UAV imaging for spectral characterization of Coffee Leaf Miner (*Leucoptera coffeella*) infestation in the Cerrado Mineiro region

Vinicius Silva Werneck Orlando ¹, Maria de Lourdes Bueno Trindade Galo ¹, George Deroco Martins ², Andrea Maria Lingua ³, Vanessa Andaló ²

¹ São Paulo State University (UNESP), Presidente Prudente, São Paulo, Brazil - (vinicius.werneck, trindade.galo)@unesp.br
 ² Federal University of Uberlândia (UFU), Uberlândia, Minas Gerais, Brazil - (deroco, vanessaandalo)@ufu.br
 ³ Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino, Italy - andrea.lingua@polito.it

Keywords: Aerial Imaging, Spectral Analysis, Sustainable Agriculture, Pest Management, Infestation Monitoring.

Abstract:

Brazil, the world's largest coffee producer, faces challenges managing the coffee leaf miner (*Leucoptera coffeella*), a significant pest. This study suggests remote sensing for pest control decisions. Two experimental areas in the Cerrado region of Minas Gerais State were analyzed to spectrally characterize infested plants and estimate the number of mines per plant. Results show the ability to differentiate infested plants with greater reflectance variance in the near infrared at 850nm. The performances of the three machine learning algorithms were compared. Determining the number of mines in the group of most infested plants demonstrated slightly higher precision, achieving an RMSE of 22.69% using the Support Vector Machine algorithm. Conversely, the group of least-infested plants obtained the best result with the Random Forest algorithm, achieving an RMSE of 32.47%. These promising results indicated that CLM can be detected using aerial multispectral imaging data.

1. Introduction

Brazil, the world's largest coffee producer, aims to reach 58 million processed coffee bags in 2024, with the state of Minas Gerais contributing to over half of this total (CONAB, 2024). In Minas Gerais, the Cerrado Mineiro stands out as one of the largest producing regions, renowned for cultivating high-quality coffees. However, the abundance of coffee plantations, combined with the dry climate and high temperatures, creates ideal conditions for the proliferation of the main pest affecting the coffee plant's aerial part in the Cerrado Mineiro region: the Coffee Leaf Miner (CLM), *Leucoptera coffeella* (Lepidoptera: Lyonetiidae) (Souza et al. 1998; Dantas et al. 2020).

The attack of the CLM has specific characteristics. The insect only causes damage in the larval stage when it feeds on the leaf mesophyll, a fundamental tissue between the two sides of the epidermis. This type of feeding causes the formation of mines on the leaves, which is the origin of its common name. The presence of mines results in necrosis of the corresponding palisade and lacunose tissues, reducing the leaf surface. Consequently, this can lead to defoliation, decreased photosynthetic rate, reduced plant productivity, and in more serious cases, plant death (Costa et al. 2012).

According to Souza et al. (1998), during the dry period of the year, leaf fall begins from the top of the plants. Oviposition and feeding on leaf structures reduce the photosynthetic area (Liu et al. 2015) and auxin levels, increasing the synthesis of ethylene, the plant hormone responsible for leaf abscission (Arteca 1996; Souza et al. 1998), triggering chlorosis and leaf abscission from the top of the plants, as shown in Figure 1.

Additionally, according to the latest climate risk indicators, the future of *Coffea arabica* L. plantations in Brazil will depend on the adoption of adaptive and effective measures (Dias et al. 2024). This is essential as climate change will also lead to an increase in the number of CLM life cycles, making attacks more aggressive (Giraldo-Jaramillo et al. 2024).



Figure 1. Attack of the coffee leaf miner on coffee leaves. In (A) the leaves are healthy, while in (B) the leaves are infested, withered, and have chlorosis.

The average number of CLM mines in the field is typically determined using a manual methodology. In this approach, a field operator manually counts the average number of mines found on the leaves of the third or fourth pair of plagiotropic branches (side branches that grow at an angle to the main stem of the plant) in the upper middle third of coffee plants (Zampiroli et al. 2017). However, this method requires specialized knowledge and is time-consuming, posing challenges for large-scale coffee-producing areas.

