Research Article
doi: 10.1590/01047760202632013570.
vol(32), 2026

CERNE

Machine learning-based assessment of leaf-cutting ant infestation in *Eucalyptus* forest plantations

Vitória Fernanda dos Santos[1], Clayton Alcarde Alvares[2], Maurício Magalhães Domingues[3], Everton Pires Soliman[3], Ítalo Ramos Cegatta[4], Edival Ângelo Valverde Zauza[4], Reginaldo Gonçalves Mafia[4]

[1] Suzano SA, Jacareí-SP, Brazil.
[2] Suzano SA, Limeira-SP, Brazil.
[3] Suzano SA, Três Lagoas-MS, Brazil.
[4] Suzano SA, São Paulo-SP, Brazil.

SIVICULTURE

ABSTRACT

Background: Brazilian planted forests play a critical role in global wood and fiber production but face significant productivity challenges from pests and diseases. Leaf-cutting ants (*Atta* spp. and *Acromyrmex* spp.) are the main pest in the Brazilian planted forests and cause significant productivity losses every year. Identifying areas most susceptible to colony establishment and growth is crucial for implementing effective management strategies. This study aims to assess the influence of edaphoclimatic and landscape variables on the establishment and expansion of leaf-cutting ant colonies and identify the key factors driving their occurrence.

Results: Based on a decade of monitoring data from 33,000 *Eucalyptus* stands across five regions, Random Forest models reached accuracies of 83% for predicting initial nests and 78% for predicting large nests.

Conclusions: The machine learning models effectively detected both initial and large nests, revealing that edaphoclimatic and landscape conditions exert varying levels of influence across macro-regions.

Keywords: *forest pests; random forest algorithm; pest management; spatial data science.*

HIGHLIGHTS

Leaf-cutting ants threaten productivity in Brazilian planted forests
A decade of monitoring assessed nests in 33,000 *Eucalyptus* stands
Random Forest models predicted initial and large nest occurrence
Edaphoclimatic and landscape factors influence infestation patterns

SANTOS, V. F.; ALVARES, C. A.; DOMINGUES, M. M.; SOLIMAN, E. P.; CEGATTA, Í. R.; ZAUZA, E. A. V.; MAFIA, R. G. Machine learning-based assessment of leaf-cutting ant infestation in Eucalyptus forest plantations. CERNE, v. 32, e103570, 2026. DOI: 10.1590/01047760202632013570.

✤Corresponding author: vfsantos@suzano.com.br
Scientific Editor: Gilvano Ebling Brondani and Nilza de Lima Pereira Sales

Received: March 27 2025
Accepted: May 19 2026

CERNE
ISSN 0104-7760
UNIVERSIDADE FEDERAL DE LAVRAS

INTRODUCTION

The Brazilian planted forest sector plays an important role in the global supply of fiber, wood, and energy. With approximately 10 million hectares designated for planted forests, predominantly *Eucalyptus* (76.9%), the country produced 25 million tons of pulp in 2022, achieving an average productivity of 33 m^3 ha^{-1} $year^{-1}$ in eucalypt plantations (INDÚSTRIA BRASILEIRA DE ÁRVORES, 2023). However, both biotic and abiotic factors can stagnate or even reduce forest productivity.

Leaf-cutting ants are the most significant pests affecting Brazilians forest plantations, causing severe losses and incurring high monitoring and control costs (Araújo *et al.*, 1997; Della Lucia and Souza, 2011). In 2016, for example, the potential economic impact of this pest considering the planting areas was US$ 8 billion (Instituto de Pesquisas e Estudos Florestais, 2016).

Leaf-cutting ants are social insects that build underground nests to protect the queen and create a suitable microenvironment for brood development (Cardoso *et al.*, 2014). During nest construction, the excavated soil is deposited on the surface, and the area of loose soil can be used to infer about nest depth and the number of chambers, providing a useful parameter to assess nest size and growth (Moreira *et al.*, 2004). Around 24 months after excavation, nests can cover up to 50 m^2 of loose soil on the surface (Grandeza *et al.*, 1999). This poses a significant challenge to forest productivity, considering that that each square meter of nest per hectare is associated with a reduction in eucalypt forest yield ranging from 0.04 to 0.13 $m^3 \cdot ha^{-1}$ (Souza *et al.*, 2011).

