Estimating peanut invisible losses using machine learning and remote sensing

Armando L. de Brito Filho[1], Franciele M. Carneiro[2], Samira L. H. de Almeida[1], Francielle M. Ferreira[3], Luciano S. Shiratsuchi[4], Rouverson P. da Silva[1]

[1]Department of Engineering and Mathematical Sciences, São Paulo State University (Unesp) - 14884-900 - Jaboticabal – SP - Brazil

[2] Federal Technological University of Paraná (UTFPR) - 85892-000 - Santa Helena - PR – Brazil

[3] State University of Mato Grosso (UNEMAT) - 78450-000 - Nova Mutum – MT - Brazil

[4] School of Plant, Enviromental and Soil Sciences, Louisiana State University (LSU) - 70808 – Baton Rouge – LA - USA

armando.brito@unesp.br, fmcarneiro@utfpr.edu.br, samira.lh.almeida@unesp.br, francielle@unemat.br, lshiratsuchi@agcenter.lsu.edu, rouverson.silva@unesp.br

Abstract. *Monitoring losses of underground crops still needs to be studied, mainly for estimating losses using new techniques that facilitate the farmer's decision-making. This study aimed to estimate peanut invisible losses using remote sensing and machine learning. The experimental area was conducted in a peanut agricultural region near Taquaritinga in São Paulo, Brazil. Seventy sample points were collected, encompassing various aspects of peanut losses, including visible and invisible losses, yield, maturation, and orbital remote sensing data. The statistical analysis used were Principal Components Analysis and Random Forest. The study concluded that the best results were RGB, RE + NIR, NIR + R, and RE to estimate peanut invisible losses.*

Keywords: Arachis hypogaea, orbital satellite images, artificial intelligence

1. Introduction

Crop losses significantly impact agricultural productivity. Among various agricultural systems, mechanized harvesting is one with substantial loss potential. It necessitates meticulous attention and proper execution to mitigate these losses effectively. Inadequate harvesting practices, often attributed to insufficient workforce training, improper machine adjustments, and factors like excessive water content in grains or pods, can drastically reduce productivity and a notable increase in operational expenses.

Addressing these issues is crucial to optimize crop yields and minimize financial burdens.

Among various crops, special attention must be given to underground crops, such as peanuts and potatoes, particularly during the harvest phase. Handling these crops poses additional challenges as they are more difficult to monitor for visible and invisible losses. Detecting losses in underground crops requires extensive labor due to the necessity of digging the soil to identify potential issues. Conventional methods for estimating losses in such crops demand considerable labor and become costlier as the area to be monitored expands.

In this context, geotechnologies, particularly remote sensing (RS) utilizing orbital images, present a promising solution for estimating invisible losses. By employing RS techniques, the monitoring process can become more efficient, cost-effective, and capable of covering larger areas, making it an invaluable tool for mitigating losses and optimizing underground crop yields.

SR is a method that obtains information without physical contact with the target [Florenzano, 2011; Jensen, 2009] and is non-destructive. Among the technologies used in agriculture, SR stands out for allowing the monitoring of temporal-spatial variability of small, medium, and large agricultural areas; estimation of crop yield [Shiratsuchi et al., 2014] and biomass; in this work, we will estimate the invisible losses of peanuts using artificial intelligence (AI), with machine learning (ML) algorithms. ML techniques are excellent statistical tools for measuring model quality and error, especially in SR, when the data may not have a linear distribution.

Given the above, we have as a novely hypothesis for this work the following question: "Is it possible to estimate invisible losses of underground cultivation through SR and AI?" We saw the benefits of using SR in agriculture, and if the hypothesis of this work is answered positively, the rural producer will reduce his operations. Thus, the objective of this study was to estimate peanut invisible losses using remote sensing and machine learning.

2. Material and Methods

The experimental area was conducted in the largest producing region in Brazil, in which São Paulo State is responsible for producing around 90% of peanuts. This experimental area was conducted in a peanut agricultural region near Taquaritinga city in São Paulo, Brazil (as depicted in Figure 1). The region is characterized by sandy soil texture, and its climate falls under the classification of Aw, indicating a tropical climate with a dry winter, following the Köppen climate classification system [Alvares et al., 2013].