As a sustainable and precise solution, remote sensing has been used to discriminate the most varied stresses in coffee trees. Martins et al. (2017), identified that the red, red edge and near-infrared (NIR) spectral bands were crucial for the discrimination of plants infected by nematodes. The multispectral classification achieved an overall accuracy of 78% and a Kappa coefficient of 0.71, highlighting the effectiveness of remote sensing in this context.

Previous studies in the Cerrado region of Minas Gerais have demonstrated the effectiveness of using remote sensing data and machine learning algorithms for various applications in coffee stresses. Using low-cost aerial images, Pereira et al. (2022) identified that the Random Forest (RF) and Support Vector Machine (SVM) algorithms demonstrated satisfactory accuracy, with a root mean square error (RMSE) of less than 26.5%, in estimating physical parameters of coffee trees, such as chlorophyll content and plant height. Also using low-cost aerial imagery, Orlando et al. (2023) estimated the Leaf Water Potential (LWP) of drip-irrigated coffee trees and obtained more accurate results using the Support Vector Machine algorithm (RMSE of 0.188) and an RMSE% of 34.18 in rainy conditions. In drought conditions, the Random Tree algorithm obtained the best result with an RMSE of 0.052 and an RMSE% of 32.

Vilela et al. (2023), selected the combination of the difference between two Sentinel-2 bands, "NIR-BLUE" and "NIR-RED", to assess CLM infestation levels. Their findings revealed that the RF algorithm outperformed the SVM algorithm in classifying the coffee leaf miner infestation. The RF algorithm achieved an overall accuracy and kappa index exceeding 0.89, whereas the SVM algorithm exhibited an overall accuracy of 81.8% and a kappa index of 0.61.

Considering the importance of detecting and monitoring the key pest affecting the aerial part of Brazilian coffee trees, it is hypothesized that high spatial resolution aerial multispectral images have the capability to detect plants infested by CLM in the Cerrado Mineiro region. The objective is to assess the feasibility of using multispectral aerial images to detect CLM infestations in the Cerrado region of Minas Gerais, as well as to estimate the average number of mines per plant.

2. Material and methods

The research was conducted within an experimental site situated on the Monte Carmelo Campus of the Federal University of Uberlândia. The site consists of two plots (Figure 2), where rows spaced 3.5 meters apart, and plants spaced 0.6 meters apart. The experimental area comprised *C. arabica*, planted in 2016 in plot 1 and 2015 in plot 2, with an average plant height recorded at 1.5 and 2 meters, respectively. In the first experimental plot, the entire area is planted with the cultivar Topázio MG-1190, susceptible to infestation by CLM. The second area consists of several cultivars, but the evaluation focused solely on plants of the same cultivar, ensuring standardization.

The location has an average annual precipitation of 1,444 mm and falls under the Aw climate classification (tropical climate with a dry winter), providing optimal conditions for the natural proliferation of the targeted pest. Both plots utilized a drip irrigation system, featuring emitters positioned at 0.6 meters apart and a flow rate of 1.6 L h-1.



Figure 2. The study area location and the spatial distribution of the 93 plants evaluated. In the first plot (1), plants were evaluated across the entire area (53 plants), while in the second plot (2), only plants of the same cultivar (Topázio MG-1190) as those in plot 1 were evaluated (40 plants).

2.1 Data Acquisition

The assessment of damage caused by CLM occurred in July 2022. The initial area encompasses around 200 coffee trees within a 650 m² space in Plot 1. In contrast, Plot 2 includes approximately 400 coffee trees within 1,300 m², featuring eight coffee cultivars arranged in a randomized block design (RBD). To maintain consistency with the cultivar assessed in the first plot, only the sections with the Topázio MG-1190 cultivar were chosen in the second plot. A total of fifty-three canopies were randomly examined in Plot 1, while Plot 2 had forty canopies analyzed, as illustrated in Figure 2.

For each plant, eight leaves on the third or fourth pair of plagiotropic branches were randomly selected from the middle/upper third of the plant. Among these, four leaves were assessed on the north and four on the south sides. The recorded metric was the average number of mines per plant, as shown in Figure 3. The plants under examination were georeferenced using a GNSS receiver, employing the real-time kinematic (RTK) positioning method.