The establishment and growth of ant colonies are influenced by factors such as precipitation, temperature, soil texture, and proximity to native forests (Camargo *et al.*, 2013; Chiles *et al.*, 2022; Zanetti *et al.*, 2000). The survival of newly founded colonies has been shown to be related to soil texture, temperature and solar irradiance. In laboratory tests, Oxisols were preferred for colony establishment and exhibited higher survival rates (Schoereder and Da Silva, 2008). In addition, shaded environments favored colony survival compared to open areas, which are characterized by intense solar irradiation and higher temperatures (Sousa *et al.*, 2022). Nest density was also positively correlated with temperature and relative humidity and negatively correlated with precipitation and proximity to native forest fragments (Canuto *et al.*, 2019; Zanetti *et al.*, 2000).

Identifying areas most susceptible to colony establishment and expansion is essential for implementing effective management strategies to ensure both production sustainability and ecosystem balance. Machine learning and modeling approaches have been widely used to understand pests-environment interactions (Aidoo *et al.*, 2022; De França *et al.*, 2024; Hentschel *et al.*, 2018; De Oliveira Aparecido *et al.*, 2020). For ant species, studies have assessed the global risk of establishing for *Solenopsis invicta* and *Anoplolepis gracilipes*, both listed among the world's 100 worst invasive species (IUCN/SSC Invasive Species Specialist Group), using modeling and anthropogenic, edaphoclimatic, and landscape variables (Chen *et al.*, 2020; Lee *et al.*, 2022). In Brazil, Schaefer *et al.* (2021) evaluated how soil, climate and land use variables affect the distribution of nine leaf-cutting ant species. Their findings revealed that climatic gradients and soil properties are the primary factors shaping the distribution patterns of these species.

However, little is known about how edaphoclimatic and landscape variables interact and influence the dynamics of leaf-cutting ant colonies establishment and growth. Therefore, this study aims to identify the key edaphoclimatic and landscape variables influencing the occurrence of initial and large leaf-cutting ant nests in managed *Eucalyptus* forests.

MATERIAL AND METHODS

Study Area

The study was conducted in managed *Eucalyptus* forests owned by Suzano S.A. Between 2010 to 2019, the presence, size, and density of leaf-cutting ant nests were monitored over 33,000 forest stands, covering a total of 729,000 ha (Figure 1). The study area covers a diverse range of edaphoclimatic conditions, with annual precipitation ranging from 741 to 2558 mm and average annual temperatures between 15.4°C and 25.8°C (Alvares *et al.*, 2013). Altitude varies from 1 m to 1600 m. The predominant soil types include mainly Ultisols (44%), Oxisols (38%), and Entisols (11%)

Monitoring of leaf-cutting ants

Monitoring of leaf-cutting ants conducted annually in forests older than one year and continued until harvesting age (5-7 years), primarily between February and August. The surveys focused exclusively on colonies of the genus *Atta*. In each stand, 5% of the total area was sampled using transects, with a minimum of two transects per 30 ha. Each transect was 21 m wide, with 10.5 m surveyed on each side of the centerline (Figure 2).

Field observations recorded the number of nests (physical mound structures), nest size (m^2 of loose soil, calculated as longest length × widest width), and the distance of each nest from the transect centerline. Data from each transect was extrapolated to estimate nest density (nests per hectare), assuming a 5% sampling intensity. However, understory vegetation density can reduce the probability of visually detecting a nest within the transect, potentially affecting extrapolated densities (Nelson Sanches Bezerra Júnior, Equilíbrio Florestal company, personal communication). Nest density per hectare was then classified into five nest size classes based on mound area: 1) 0.1 – 0.50; 2) 0.51 – 1.00; 3) 1.01-10; 4) 10.01-50; 5) > 50.1 m^2 ha^{-1}

Class 1 represents newly established colonies with little or no surrounding loose soil. Class 2 nests area also in the early stages, but they exhibit up to 1 m^2 of loose soil. Both class 1 and 2 were considered indicative of recent

