty of Agricultural Sciences, Federal University of Grande Dourados (FCA/UFGD), Dourados/MS, Brazil

Regular Section

Academic Editor: *Celso Antonio Goulart*

Statements and Declarations

Data availability

The data presented in this study are available on request from the corresponding author.

Institutional Review Board Statement

Not applicable.

Conflicts of interest

The authors declare no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Code/Software availability

Not applicable.

Use of Generative AI

During the preparation of this work, the authors used ChatGPT (OpenAI, GPT-5.5) to support language editing and grammar correction. The authors reviewed and edited all AI-generated output and take full responsibility for the content of this publication.

Author contribution (CRediT)

A.K.S.S.: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – original draft; J.W.C.: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

Abstract

The spatial variability of yield components and harvest losses represents a determining factor for production efficiency in agricultural systems managed under the principles of precision agriculture. In this context, the present study aimed to evaluate the spatial variability of soybean yield components and harvest losses, as well as to investigate their relationship with the Normalized Difference Vegetation Index (NDVI). The study was conducted on a commercial farm in two production fields with areas of 70.46 ha and 63.14 ha. A regular sampling grid was established with the addition of 20% random points, totaling 82 sampling points per field. The variables analyzed were NDVI, plant height, first pod insertion height, grain yield, and harvest losses, classified as pre-harvest losses, total losses, and harvester losses. The data were subjected to descriptive statistical analysis, geostatistical analysis, spatial interpolation, and Pearson correlation. The results showed spatial variability for NDVI, plant height, first pod insertion height, and yield. The mean yield was 2846.2 kg ha^{-1} in field T1 and 4402.5 kg ha^{-1} in field T2. A positive correlation between NDVI and yield was observed in field T1. Total grain losses were above the acceptable limit of 60 kg ha^{-1}, ranging predominantly between 300 and 600 kg ha^{-1} (6.81 to 13.63% of production), with regions reaching 600 to 1200 kg ha^{-1} (13.63 to 27.26%). A strong relationship between harvester losses and total losses was also found, highlighting the influence of harvester adjustment and operating conditions on harvest efficiency.

Keywords

Precision agriculture; NDVI; Soybean yield

1. Introduction

Precision agriculture (PA) is a management approach that seeks to optimize processes and reduce costs, using advanced technologies to improve efficiency at all stages, from sowing to harvest, taking into account the spatial variability of crops, which promotes greater profitability and environmental conservation (Comparin and Cortez, 2023; Simas et al., 2023). Among the technologies used in PA, the use of satellite images for preparing vegetation index (VI) maps is a widely used technique. One of the most widely used is the Normalized Difference Vegetation Index (NDVI). VIs calculated from the spectral bands of images are effective tools for identifying agricultural problems (Hernández-López et al., 2021), provided that they are free from interference such as clouds (Ramadhani et al., 2021). Ré et al. (2025) demonstrated the effectiveness of VIs in detecting stress in sugarcane, although their interpretation may be affected by multiple factors. Lopes et al. (2024) identified significant correlations between NDVI, WDRVI (Wide Dynamic Range Vegetation Index), and maize yield. Filla et al. (2023) used NDVI to estimate common bean yield, obtaining R^2 values of up to 78% in the reproductive stages. Similarly, Batistella et al. (2023) studied soybean yield estimation using the EVI and NDVI indices obtained from sensors onboard satellites, concluding that the Random Forest algorithm showed promising results in estimating soybean yield, with greater influence of the images corresponding to crop maturation. Thus, these studies demonstrate the potential of VIs in modeling agricultural yield, including soybean crops. In addition to the use of spectral indices, PA also employs technologies embedded in modern grain harvesters, equipped with yield monitors, which, when integrated with several sensors, such as grain mass flow sensor, allow grain production values to be obtained in real time and yield maps to be generated, an essential tool for smart agriculture (Inacio and Cortez, 2023). However, one of the main challenges of mechanized soybean harvesting is associated with grain losses, which may occur both in systems that use precision agriculture technologies and in conventional systems. In general, these losses represent a critical factor for production efficiency and may compromise the economic sustainability of agricultural activity. In this context, the systematic monitoring of losses during the harvesting operation becomes essential, as well as the adoption of strategies for their reduction. Among the main measures for mitigating these losses, adequate operator training and the correct adjustment and calibration of the harvester stand out, factors directly related to the operational performance of the machine (Carreira et al., 2024). The spatial variability of soybean yield components and harvest losses represents one of the main challenges for production efficiency in commercial systems. In addition, the integration of vegetation indices, such as NDVI, makes it possible to anticipate spatial patterns of vegetative vigor, contributing to the predictive monitoring of yield and harvest efficiency. It is assumed that areas with greater vegetative vigor (NDVI) tend to show higher grain yield and lower

*Corresponding author

E-mail address: jorgecortez@ufgd.edu.br

https://doi.org/10.18011/bioeng.2025.v19.1261

Received: 03 February 2025 / Accepted: 12 March 2026 / Available online: 14 May 2026

relative losses, although increased plant biomass may influence the operational performance of the harvester. Given this context, the present study aimed to analyze the spatial variability of soybean yield components and harvest losses in a commercial area, as well as to investigate their spatial relationship with NDVI, aiming to support improvements in productive and operational efficiency.