Figure 3. Assessment of the average number of mines per plant. (A) An example of a leaf with 4 mines and (B) a coffee plant.

On the same date, as the other assessments, an aerial survey was conducted by affixing an AgroWing multispectral camera (Figure 4A) to an unmanned aerial vehicle (UAV) (Figure 4B). The camera is equipped with 14 spectral bands, responsive to radiation within wavelengths centered at 405, 430, 450, 490, 525, 550, 560, 570, 630, 650, 685, 710, 735, and 850 nm. The flight was executed at 1 p.m. at an altitude of 60 meters, ensuring a Ground Sampling Distance (GSD) of 1 cm.



Figure 4. Aerial multispectral data collection. In (A) the AgroWing camera with six optical lens heads, in (B) the unmanned aerial vehicle, and in (C) the targets fixed in the field to carry out the empirical line calibration.

For the radiometric calibration of the images, the spectral signatures of three non-reflective targets (Figure 4C) were measured on the day of the aerial survey, using a portable FieldSpec® spectroradiometer manufactured by Analytical Spectral Devices (ASD). The instrument operates within a spectral range of 325-1075 nm and boasts a spectral resolution of 0.016 μ m. A 10° field of view (FOV) was achieved by incorporating a specific filter. Furthermore, the distance from the spectrum to the target was maintained at approximately 11.4 cm, resulting in an instantaneous field of view (IFOV) of 1 cm. For each radiance measurement, an average of ten repetitions of the target radiance reading and the radiance of a reference Lambertian surface (Spectralon plate) (Jackson et al. 1992) were

simultaneously measured under identical lighting and observation conditions.

2.2 Aerial image processing

The raw images underwent alignment and band co-registration using the AgroWing Basic software, adhering to the recommended processing flow by the manufacturer. Subsequently, the 14 bands were stacked to generate a spectral cube. The resultant images were exported in tiff format and the orthomosaic was generated and georeferenced using Agisoft Metashape software.

For the radiometric calibration of the orthomosaic, the empirical line method was used. For this, a linear regression was generated for each band, relating the digital number of the pixel to the reflectance of the targets used (Porto et al. 2023). The empirical values of gain and offset were calculated in the ENVI software. Then, the Hemispherical Conical Reflectance Factor (HCRF) of the targets, as defined by Equation 1, was computed. This calculation involved taking the ratio of the average radiant flux reflected by the target to that of an ideal diffusing surface, represented by a Spectralon plate, under identical geometric and lighting conditions.

$$HCRF = \left(\frac{\text{Average of target radiance}}{\text{Average of corrected plate radiance}}\right) * k \tag{1}$$

Acknowledging that this plate utilized in the field may deviate from the ideal conditions of a laboratory diffusing surface, a calibration factor (k) was determined by referencing it to a meticulously conditioned plate available in the laboratory. This calibration factor accounts for any discrepancies introduced by the field conditions, thus ensuring the accuracy of the HCRF calculations in the specific context of the study.

For each plant evaluated in the field, the spectral averages of the reflectance values of the top of the plants, obtained in the images, were extracted and tabulated. Thus, the ideal range to discriminate CLM in plants was identified through an analysis that focused on determining the spectral region that presents the greatest variance in spectral signatures between the reflectances extracted from the images. This spectral characterization was also done individually between the plots due to their varying infestation conditions.

2.3 Estimation of the number of CLM mines per coffee plant

To assess the feasibility of estimating the number of mines per plant, regression models were developed for both plots. Based on the main algorithms used to estimate stresses in coffee trees from remote sensing data, three regression algorithms available in the WEKA 3.9.4 software (Waikato Environment for Knowledge Analysis) were trained: Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF). It was standardized to use fixed and pre-established hyperparameters provided by the software for each of the algorithms. Due to the different agronomic conditions of the Plots, each Plot was evaluated individually, with 80% of the samples used for training and 20% for testing. This approach ensured that the models were tailored to the specific characteristics of each plot, thereby enhancing the accuracy of the evaluations.