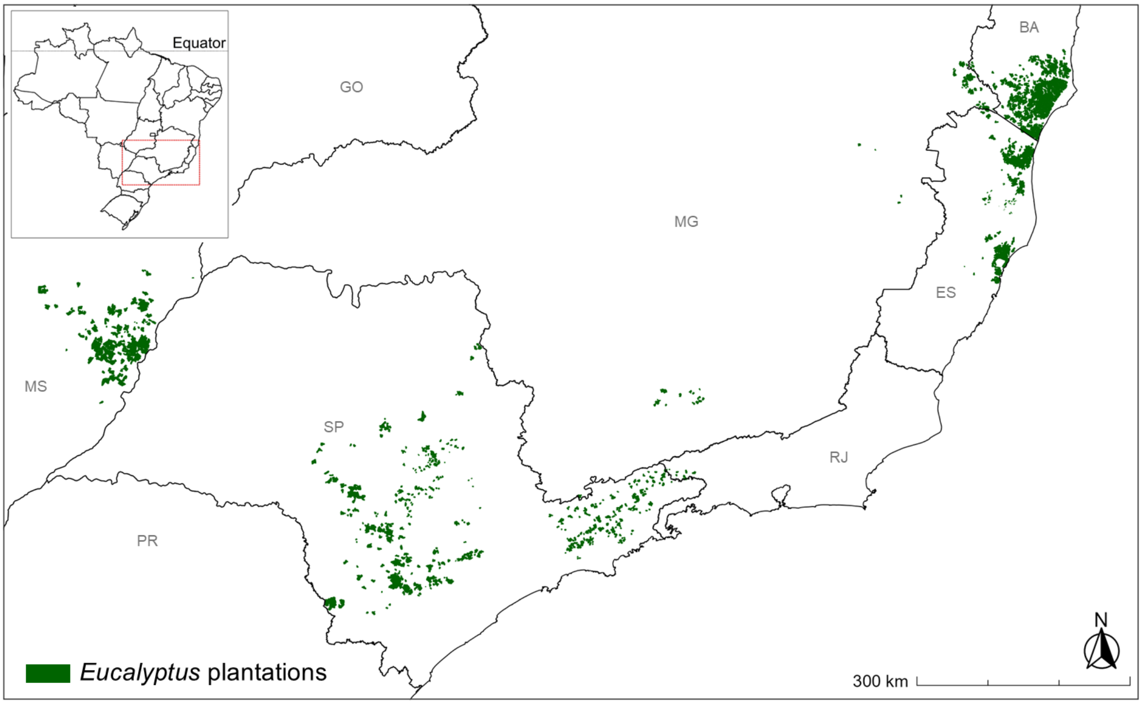

Figure 1: *Eucalyptus* plantations monitored from 2010 to 2019.

nuptial-flight events (infestation). Class 3 includes nests in intermediate development, while classes 4 and 5 represent mature colonies, with class 5 being the most challenging to control.

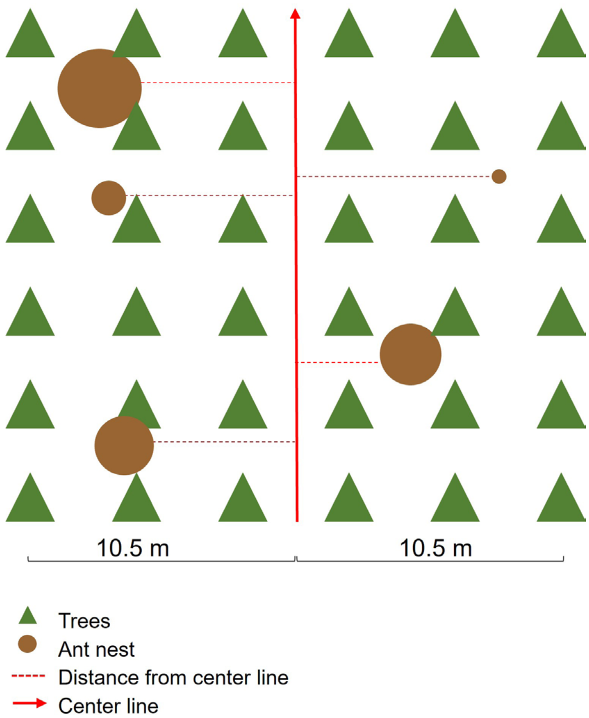

Figure 2: Illustration of the 21 m wide transect used for monitoring leaf-cutter ant nests.

Edaphoclimatic and landscape data

Monthly meteorological data from 2010 to 2019, including precipitation (PPT), potential evapotranspiration (ETP), average temperature (Tmed), minimum temperature (Tmin), maximum temperature (Tmax), relative humidity (RH), and radiation (RAD), were obtained from public meteorological stations (INMET - National Institute of Meteorology, https://mapas.inmet.gov.br) and supplemented with data from private company-owned stations (Alvares *et al.*, 2023). Data were interpolated at the stand level using the Inverse Distance Weighting (IDW) method with a 50 km radius. Where data gaps existed, missing values were supplemented using meteorological information from the NASA POWER platform (National Aeronautics and Space Administration/Prediction of Worldwide Energy Resources, https://power.larc.nasa.gov (Sparks *et al.*, 2022; Stackhouse *et al.*, 2018), which integrates satellite-based observations and modeling.