2. Materials and methods

2.1 Location

The study was conducted on a commercial farm located in the municipality of Sidrolândia, MS, at latitude 20º55'55" South and longitude 54°57'39" West. The arable areas of the farm have been managed for more than 15 years under a no-tillage system (NTS). Two fields were selected on the property for carrying out the study. The first, designated as Field 1 (T1), has 70.46 ha, 64.32% clay, and a mean altitude of 360 m; the second, designated as Field 2 (T2), has 63.14 ha, 63.38% clay, and a mean altitude of 347 m. Both fields have soil classified as Oxisol, with a clayey texture. Soil sampling was performed to verify fertility levels, with sampling carried out according to the property's planning. Based on the results (Table 1), the area was previously corrected with the aim of reaching 70% base saturation, with dolomitic limestone applied at a variable rate. In general, a mean rate of 350 kg ha^{-1} was applied. Rainfall data were monitored using a rain gauge installed on the property, and it was found that monthly rainfall from November to March was below 100 mm; only in January 2021 did rainfall exceed 500 mm.

Table 1. Data on soil chemical attributes of fields 1 and 2.

Chemical attribute	pH in water	O. M.	P	K	Ca	Mg	Total CEC	V%
Field 01	6.16	29.38	8.73	3.59	34.20	13.70	83.04	61.86
Field 02	6.03	33.34	30.96	9.73	52.39	19.55	18.30	68.69

2.2 Crop establishment

The soybean crop was established under a no-tillage system in both fields. Field T1 originated from the sowing of second-crop maize in 2020, whereas in T2, Brachiaria brizantha cv. Marandu had been used as cover crop for two years. The area was previously desiccated fifteen days before sowing, and Bmx Compacta Ipro soybean seeds were used. The seeder was adjusted to distribute 14.61 seeds per meter (292,200 plants per ha) at a row spacing of 0.50 m in both fields, resulting in 12.23 plants per meter in T1 and 13.41 plants per meter in T2.

2.3 Vegetation index

For NDVI calculation, images were used based on the sowing date and availability from the Sentinel-2A imager, as follows: 01/03/2021 for T1 (51 days after sowing) and 12/19/2020 for T2 (45 days after sowing). NDVI was calculated using band B4 Red and band B8, corresponding to the near infrared (NIR). The NDVI interpretation classes were based on Chedid et al. (2024), as follows: ≤ 0.20 (Class 1); 0.20–0.40 (Class 2); 0.40–0.60 (Class 3); 0.60–0.80 (Class 4); and >0.8 (Class 5), indicating, respectively, exposed soil and straw, straw and early vegetative development, partial vegetative development, vegetative development, and full development.

2.4 Yield components

For the collection of data on plant height, first pod insertion height, and harvest losses, a regular sampling grid with 20% random points was prepared (adapted from Molin et al., 2015), with adjustments so that each field had 82 sampling points (Figures 1).

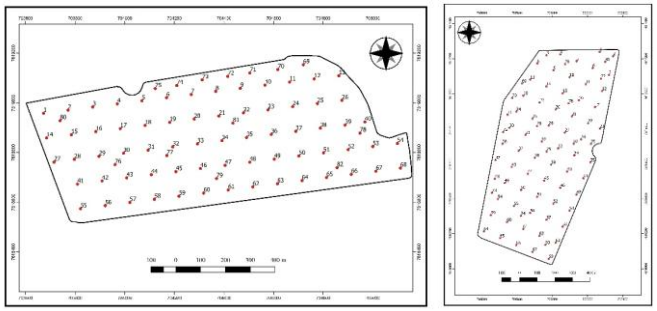

Figure 1. Sampling schemes of T1 (left) and T2 (right).

Plant height was determined by measuring, at the center of each point of the sampling grid, a total of 10 consecutive plants, from the soil surface to the insertion of the last fully expanded leaf (trifoliate), using a measuring tape (Ratke et al., 2024). To determine the first pod insertion height (FPIH), 10 consecutive plants were measured at each point of the sampling grid, given by the distance between the soil level and the productive FPIH, using a measuring tape. Yield data were obtained from grain harvesters equipped with installed sensors, namely a grain mass flow sensor and a moisture sensor, to measure the mass of harvested grains. During the harvesting process, the harvested grain data were measured by the mass sensor, and the moisture content of the harvested grains was also measured. The monitor was calibrated using grain mass sensors installed in the grain tank, using the technology system available in the harvester. The grain mass values obtained by the mass sensor were corrected to 13% moisture, using the data originally described by the harvester moisture sensor; thus, the final yield data used in this study were standardized at 13% moisture.

2.5 Harvest losses

Grain losses were determined at each point of the sampling grid (Figure 1) within a rectangular frame with an area of 2 m^2 (Mesquita et al., 1998), constructed with two PVC bars and two nylon cords, measuring 10.5 × 0.019 m, which coincided with the width of the harvester header. Losses were determined as: pre-harvest loss, total losses (after the harvester pass), and machine losses, given by the difference between total loss and pre-harvest loss.