To validate the quality and ascertain the optimal prediction model, the Root Mean Square Error (RMSE) evaluation metric

was employed. The selection criterion for the best estimation model involved identifying the architecture that demonstrated the lowest RMSE and RMSE% (Equations 2 and 3, respectively) values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - x_m)^2}{n}}$$
(2)

RMSE % =
$$\sqrt{\frac{\sum_{i=1}^{n} (x_i - x_m)^2}{n}} X\left(\frac{100Xn}{\sum_{i=1}^{n} x_m}\right)$$
 (3)

where,

 x_i represents the estimated number of mines. x_m represents the measured number of mines. n the number of samples.

3. Results and discussion

3.1 Determination of plot infestations

Regarding the infestation levels, it is important to acknowledge that different areas exhibit varying conditions of CLM infestation. In the first plot, there is an initial trend towards a normal distribution, however, as values increase, a decline in data frequency is noted, along with a greater dispersion towards the upper extremes of the range. The wide range of data amplitudes also indicates significant variability within the sample, which has an average of approximately 20 mines per plant, ranging from a minimum of 4 and maximum of 40. The rightward tilt of the histogram (Figure 5) suggests a non-symmetrical distribution, with a higher concentration of lower values and a prolonged tail towards higher values.



Figure 5. Distribution of the average number of mines per plant evaluated in the two plots. The bars in pink represent Plot 1, while the bars in light blue represent Plot 2. The dark color in the center shows the overlay of the histograms from both plots.

In Plot 2, a perfectly normal distribution was observed, indicating symmetry around the mean. The data range is relatively narrow compared to Plot 1, with values ranging from 2 to 21, and the average of the data is approximately 10. This concentration of values around the mean suggests a significant clustering of data points around this central value.

Furthermore, as depicted in Figure 5, the second plot exhibits a smaller range of number of mines per plant compared to Plot 1, indicating that infestations are more advanced in the first plot.

Notably, all coffee trees on both plots were infested. Additionally, in both plots, the absence of outliers and the tendency towards a normal distribution curve highlights the consistency of the data collected.

3.2 Characterization of spectral signatures

Figure 6 illustrates the spectral signatures of all coffee trees in the first plot. Notably, the variability in spectral signatures begins in the 685 nm region of the spectrum. Therefore, the red edge and near-infrared are the regions of interest to explore the variability of the canopy of the most infested plot.



Figure 6. Spectral curves of the 53 plants evaluated in Plot 1.

In Figure 7 we can observe the spectra of all coffee trees in the least infested area (Plot 2). Once again, the variability in spectral signatures begins in the 685 nm region; however in this case there is a greater intensity of reflectance in these spectral bands. Furthermore, in this region, there is a greater variation in the spectral signatures of the plants in Plot 2 than in Plot 1.



Figure 7. Spectral curves of the 40 plants evaluated in Plot 2.

When analyzing the Plots in the bands of greatest variance (bands centered at wavelengths 710 nm, 735 nm and 850 nm), some differences arise when altering the quantity of heavily or lightly infested plants factored into the average calculation. In Plot 1 (Figure 8), the effect of the CLM infestations is more evident when calculating the spectral average of a few plants, since they were more infested in the extreme groups. That is, there is greater interference from heavily infested plants. This leaf wilting effect is also observed in the progressive drying of magnolias, where there is an increase in reflectance throughout the infrared region as the moisture content decreases (Jensen 2009). As more plants in the sample group are averaged, there are also more vigorous plants in the sample group, resulting in the spectral signature of the less-infested plant group having greater reflectance.

As more observations are included at the extremes, the behavior tends to converge towards the average. However, in the third condition, this trend is reversed as the impact of plants with numerous mines becomes more subdued, causing their values to cluster around the mean.





A similar analysis can be carried out for Plot 2 (Figure 9). When there are fewer plants in the sample group, the less-infested plants exhibit higher reflectance in the near-infrared region, consistent with the expected spectral behaviour when comparing healthy and stressed coffee leaves (Martins et al. 2017). As it is a less infested area and with a normal distribution of the number of mines (Figure 5), as more plants are incorporated into the sample group, there is a tendency for spectral signatures to show an average behaviour.