From these meteorological variables (PPT, ETP, Tmed, Tmin, Tmax, RH, and RAD), annual summary statistics were derived to represent seasonal variations, amplitude, and extreme values. A total of 83 climatic variables were generated (Table 1). Additional environmental data included altitude (Farr and Kobrick, 2000), slope (derived from altitude), content of clay and sand in the soil at 0-30 cm depth (Vasques *et al.*, 2021), and soil available water capacity (AWC) (National Water Agency, 2021, https://metadados.snirh.gov.br).

To assess the influence of surrounding land use, a 600-m buffer was delimited around each stand. The proportion at different land use categories was extracted annually from

MapBiomas (Projeto MapBiomas – Coleção 6, Souza *et al.* (2020)). The classification and groupings of land use categories are described in Table S1. In total, 97 edaphoclimatic and landscape variables were considered (Table 1).

Table S1: MapBiomas Land use classification (Collection 6) and the respective groupings.

Landuse (Mapbiomas)	Group
Agriculture	agriculture
Temporary Crop	agriculture
Soybean	agriculture
Sugar cane	agriculture
Rice	agriculture
Other temporary Crops	agriculture
Mosaic Agriculture and Pasture	agriculture_pasture
Perennial Corp	perennial_crop
Coffee	perennial_crop
Citrus	perennial_crop
Other Perennial Crop	perennial_crop
Wetlands	wetlands
Grassland	grassland
Forest	forest
Forest Formation	forest
Savanna Formation	forest
Mangrove	forest
Wooded Restinga	forest
Other non Forest Formations	others_occupations
Farming	pasture
Pasture	pasture
Forest Plantation	forest_plantation

Data analysis

To identify locations with similar edaphic and climatic conditions, the forest stands were divided into five groups using the K-means clustering method (Diday, 1972). The clustering process was based on clay (%) and sand (%) content (Vasques *et al.*, 2021), as well as historical climate: annual precipitation (mm), precipitation coefficient of variation (%), average temperature (°C), average temperature coefficient of variation (%), minimum temperature (°C), and maximum temperature (°C). These climate variables were derived from monthly averages for the period 1980 to 2009 (Xavier *et al.*, 2016). Prior to clustering, all variables were normalized to a 0-1 scale using to equation Xnorm = (X – Xmin) / (Xmax – Xmin).

Modeling process

The influence of edaphoclimatic and landscape variables on the presence of initial and large nests in each cluster was analyzed, considering the presence or absence of initial and large nests as the dependent variable. The presence of initial nests was determined by the sum of the densities of nests in classes 1 and 2 (present = ∑class 1+class 2> 0). The presence of large nests was determined by the sum of the densities of nests in classes 4 and 5 (present = ∑class 4+class 5 > 0). These dependent variables were analyzed related to 97 edaphoclimatic and landscape variables using the RandomForest machine learning model, implemented in Python with the Scikit-learn package (Pedregosa *et al.*, 2011; Python Software Foundation, 2022). The model was trained on 80% of the data (training base) and validated on the remaining 20% (test base).

Since the database is imbalanced between the present and absence classes, a subsampling technique was tested in the training set. This approach adjusted the number of observations in the larger class to match the number in the smaller class through random selection within the original dataset.

In addition to specific models (with and without subsampling) for each cluster, a general model (without clusters distinction) was also evaluated with and without subsampling. For model training, the default parameters of the Scikit-learn package were used, except for the n_estimator and max_depth parameters, which were set to 500 and 10, respectively. Models performance was evaluated based on accuracy, precision, sensitivity, and specificity, calculated from the confusion matrix using the following equations:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

Where: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

RESULTS

The cluster analysis identified five forest areas with distinct climates and soils (Figure 3). The characteristics of these clusters are summarized in Table 2.

During the evaluation period, approximately 39%, 78%, 90%, 73%, and 48% of the stands presented initial nests (nests class 1 or 2) in clusters 1, 2, 3, 4, and 5, respectively. Meanwhile, 26% (C1), 49% (C2), 59% (C3), 60% (C4) and 44% (C5) of the stands presented large nests (class 4 or 5) at some stage of the eucalypts growth rotation. Figure 4 shows the spatial dispersion of initial and large nests from 2010 to 2019.