2.6 Data analysis

The data were analyzed using the coefficient of variation (CV) and classified according to Lopes et al. (2021), using the mean data, which indicate low CV when below 9%; medium CV when from 9 to 27%; high CV when between 27 and 36%; and very high CV when above 36%. The yield data obtained by the harvester sensors were downloaded from the monitor and subjected to filtering, removing data with positioning errors, platform-width errors, and very high or very low value errors; data with values lower than 75% of the total platform width (<9.01 m) were removed; data with speed lower than 1.0 km h^{-1} were also removed. Subsequently, the data were

interpolated using the IDW method on a grid of approximately 9 × 9 m. Harvest loss data were initially subjected to descriptive analysis to obtain measures of central tendency and dispersion. All spatial analyses were performed in the free software QGIS (QGIS, 2024), and geostatistics followed by kriging were performed using the Smart-Map plugin (Pereira et al., 2022). Data not fitted to semivariograms were interpolated using inverse distance weighting (IDW). Therefore, the following were adjusted by IDW: FPIH and total losses in T1; yield and losses from the harvesting process for both fields. After interpolation of all data, they were grouped into a single matrix for each field and subjected to Pearson's correlation matrix.

3. Results and discussion

3.1 Vegetation index

The NDVI data for Field 1 (T1) showed medium dispersion (CV equal to 15.53%), whereas for Field 2 (T2), dispersion was low (CV equal to 8.34%), according to Lopes et al. (2021), respectively. NDVI for the fields showed a high range of values, varying between 0.16 and 0.95 (Figure 2). As the crop develops, NDVI values increase, approaching the dark-green color, indicating that the area was under full vegetation cover, which would correspond to the saturation point. Silva et al. (2022) observed that NDVI is sensitive to changes in the coloration of soybean plants, showing stand gaps across the cropped area, areas of healthy vegetation, and exposed soil.

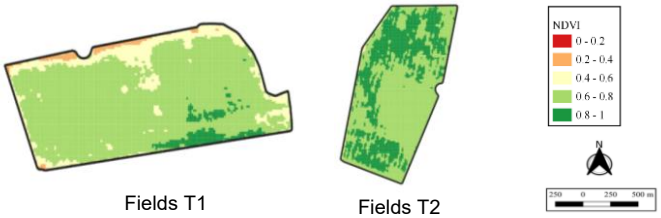

Figure 2. Spatial distribution of NDVI in Fields T1 and T2.

In T1 showed lower NDVI values (Figure 2), mainly along the borders of the area, indicating lower plant density, which may be attributed to the presence of cattle in the neighboring areas and the proximity to a permanent preservation area, where the presence of wild animals caused trampling along the border of the area. In general, approximately 75% of the area had a predominance of NDVI between 0.6 and 0.8, indicating vigorous vegetation, with approximately another 10% above 0.8. In T2, the amount of area with NDVI between 0.6 and 0.8 was around 53%, while the amount of area above 0.8 was approximately 33%. The vegetation cover of this field, with the presence of Brachiaria brizantha cv. Marandu before crop sowing, favored greater crop plant biomass. Thus, the results demonstrate that crop and vegetation monitoring using remote sensing satellites is important for monitoring agricultural production (Kasimati et al., 2023), but they highlight that proximal sensors better explain variability in the initial and final growth stages.

3.2 Yield components

Plant height (Figure 3) showed a coefficient of variation (CV) classified as medium for T1 (10.04%) and low for T2 (8.26%), according to Lopes et al. (2021). First pod insertion height (FPIH) had a CV classified as low for T1 (5.08%) and medium for T2 (10.3%), whereas grain yield showed a medium CV for T2 (14.26%) and a high CV for T1 (28.76%), according to Lopes et al. (2021) (Figure 3)

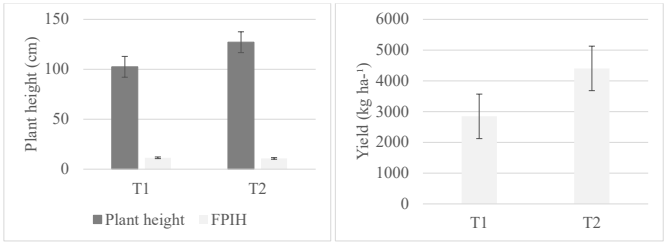

Figure 3. Plant height (cm), first pod insertion height (FPIH - cm), and yield (kg ha^{-1}) in Fields T1 and T2.

In field T1 (Figure 3), excessive rainfall in January 2021 during the growing season made it necessary to delay harvest, which negatively affected yield (2846.2 kg ha^{-1}) due to physiological deterioration. In field T2 (Figure 3), the presence of Brachiaria brizantha cv. Marandu may have favored higher grain yield (4402.5 kg ha^{-1}). The cover crop provides benefits to the production system, including a deep and aggressive root system, which contributes to nutrient cycling. Another benefit of soil vegetation cover is its capacity to promote an increase in organic matter, as can be verified in Table 1. The spherical model showed the best fit for the semivariogram of plant height and FPIH height (Table 2). Thus, the spatial dependence of the plants was observed based on the semivariogram range. The coefficients of determination (R^2) ranged from 0.68 to 0.96 for plant height and were 0.81 for FPIH. The FPIH variable for T1 did not fit any semivariogram model and was therefore interpolated using the IDW method (Inverse Distance Weighting).

Table 2. Parameters fitted to the semivariograms of plant height and first pod insertion height (FPIH) for Fields 1 (T1) and 2 (T2).

Variable	Model	C0	C0+C	Range (m)	RMSE	R^2	CV - SC
Height – T1	SP	45.94	107.30	383.10	896.30	0.68	1.21
Height – T2	SP	34.51	115.19	394.07	85.71	0.96	1.08
FPIH – T2	SP	0.61	1.29	323.18	0.02	0.81	0.74

SP: spherical model; C0: nugget effect; C0+C: sill; RMSE – residual; CV: cross-validation; SC: slope coefficient of the line.