3.3 Models for estimating the number of mines per plant

Table 1 illustrates the RMSE and RMSE% obtained for the models utilized in estimating the number of mines across both plots. Plot 1, the most heavily infested, exhibits plants with high numbers of mines. In this context, employing the most significant bands (ranging between 650 and 850 nm) results in an error rate of 22.69%, marginally lower than when utilizing all bands, which yields an error rate of 25.94%.

Features	Algorithm	Plot 1		Plot 2	
		RMSE	RMSE %	RMSE	RMSE %
All bands	MLP	7.51	28.07	8.09	76.14
	SVM	6.94	25.94	4.09	38.49
	RF	6.84	25.57	3.45	32.47
Bands	MLP	6.42	24.00	4.47	42.07
650 to	SVM	6.07	22.69	3.56	33.50
850nm	RF	7.10	26.54	3.59	33.78

Table 1. RMSE and RMSE% results for each algorithm for Plot 1 (most infested) and Plot 2 (least infested). MLP: multilayer perceptron; SVM: Support Vector Machine; RF: Random Forest. In Plot 2, the models exhibit fewer errors in absolute numbers, as evidenced by smaller mean squared errors, when compared to Plot 1. However, the mean squared error percentage of the second plot, being less infested, makes it more difficult to characterize the extremes.

Random Forest was the only algorithm that showed an increase in RMSE with the selection of the most important bands as features. This algorithm, although robust, may be less effective with highly correlated data, as evidenced by the increase in RMSE when using the most relevant bands.

Despite the promising results of the other algorithms, the Multilayer Perceptron algorithm may have been sensitive to the hyperparameters that were pre-fixed and therefore may suffer from overfitting, especially with smaller data sets.

4. Conclusion

The spectral region starting at 650 nm, extending towards the near infrared, has been demonstrated to be ideal for discriminating CLM infestations in plants through the analysis of aerial multispectral images. This capability can be crucial to discriminate the levels of infestations.

The comparison of algorithm performance revealed that the SVM algorithm achieved slightly higher accuracy in determining the number of mines in the most infested plant group. Conversely, the Random Forest algorithm demonstrated superior performance for the group of less-infested plants. The coffee production areas, even if under the same cultivar, exhibit differences in canopy structures due to age disparities and variations in the average number of mines. This discrepancy significantly impacts the spectral response of the canopies, potentially increasing errors in estimating the average number of mines, particularly with a single aerial survey.

In this context, CLM can be detected using aerial multispectral image data. However, for more accurate detection of infestations, continuous monitoring of areas is recommended. This allows for the evaluation of spectral differences in canopy structures over time as the infestation progresses.

References

- Arteca, R. N. 1996. Plant growth substances: principles and applications. https://doi.org/10.1007/978-14757-2451-6
- CONAB COMPANHIA NACIONAL DE ABASTECIMENTO. 2024. Acompanhamento da safra brasileira de café. Brasília, DF. https://www.conab.gov.br/info-agro/safras/cafe. Accessed 11 March 2024
- Costa, J. N. M., Teixeira, C. A. D., Júnior, J. R. V., Rocha, R. B., & Fernandes, C. de F. 2012. Informações para facilitar a identificação das diferentes fases do bichomineiro (*Leucoptera coffeella*) em campo. *Embrapa*, *Comunicado Técnico 384*, 1(1), 4.
- Dantas, J., Motta, I., Vidal, L., Bilio, J., Pupe, J. M., Veiga, A., et al. 2020. A comprehensive review of the coffee leaf miner Leucoptera coffeella (Lepidoptera : Lyonetiidae), with special regard to neotropical impacts, pest management and control. *Preprints*, *1*(October), 1–25. https://doi.org/10.20944/preprints202010.0629.v1
- Dias, C. G., Martins, F. B., & Martins, M. A. 2024. Climate risks and vulnerabilities of the Arabica coffee in Brazil under current and future climates considering new CMIP6 models. *Science of The Total Environment*, 907,

167753.