Seventy sample points were collected, encompassing various aspects of peanut losses, including visible and invisible losses, yield, and maturation. Additionally, free orbital remote sensing data from the Sentinel 2 (S2) satellite imagery were acquired from the Copernicus Open Access Hub website (https://scihub.copernicus.eu/dhus/#/home).

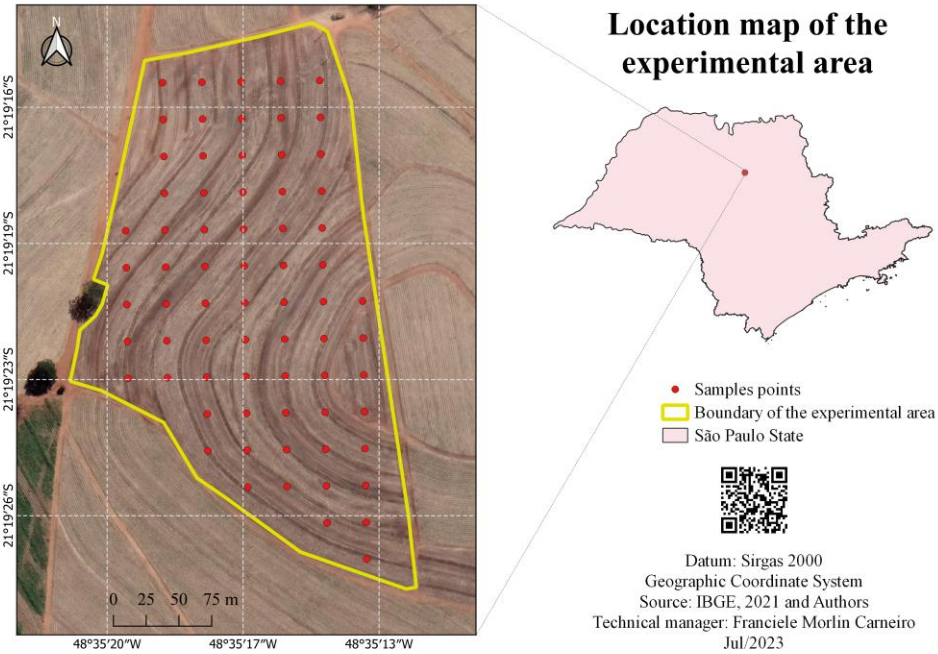


Figure 1. Location map of the experimental area in a peanut agricultural

Data processing was meticulously organized into three steps: pre-processing, processing, and data analysis.

1. Pre-processing: The data collected were categorized into two main groups: input data, which included Sentinel 2 (S2) satellite imagery, visible losses, yield, and maturation data, and output data, focusing on invisible losses. The remote sensing data were gathered on six specific dates, as outlined in Table 1. Detailed information regarding sowing and harvesting data and the processing of orbital imagery data was also recorded. The spectral bands (Table 2) evaluate were: blue (B), green (G), red (R), redegde (RE), near-infrared (NIR), and combinations between them (RE + NIR, NIR + R)

2. Processing: The software QGIS®, version 3.22.10, was employed for data processing. To extract the sample points obtained across all data categories within the designated area, the "Points Sampling Tool" plugin was utilized. Subsequently, the collected data were saved in CVS format and further organized using the Excel program.

3. Data analysis: For the data analysis phase, Google Collaboratory, known as Colab (https://colab.research.google.com/), was employed using Python. This platform facilitated the implementation of the Random Forest Regression (RFR) and Principal Components Analysis (PCA) techniques to process the data effectively. In addition, the dataset was trained (70%) and validated (30%) by RFR.