In fields T1 and T2 (Figure 4), it can be observed that the tallest plants are located in the center of the field, whereas the shortest plants are concentrated in the headland area. This pattern may be associated with the fact that the central regions are less trampled areas, favoring plant development when compared with plants along the border. This result is in agreement with Moraes et al. (2020), who highlighted that the interaction between the environment and the physical, chemical, and biological soil conditions alter plant growth and development.

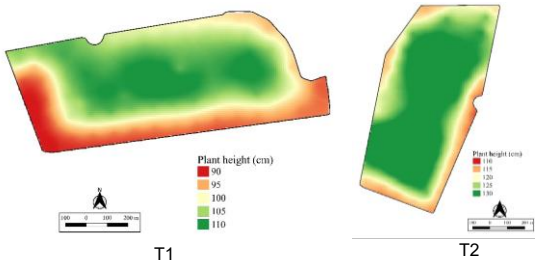

Figure 4. Spatial distribution of plant height (cm) in T1 and T2.

For FPIH, it can be observed that in T1 (Figure 5), there is a predominance between 11 and 12 cm in height, with some patches in which plant height varied between 10 and 13 cm. In field T2, FPIH predominated between 10 and 12 cm, with some patches of plants lower than 10 cm and one region with plants between 12 and 13 cm of FPIH. This can be explained by the illumination period to which the plants in each field were exposed during their emergence and vegetative periods. Fields with greater intensity influence plants to enter the reproductive stage more rapidly, resulting in lower FPIH.

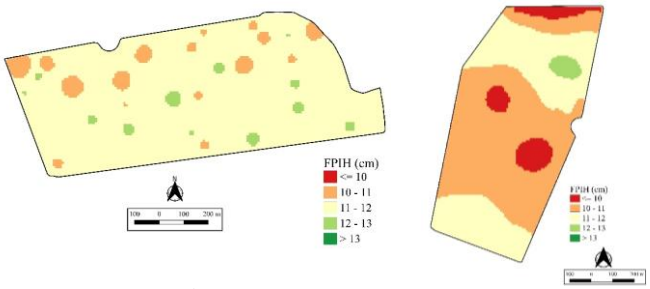

Figure 5. First pod insertion height (FPIH - cm) in T1 (IDW) and T2 (Kriging).

The spatial distribution of yield in fields T1 and T2 (Figure 6) showed higher grain yield in T2. In T1, yield was negatively affected by excessive rainfall during harvest, making it necessary to delay the harvesting operation, which affected the yield of the field.

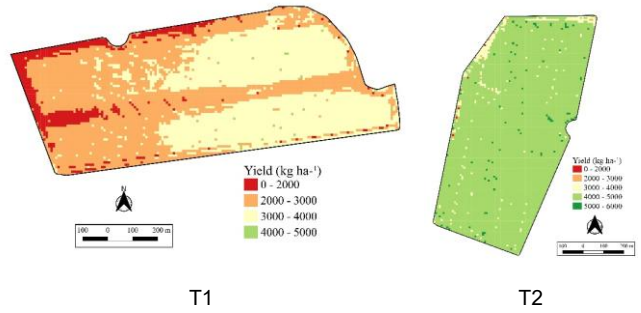

Figure 6. Spatial distribution of yield (kg ha-¹) in T1 and T2.

Another aspect to observe is that plant height (Figure 4) was lower in T1 compared to T2, coinciding with the area of lower yield observed in T1. Population density is a factor that may influence plant height. The data from this study show that field T2 had a higher final population than T1, which may have favored greater height. This result is in agreement with Dörr et al. (2023), who found that increasing plant density favors taller plants. However, Souza et al. (2010) state that greater population density reduced the number of pods per plant and grains per pod, but total grain yield was not altered. Wei and Molin (2020) point out that plant height can serve as an indirect indicator of the productive potential of the crop, and when associated with other indicators, it can help identify the efficiency of vegetative growth and its relationship with yield. Corassa et al. (2019) demonstrated that taller varieties should preferably be cultivated in low-potential areas, whereas shorter plants are more suitable in high-potential areas, because if they grow too much, it may result in excessive shading and reduce yield. However, if the actual plant density is very high as close to the value obtained in T2 (13.41 plants.m^{-1}), even shorter hybrids may show reduced yield due to mutual shading and greater intraspecific competition. Thus, based on the results of this study, it is suggested that sowing quality influenced the dynamics of competition among plants in the fields. This variation in spatial arrangement, together with the stature of the varieties, may have affected light interception and, consequently, yield. In the present study, T2 better converted the planned density into a functional canopy, whereas T1 suggests losses in stand establishment quality that reduced the use of productive potential. In T2, Brachiaria brizantha cv. Marandu was present as vegetation cover, which may have contributed to better crop establishment and yield.

3.3 Harvest losses

The analysis of pre-harvest losses (Table 3), according to field observations, occurred mainly due to pod dehiscence caused by delayed harvest and grain germination while still in the pods.

Table 3. Descriptive statistics of harvest losses (kg ha^{-1}).