https://doi.org/10.1016/J.SCITOTENV.2023.167753

- Giraldo-Jaramillo, M., Quiroga-Mosquera, A., & Fernandes, F. L. 2024. Thermal requirements and estimation of the number of generations of Leucoptera coffeella (Guérin-Méneville, 1842) (Lepidoptera: Lyonetiidae) in Minas Gerais state, Brazil. Crop Protection, 175, 106483. https://doi.org/10.1016/J.CROPRO.2023.106483
- Jackson, R. D., Clarke, T. R., & Susan Moran, M. 1992.
 Bidirectional calibration results for 11 spectralon and 16
 BaSO4 reference reflectance panels. *Remote Sensing of Environment*, 40(3), 231–239.
 https://doi.org/10.1016/0034-4257(92)90005-5
- Jensen, J. R. 2009. Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres. *Information Systems Journal*, 21, 587.
- Liu, W. H., Dai, X. H., & Xu, J. S. 2015. Revisión de las influencias de los insectos minadores de hojas en sus plantas huésped. *Collectanea Botanica*, 34, 005. https://doi.org/10.3989/collectbot.2015.v34.005
- Martins, G. D., Galo, M. D. L. B. T., & Vieira, B. S. 2017. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. *IEEE Journal of Selected Topics in Applied Earth Observations* and Remote Sensing, 10(12). https://doi.org/10.1109/JSTARS.2017.2737618
- Menezes Freitas, R. A. S., Martins, G. D., Assis, G. A.,
 Siquieroli, A. C. S., dos Santos Fernandes, M. I., Soares,
 M. O. Q. S., & da Silva Pinheiro, B. E. C. C. 2023.
 Hyperspectral characterization and estimation models for agronomic parameters of coffee cultivars after pruning. *Precision Agriculture*, 24(6), 2374–2394.
 https://doi.org/10.1007/s11119-023-10044-6
- Orlando, V. S. W., Martins, G. D., Fraga Júnior, E. F., Marra, A. B., Pereira, F. V., & Galo, M. d. L. B. T. 2023.
 Potential of multispectral images taken by sensors embedded in uavs for monitoring the coffee crop irrigation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-1/W1-2023*, 91–96. https://doi.org/10.5194/isprs-annals-X-1-W1-2023-91-2023
- Pereira, F. V., Martins, G. D., Vieira, B. S., de Assis, G. A., & Orlando, V. S. W. 2022. Multispectral images for monitoring the physiological parameters of coffee plants under different treatments against nematodes. *Precision Agriculture*. https://doi.org/10.1007/S11119-022-09922-2
- Porto, L. R., Berveglieri, A., & Imai, N. N. 2023. Calibração radiométrica em imagens multiespectrais aéreas utilizando linha empírica e placa colorimétrica. In Anais do XX simpósio brasileiro de sensoriamento remoto (pp. 2993–2997). São José dos Campos: INPE. https://proceedings.science/sbsr-2023/trabalhos/calibracao-radiometrica-em-imagensmultiespectrais-aereas-utilizando-linha-empir?lang=pt-br. Accessed 23 February 2024
- Souza, J. C., Reis, P. R., & Rigitano, R. L. O. 1998. Bichomineiro do cafeeiro: Biologia, Danos e Manejo Integrado (2nd ed., Vol. n54).
- Vilela, E. F., Ferreira, W. P. M., Castro, G. D. M. de, Faria, A. L. R. de, Leite, D. H., Lima, I. A., et al. 2023. New Spectral Index and Machine Learning Models for Detecting Coffee Leaf Miner Infestation Using Sentinel-2 Multispectral Imagery. *Agriculture*, 13(2), 388. https://doi.org/10.3390/agriculture13020388
- Zampiroli, R., Alvarenga, C. B. de, Andaló, V., Rinaldi, P. C., & Assis, G. A. de. 2017. Application technology for chemically controlling coffee leaf miner in the cerrado of

Minas Gerais State. *Revista de Ciências Agrarias -Amazon Journal of Agricultural and Environmental Sciences*, 60(3), 256–262. https://doi.org/10.4322/rca.2583