Table 1. Data information about sowing and harvesting date and also orbital satellite imagery date

Data collection date information	Date	Number of imagery date
Sowing date	October 25th, 2022	-----
S2 imagery	October 26th, 2022	1
S2 imagery	November 10th, 2022	2
S2 imagery	November 20th, 2022	3
S2 imagery	November 30th, 2022	4
Harvest date	March 24th, 2023	----
S2 imagery	March 25th, 2023	5
S2 imagery	March 30th, 2023	6

S2: Sentinel 2 satellites imageries

Table 2. Orbital remote sensing used was Sentinel-2 spectral bands

Sentinel spectral band	**Abbreviation**	**Central wavelength (nm)**	**Spatial resolution (m)**	**Abbreviation**
Blue	B	490	10	B
Green	G	560	10	G
Red	R	665	10	R
Red Edge	RE	740	20	RE
Near-infrared	NIR	842	10	NIR

The variable collection as pod maturation was determined using the Hull scrape method (Williams and Drexler., 1981), where eight plants were collected, containing approximately 200 pods per point. Yield and losses were collected in the same frame of 2 m^2. The yield was corrected to 8% of the water content in a pod.

The metric parametric of accuracy and precision used were Determination coefficient (R2), Root Mean Square Error (RSME), and Mean Absolute Percentage Error (MAPE), the equations 1 - 3 for each metric parameter below.

$$\mathrm{RSME}=\sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{2}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \times 100 \tag{3}$$

On what:

y_i: estimated values

$\hat{y}_i$: observed values

N: number of samples

$\bar{y}_i$: mean of the observed values

3. Results and Discussion

The most effective spectral bands for estimating peanut invisible losses during crop growth were found to be RGB and RE, as depicted in Figures 2a, 2b, and 2c. Subsequently, the most accurate results were obtained one day after the harvesting date using the NIR + R and RE + NIR spectral bands, as shown in Figure 2e. Figure 2d shows us that almost one month after sowing, plant reflectance was affected by soil reflectance effects due to smaller peanut plants and little plant coverage on the ground, which was observed by Povh et al. [2008] and Carneiro et al. [2020]. Moreover, six days after the harvest, the spectral bands that yielded the best results were RGB, NIR + R, and NIR + RE, as demonstrated in Figure 2f.

We conducted an intriguing study concerning peanut crop growth, wherein we utilized various spectral bands such as RGB, NIR + R, and NIR + RE to assess potential imperceptible losses at an early stage. This early estimation of peanut losses proved advantageous for farmers, as it directly pertains to crop yield and facilitates more accurate decision-making to address yield-affecting factors, as evident from the observed losses.

Moreover, we found that employing RGB, NIR, and RE + NIR spectral bands yielded superior results for monitoring crop variability and estimating overall output. The rationale behind these outcomes lies in the close association between remote sensing and plant physiology. Specifically, the RGB bands are absorbed by chloroplasts, which play a crucial role in the photosynthetic process, particularly in the case of the red and blue components. Additionally, NIR and RE exhibit greater reflectance compared to the red and blue components, further contributing to the efficacy of these spectral bands in our study.

(a) (b)