	Pre-harvest losses		Total losses		Harvester losses	
Parameter	Field 01	Field 02	Field 01	Field 02	Field 01	Field 02
Mean	160.40	131.65	606.50	481.46	446.10	349.80
SD[1]	118.90	89.59	347.70	79.48	312.70	109.70
CV (%)[2]	74.15	68.05	57.34	16.51	70.1	31.36
Minimum	0	0	110	265	65	80
Median	117.5	130	547.5	487.5	380	335
Maximum	455	425	2685	740	2290	675
Sk[3]	1.01	0.74	2.84	0.04	2.74	0.22
K[4]	-0.04	1.14	14.83	1.89	13.88	0.62
RJ[5]	<0.01*	0.02*	<0.01*	<0.01*	<0.01*	>0.10^{ns}

[1] SD: standard deviation; [2] CV (%): coefficient of variation; [3] Sk: Skewness; [4] K: kurtosis; [5] RJ: Ryan–Joiner test, where (*) indicates significance at $p < 0.05$ and (ns) indicates a non-significant distribution. When significant, it indicates that the hypothesis of normal distribution is rejected.

The CV was medium only for T2 in total losses; high only for T2 in harvester losses; and very high for the other losses (Table 3), according to Lopes et al. (2021). Jasper et al. (2021) also found high CV values (64.52%) for harvest losses, justified by the influence of uncontrollable factors, such as climatic variations and soil conditions. The analysis of skewness (Sk) and kurtosis (K) data indicates that most losses do not follow a normal frequency distribution, which was confirmed by the Ryan–Joiner test (Table 3). The exception occurs for harvester losses in T2, which showed a normal distribution. The non-normality of the loss data was confirmed by the skewness (Sk) and kurtosis (K) coefficients being distant from zero, which suggests, for future studies, the use of non-parametric statistical methods for these variables. The analysis of total losses (Table 3) indicates high values, due to climatic adversities, lack of monitoring during sowing and harvest, and probable lack of machinery adjustment. Jasper et al. (2021) found a 5.29% increase in losses when harvester speed increased by 1 km h^{-1}, and suggested that this increase was due to lodged and unharvested plants. In T1, one of the factors that may have caused the increase in losses was excessive rainfall during the harvest stage, which had to be delayed, generating increased losses due to pod dehiscence and grain deterioration. Harvester losses (Table 3) were determined by the difference between total loss and pre-harvest loss. These values are directly related to poor harvester

adjustment, influenced by reel speed, cutting height of the harvester platform, threshing cylinder rotation, clearance between cylinder and concave, and machine travel speed. In addition to these factors, harvest flow affects feeding of the threshing system, caused by uneven crops. Carreira et al. (2024) state that attention should be given to measures for reducing harvest losses, such as training operators and configuring the harvester to reduce losses. Regarding the parameters fitted to the semivariograms of losses (Table 4), the spherical model showed the best fit. A moderate fit was observed for pre-harvest losses, and a weak fit for total loss in field T2. The highest value of the range was observed for pre-harvest losses in T2, which represents greater spatial dependence of these losses. The range represents the maximum distance for two points to have spatial dependence; that is, points within an equal or smaller radius are correlated with each other (Lopes et al., 2020).

Table 4. Parameters fitted to the semivariograms of pre-harvest and total losses.

Variable	Model	C0	C0+C	Range (m)	RMSE	R^2	CV - SC
PHL – T1	SP	9411.20	14634.79	327.87	1589315	0.80	0.73
PHL – T2	SP	6568.64	8836.62	969.95	2662023	0.61	0.53
TL – T2	SP	5973.55	6452.94	305.29	31262.28	0.53	0.67

PHL: pre-harvest losses; SP: spherical model; C0: nugget effect; C0+C: sill; RMSE – residual; CV: cross-validation; SC: slope coefficient of the line.

The spatial evaluation of pre-harvest grain losses indicated that, in most of the area, values were 120 kg ha^{-1} for both fields, with a loss of 4.22% for T1 and 2.73% for T2. However, some areas in T1 showed losses from 240 to 300 kg ha^{-1} (Figure 7), representing losses from 8.43% to 10.54%. The delay in harvest determined the increase in losses due to rainfall that occurred after the R8 stage, in addition to the genetic characteristics of the cultivar and pod dehiscence. In T1, grain germination still inside the pod also occurred at some points. It was observed that the region of lower yield in T1 overlapped the areas with the highest pre-harvest losses, indicating a relationship between yield and losses. Zuffo et al. (2020) found that harvesting at advanced stages promotes losses in seed quality, mainly when heavy rainfall occurs. The authors reported that rainfall of 50 mm causes deterioration of soybean seeds, a phenomenon that was also recorded in this study.

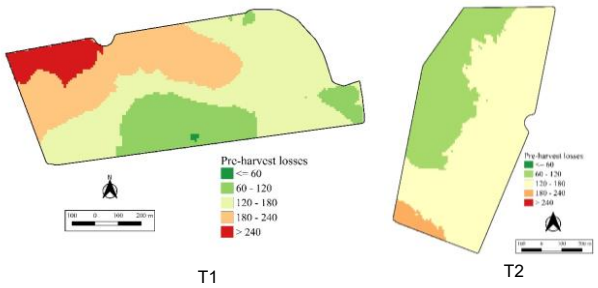

Figure 7. Spatial distribution of pre-harvest losses (kg ha^{-1}) in T1 and T2.

Total losses (Figure 8) ranged from 300 to 600 kg ha^{-1} in T2 (6.81% to 13.63%) and were even higher in T1 over a large part of the area. Grain losses generate substantial economic losses for farmers. However, many do not pay attention to this highly important factor, despite its direct impact on yield reduction.

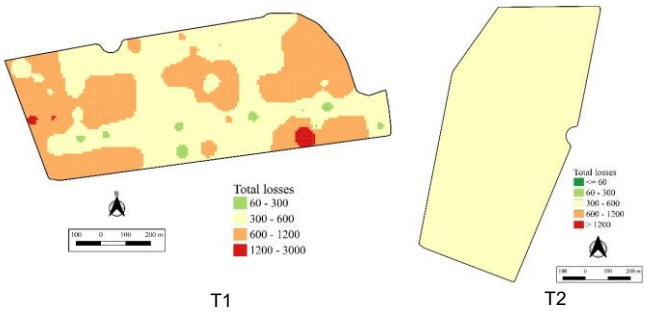

Figure 8. Spatial distribution of total losses (kg ha^{-1}) in T1 (IDW) and T2 (Kriging).

Machine grain losses (Figure 9) result in high losses for farmers, as they cause reductions in yield. In most of the area, losses ranged between 300 and 600 kg ha^{-1} (6.81% to 13.63%); however, in field T1, some regions showed losses on the order of 600 to 1200 kg ha^{-1} (13.63% to 27.26%), which can be considered areas of extreme loss for the producer.

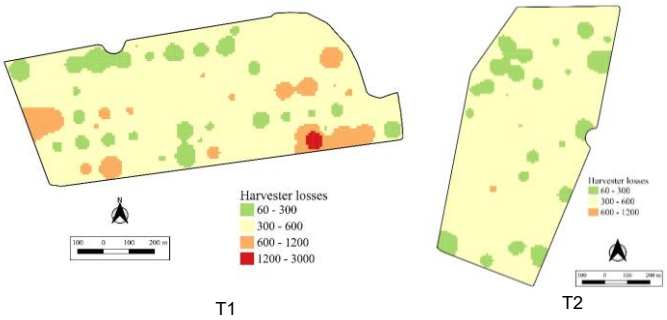

Figure 9. Spatial distribution of harvester losses (kg ha^{-1}) in T1 and T2, by IDW.

3.4 Pearson correlation

The results of Pearson's correlation analysis (Table 5) indicate that, in T1, there is a positive correlation between yield and NDVI. Spectral indices, including NDVI, explain an important part of yield variation in soybean, showing good predictive capacity of the spectrum/NDVI for yield. Shammi and Meng (2021) observed that models based on NDVI and EVI showed better performance in predicting soybean yield at 70 and 85 days after sowing. Similarly, Crusciol et al. (2021), evaluating three consecutive growing seasons, identified that the highest accuracy in yield estimation occurred at 89, 96, and 94 days after sowing, corresponding to the R5 phenological stage. These results reinforce that leaf reflectance constitutes a tool for predicting soybean grain yield. In field T2, no significant correlation was observed between NDVI and yield (Table 5), nor among the other attributes evaluated. In addition, NDVI, plant height, and first pod insertion height did not show a significant relationship with grain losses. This absence of correlation may be related to NDVI saturation in more dense canopies, which reduces its sensitivity to detect biomass variations. According to Xu and Katchova (2019), the response of soybean yield to NDVI throughout the cycle is nonlinear, which helps explain the limitation of the index at certain stages of crop development. Therefore, the correlation pattern in T1 and T2 is consistent with evidence that NDVI may correlate with yield when there is contrast in vigor/stress and when measured at appropriate stages but tends to saturate in dense canopies.

Table 5. Pearson correlation for the variables analyzed in fields T1 and T2.

T1	Yield	NDVI	Height	FPIH	PHL	TL
NDVI	0,478					
Height	0,154	0,106				
FPIH	0,156	0,247	-0,004			
PHL	-0,390	-0,350	0,264	-0,286		
TL	-0,109	0,050	-0,175	-0,000	0,263	
HL	0,007^{ns}	0,152	-0,286	0,118	-0,056	0,926
T2	**Yield**	**NDVI**	**Height**	**FPIH**	**PHL**	**TL**
NDVI	0,133					
Height	0,138	0,251				
FPIH	0,000^{ns}	0,356	-0,005^{ns}			
PHL	0,108	-0,200	-0,116	0,282		
TL	0,175	-0,175	0,274	-0,169	0,505	
HL	-0,023	-0,100	0,209	-0,277	-0,321	0,247

^{ns}: not significant at 5% probability. Yield: yield; NDVI: Normalized Difference Vegetation Index; FPIH: first pod insertion height; PHL: pre-harvest losses; TL: total losses; HL: harvester losses.

Harvester losses (HL) and total losses showed a very strong correlation in T1 (Table 5); that is, they are directly correlated variables, and it was therefore verified that the increase in MACHL is directly associated with the increase in total losses. In field T2, there was a significant correlation between the variables pre-harvest losses (PHL) and total losses (TL). This suggests that pre-harvest losses directly influenced total losses. Pre-harvest loss increases TL when harvest is delayed, especially in cultivars more susceptible to dehiscence and/or weather conditions, because it composes the total and may predominate in certain scenarios. Pre-harvest losses may dominate total losses when harvest is delayed or when adverse weather conditions occur. For this study, it is highlighted that pre-harvest losses were above the acceptable limit, affecting the values of total losses.