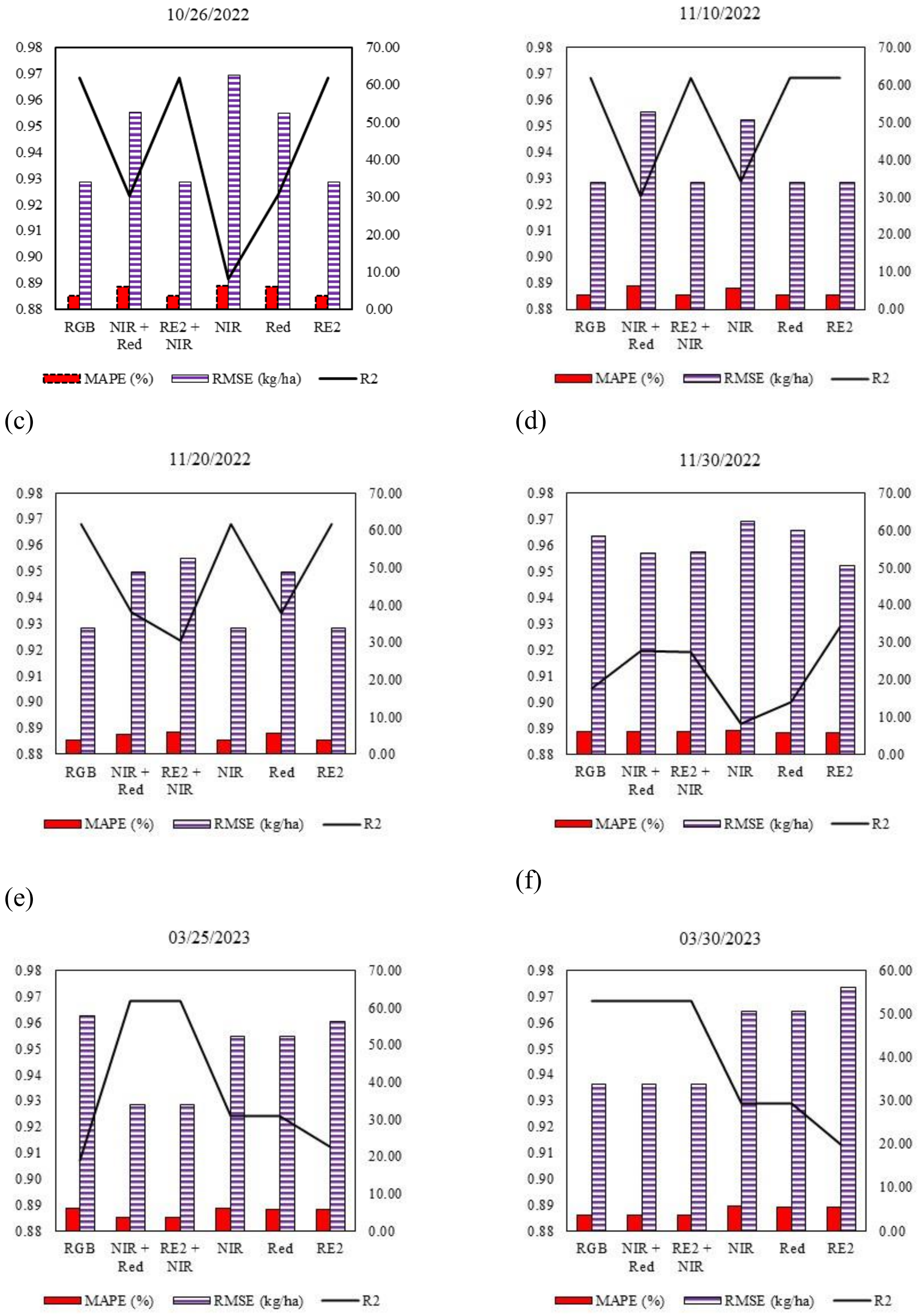
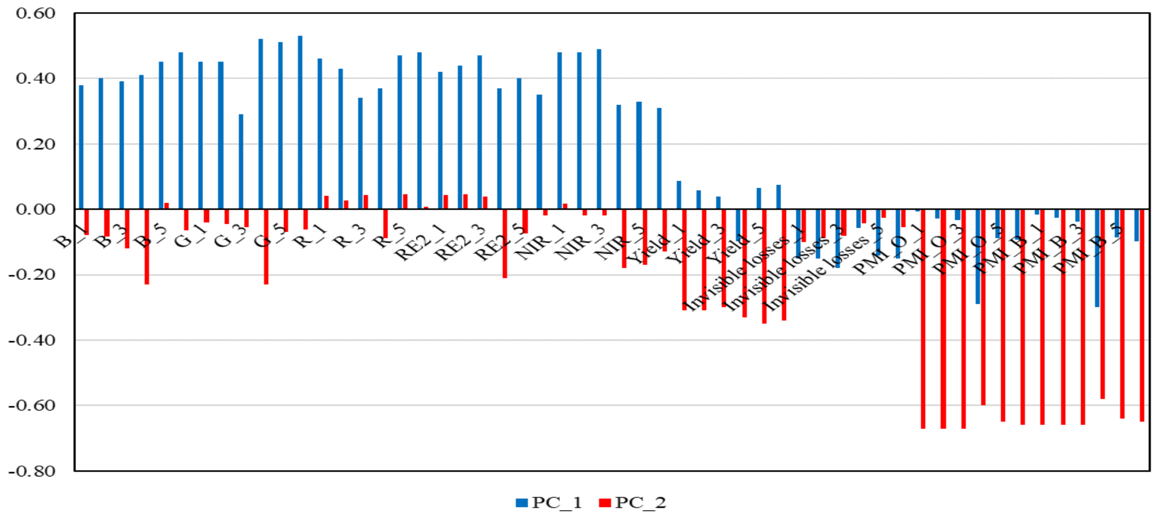


Figure 2. Metric parameters used for estimation of peanut invisible losses

Carneiro et al. (2023), using Random Forest, orbital remote sensing, and soil data, got more excellent results using this machine learning algorithm, demonstrating a good statistical analysis.