4. Conclusions

The soybean crop showed spatial variability in the attributes evaluated, including NDVI, plant height, first pod insertion height (FPIH), and yield. A positive correlation occurred between NDVI and yield in one of the areas, highlighting the potential of this index as an indicator of the vegetative and productive vigor of the crop. Plant height and first pod insertion height showed spatial variability, but with low correlations with yield and grain losses. Mean yield was 2846.2 kg ha^{-1} in field 1 and 4402.5 kg ha^{-1} in field 2, indicating marked differences between the fields due to factors such as delayed harvest and rainfall occurrence during the maturation period. Harvest losses were above the acceptable limit of 60 kg ha^{-1}. Total losses ranged from 300 to 600 kg ha^{-1} (6.81–13.63% of production), with regions reaching 600–1200 kg ha^{-1} (13.63–27.26%).

Acknowledgements

The authors acknowledge CNPq for granting the Master's scholarship, and the team of Fazenda União, belonging to the Fabiani Agropecuária Ltda. group, for the opportunity and support provided during the development of this study.

References

Batistella, D., Modolo, A. J., Campos, J. R. da R., & Lima, V. A. de. (2023). Comparative analysis of orbital sensors in soybean yield estimation by the random forest algorithm. Ciência e Agrotecnologia, 47:e002423. https://doi.org/10.1590/1413-7054202347002423

Carreira, V. dos S., Aleixo, E. V., Ribeiro, N. M., Santos, J. do N., & Silva, R. P. da. (2024). A systematic and meta-analytical review of soybean mechanized harvesting in South America. Revista Brasileira De Engenharia Agrícola E Ambiental, 28(1):e265804. https://doi.org/10.1590/1807-1929/agriambi.v28n1e265804

Chedid, V., Cortez, J. W., & Arcoverde, S. N. S. (2024). Monitoring the vegetative state of coffee using vegetation indices. Engenharia Agrícola 44:e20220212. https://doi.org/10.1590/1809-4430-Eng.Agric.v44e20220212/2024

Comparin, P.J.S., & Cortez, J.W. (2023). Soil sample densities combined with additional points in the variability of soil chemical attributes. Rev. Gest. Soc. Ambient. 17(8):e03218. https://doi.org/10.24857/rgsa.v17n8-029

Corassa, G. M., Santi, A. L., Amado, T. J. C., Reimche, G. B., Gaviraghi, R., Bisognin, M. B., & Pires, J. L. F. (2019). Performance of soybean varieties differs according to yield class: A case study from Southern Brazil. Precision Agriculture, 20:520–540. https://doi.org/10.1007/s11119-018-9595-0

Crusiol, L.G.T., Nanni, M.R., Furlanetto, R.H., Sibaldelli, R.N.R., Cezar, E., Sun, L., Foloni, J.S.S., Mertz-Henning, L.M., Nepomuceno, A.L., Neumaier, N., & Farias, JRB (2021). Yield Prediction in Soybean Crop Grown under Different Levels of Water Availability Using Reflectance Spectroscopy and Partial Least Squares Regression. Remote Sens, 13 (5):977. https://doi.org/10.3390/rs13050977

Filla, V. A., Coelho, A. P., Bettiol, J. V. T., Leal, F. T., Lemos, L. B., & Rosalen, D. L.. (2023). Model performance in estimating the yield of common bean cultivars. Revista Ciência Agronômica, 54:e20217835. https://doi.org/10.5935/1806-6690.20230002

Hernández-López, D., Piedelobo, L., Moreno, M. A., Chakhar, A., Ortega-Terol, D., & González-Aguilera, D. (2021). Design of a local nested grid for the optimal combined use of Landsat 8 and Sentinel 2 data. Remote Sensing, 13(8):1546. https://doi.org/10.3390/rs13081546

Inacio, K. A. M., & Cortez, J. W. (2023). Variabilidade espacial da produtividade da soja e sua correlação com atributos químicos e textura do solo. Rev. Gest. Soc. Ambient. 17(2):e03226. https://doi.org/10.24857/rgsa.v17n2-030

Jasper, SP, Zimmermann, GG, Savi, D., Strapasson Neto, L., Kmiecik, LL, & Sobenko, LR. (2021). Desempenho operacional e eficiência energética de colhedoras axiais com sistemas de rotor simples e duplo na colheita de sementes de soja. Ciência E Agrotecnologia, 45:e031720. https://doi.org/10.1590/1413-7054202145031720

Kasimati, A., Psiroukis, V., Darra, N., Kalogrias, A., Kalivas, D., Taylor, J.A., & Fountas, S. (2023). Investigation of the similarities between NDVI maps from different proximal and remote sensing platforms in explaining vineyard variability. Precision Agriculture 24:1220–1240. https://doi.org/10.1007/s11119-022-09984-2

Lopes, B.G., Faria, GA, Maltoni, K.L., Rocha, P.S., Peixoto, A.P.B., Oliveira, T.A. de., Fonseca, A.D. da., & Felizardo, L.M. (2021). Classification of the coefficient of variation for experiments with eucalyptus seedlings in greenhouse. Revista Ciência Agronômica, 52(4):e20207587. https://doi.org/10.5935/1806-6690.20210050

Lopes, K. A. L., Pinto Junior, F. F, Aguiar, F. I. S., Sousa, A. E. S.; Oliveira, I. R., & Dantas J. S. (2020). Diferentes densidades amostrais na caracterização da variabilidade espacial de atributos granulométricos de um Argissolo amarelo distrocoeso típico. Cultura Agronômica 29(1):50-60. https://doi.org/10.32929/2446-8355.2019v29n1p50-60

Lopes, A. da S., Andrade Júnior, A. S. de., Bastos, E. A., Sousa, C. A. F. de., Casari, R. A. das C. N., & Moura, M. S. B. de. (2024). Assessment of maize hybrid water status using aerial images from an unmanned aerial vehicle. Revista Caatinga, 37:e11701. https://doi.org/10.1590/1983-21252024v3711701rc

Dörr, C. S., Pinz, E. R., Bratz, I. S., Spinelli, V. M., Martins, A. B. N., Drews, G. K., Magano, D. A., & Panozzo, L. E. (2023). Densidade de plantas e desempenho produtivo da soja. Revista De Gestão e Secretariado, 14(10):18821–18831. https://doi.org/10.7769/gesec.v14i10.3083

Mesquita, C.M.; Costa, N.P.; Mantonvani, E.C.; Andrade, J.C.M. De A.; França-Neto, J.B.; Silva, J.G. De; Fonseca, J.R.; Portugal, F.A.F.; Guimarães Sobrinho, J.B. (1998). Manual do produtor: como evitar desperdício nas colheitas de soja, do milho e do arroz. Londrina: EMBRAPA-CNPSo. 31 p. (EMBRAPA-CNPSo, Documentos, 112).

Moraes, M. T. Debiasi, H., Franchini, J. C., Mastroberti, A. A., Levien, R., Leitner, D., & Schnepf, A. (2020). Soil compaction impacts soybean root growth in an Oxisol from subtropical Brazil. Soil & Tillage Research 200(104611). https://doi.org/10.1016/j.still.2020.104611

Molin, J. P.; Amaral, L. R. Do; Colaço, A. (2015). Agricultura de precisão. São Paulo: Oficina de Textos. 224 p.

Pimentel Gomes, F. (2000) Curso de estatística experimental. 14. ed. Piracicaba: Degaspari. 477 p.

Pereira, G. W., Valente, D. S. M., Queiroz, D. M. D., Coelho, A. L. D. F.; Costa, M. M.; Grift, T. (2022). Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. Agronomy 12(6):1350. https://doi.org/10.3390/agronomy12061350

QGIS (2024). QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Ramadhani, F., Pullanagari, R., Kereszturi, G., & Procter, J. (2021). Mapping a cloud-free rice growth stages using the integration of PROBA-V and Sentinel-1 and its temporal correlation with sub-district statistics. Remote Sensing 13(8):e1498. https://doi.org/10.3390/rs13081498

Ratke, RF, Aguilera, JG, Zuffo, AM, Baio, FHR, Theodore, PE, Yokota, LA, Viana, PRN, Ratke, LPT, & Oliveira, JT de.. (2024). Spatial dependence of soybean cultivation, in a lowcarbon production system, integrated with eucalyptus forest. Ciência Florestal, 34(3)e73889. https://doi.org/10.5902/1980509873889

Ré, N. C., Silva, C. A. A. C., Oliveira, A. K. da S., Caron, M. L., Nilsson, M. S., Duft, D. G., & Fiorio, P. R.. (2025). VIS-NIR-SWIR spectroscopy in sugarcane (Saccharum officinarum L.) cultivation for phytosanitary purposes. Ciência Rural, 55(3):e20230484. https://doi.org/10.1590/0103-8478cr20230484

Shammi, S. A, & Meng, Q. (2021). Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. Ecological Indicators, 121:107124 https://doi.org/10.1016/j.ecolind.2020.107124

Silva, M. H. Da, Elias, A. R., & Rosário, L. L. do. (2022). Análise da cultura da soja a partir de índices de vegetação (ExG - GLI - TGI - VEG) advindos de imagens RGB obtidas com ARP. Revista Brasileira de Geomática 10(2):140-154. https://doi.org/10.3895/rbgeo.v10n2.15042

Simas, G.H.N., Fiedler, S., & Cortez, J.W. (2023). Uso de imagens aéreas com drones para identificação de falhas no estabelecimento da soja. In: Zuffo, A. M., & Aguilera, J.G. Inovações em pesquisas agrárias e ambientais - Volume I (pp.6-15). Nova Xavantina-MT: Pantanal. https://doi.org/10.46420/9786585756143cap1

Souza, C. A., Gava, F., Casa, R. T., Bolzan, J. M., & Kuhnem Junior, P. R. (2010). Relação entre densidade de plantas e genótipos de soja roundup readyTM. Planta Daninha, 28(4):887–896. https://doi.org/10.1590/S0100-83582010000400022

Wei, M.C.F & Molin, J. P. (2019). Soybean Yield Estimation and Its Components: A Linear Regression Approach. Agriculture 10(8):348. https://doi.org/doi:10.3390/agriculture10080348

Xu, C., & Katchova, A. L. (2019). Predicting Soybean Yield with NDVI Using a Flexible Fourier Transform Model. Journal of Agricultural and Applied Economics, 51(3):402–416: https://doi.org/10.1017/aae.2019.5

Zuffo, A. M., Aguilera, J. G., Carvalho, E. R., & Teodoro, P. E. (2020). Harvest times with chemical desiccation and the effects on the enzymatic expression and physiological quality of soybean seeds. Revista Caatinga 33(2):361–370. https://doi.org/10.1590/1983-21252020v33n209rc