Principal components analyses were realized (Figure 3); the best results were G (dates 1, 2, 4, 5, and 6), B (dates 2, 4, and 6), R (dates 1, 2, 5, and 6), RE (dates 1, 2, 3, and 5), and NIR (dates 1, 2, and 3) spectral bands. PCA analysis is an excellent tool for selecting the best inputs to determine the output that you would like to know how in our case, to estimate the peanut invisible losses. These results corroborate those found in the literature, in which the wavelengths most used in agriculture are in the visible, as RGB, and NIR ranges [Molin et al., 2015]. Moreover, we can see that the peanut biophysical characteristics (maturation, yield, and invisible losses) of the cultures obtained high explanatory values above 60% but were inversely proportional to the remote sensing data.



PC_1: Principal Components 1, PC_1: Principal Components 2, _1: data 1, _2: data 2, _3: data 3, _4: data 4, _5: data 5, _6: data 6, B: blue, R: red, G: green, RE: rededge, NIR infrared, PMI_O: Peanut Maturity Index orange classification, PMI_B: Peanut Maturity Index black classification

Figure 3. Principal Components Analyses to estimate peanut invisible losses

4. Conclusions

The study concluded that the best results were RGB, RE + NIR, NIR + R, and RE to estimate peanut invisible losses. Future works, we used more fields to validate a model created to estimate peanut invisible losses for different scenarios. The earlier estimating the losses, the better to avoid increasing costs and ensure crop yield is not harmed. In this way, we saw in this work that we obtained good results right at the beginning of the development of the culture.

5. References

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. de M., and Sparovek, G. (2013). "Köppen's climate classification map for Brazil". *Meteorologische Zeitschrift*, 22(6), pages 711-728. doi: 10.1127/0941-2948/2013/0507

Carneiro, F. M., Furlani, C. E. A., Zerbato, C., Menezes, P. C., Gírio, L. A. S., Oliveira, M. F. (2020). Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. *Precision Agriculture*, 21, 979-1007. doi: 10.1007/s11119-019-09704-3

Carneiro, F. M., de Brito Filho, A. L., Ferreira, F. M., Junior, G. D. F. S., Brandao, Z. N., da Silva, R. P., & Shiratsuchi, L. S. (2023). Soil and satellite remote sensing variables importance using machine learning to predict cotton yield. *Smart Agricultural Technology*, 5, 100292. doi: 10.1016/j.atech.2023.100292

Florenzano, T. G. (2011) *"Iniciação em Sensoriamento Remoto*". 3 ed. ampl. e atual. São Paulo: Oficina de Textos, 128p.

Jensen, J. R. (2009) "*Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres"*. São José dos Campos: Parêntese, 598p.

Molin, J. P., Amaral, L. R., and Colaço, A. F. (2015) *"Agricultura de precisão"*. Oficina de Textos: São Paulo, p.119-153.

Povh, F. P., Molin, J. P., Gimenez, L. M., Pauletti, V., Molin, R., Salvi, J. V. (2008) Comportamento do NDVI obtido por sensor ótico ativo em cereais. *Pesquisa agropecuária brasileira* 43, pages 1075-1083.

Shiratsuchi, L. S., Brandão, Z. N., Vicente, L. E., Victoria, D. C., Ducati, J. R., Oliveira, R. P., and Vilela, M. F. (2014) "Sensoriamento Remoto: conceitos básicos e aplicações na Agricultura de Precisão". In: Bernardi ACC, Naime JM, Resende AV, Bassoi LH, Inamasu RY (2014) *Agricultura de precisão*: resultados de um novo olhar. Brasília: EMBRAPA, 2014. Cap. 4. p.58-73.

Williams, E.J., and Drexler, J.S. (1981). "A non-destructive method for determining peanut pod maturity". *Peanut Science,* 8, pages 134-141.