The models for initial nests prediction without subsampling showed high accuracy across all clusters: 84% (General), 89% (C1), 86% (C2), 81% (C3), 83% (C4), and 88% (C5). However, in clusters with an unbalanced number of positive (present) and negative (absent) observations, sensitivity and specificity were lower. After applying subsampling, model accuracy decreased across all clusters, but the mean sensitivity and specificity increased for C1, C3, and C4, while remaining stable for C2, C4, and

Table 1: Description of edaphoclimatic and landscape variables used for modelling

Group	Variable	Description	Unit
C	min_day_legth	Minimum day length in the year	hours
C	range_etp	Range of ETP in the year	mm
C	range_etp_py	Range of ETP in the previous year	mm
C	max_etp	Maximum ETP in the year	mm
C	max_etp_py	Maximum ETP in the previous year	mm
C	sd_etp	Sd of ETP in the year	mm
C	sd_etp_py	Sd of ETP in the previous year	mm
C	months_w/o_rain	Number of months with PPT below 20 mm in the year	un
C	months_w/o_rain_py	Number of months with PPT below 20 mm in the previous year	un
C	ppt_4qtr	PPT in the 4th quarter of the year (Oct-Dec)	mm
C	ppt_4qtr_py	PPT in the 4th quarter of the previous year (Oct-Dec)	mm
C	acumul_ppt	Accumulated PPT in the year	mm
C	acumul_ppt_py	Accumulated PPT in the previous year	mm
C	range_ppt	Range of PPT in the year	mm
C	range_ppt_py	Range of PPT in the previous year	mm
C	max_ppt	Maximum monthly PPT in the year	mm
C	max_ppt_py	Maximum monthly PPT in the previous year	mm
C	min_ppt	Minimum monthly PPT in the year	mm
C	min_ppt_py	Minimum monthly PPT in the year previous	mm
C	sd_ppt	Sd of PPT in the year	mm
C	sd_ppt_py	Sd of PPT in the previous year	mm
C	ppt_cold_quarter	PPT accumulated in the coldest quarter of the year	mm
C	ppt_cold_quarter_py	PPT accumulated in the coldest quarter of the previous year	mm
C	ppt_warm_quarter	PPT accumulated in the warmest quarter of the year	mm
C	ppt_warm_quarter_py	PPT accumulated in the warmest quarter of the previous year	mm
C	ppt_dri_quarter	PPT accumulated in the driest quarter of the year	mm
C	ppt_dri_quarter_py	PPT accumulated in the driest quarter of the previous year	mm
C	ppt_wett_quarter	PPT accumulated in the wettest quarter of the year	mm
C	ppt_wett_quarter_py	PPT accumulated in the wettest quarter of the previous year	mm
C	range_rad	Range of RAD in the year (Maximum RAD - Minimum RAD)	MJ
C	range_rad_py	Range of RAD in the previous year (Maximum RAD - Minimum RAD)	MJ
C	rad_max	Maximum RAD in the year	MJ
C	rad_max_py	Maximum RAD in the previous year	MJ
C	rad_mean	Average RAD in the year	MJ
C	rad_mean_py	Average RAD in the previous year	MJ
C	rad_min	Minimum RAD in the year	MJ
C	rad_min_py	Minimum RAD in the previous year	MJ
C	rad_sd	Sd of RAD in the year	MJ
C	rad_sd_py	Sd of RAD in the previous year	MJ
C	temp_range	Tmax - Tmin in the year	°C
C	temp_range_py	Tmax - Tmin in the previous year	°C
C	range_tmax	Range of Tmax in the year	°C
C	range_tmax_py	Range of Tmax in the previous year	°C
C	high_tmax	Highest Tmax of the year	°C
C	high_tmax_py	Highest Tmax of the previous year	°C
C	tmax_mean	Average Tmax in the year	°C
C	tmax_mean_py	Average Tmax in the previous year	°C
C	low_tmax	Lowest Tmax of the year	°C
C	low_tmax_py	Lowest Tmax of the previous year	°C

Group	Variable	Description	Unit
C	range_tmean	Range of Tmed in the year	°C
C	range_tmean_py	Range of Tmed in the previous year	°C
C	high_tmed	Highest Tmed in the year	°C
C	high_tmed_py	Highest Tmed in the previous year	°C
C	tmed_mean	Average monthly temperature in the year	°C
C	tmed_mean_py	Average monthly temperature in the previous year	°C
C	low_tmed	Lowest mean temperature in the year	°C
C	low_tmed_py	Lowest mean temperature in the previous year	°C
C	tmed_cold_quarter	Average temperature in the coldest quarter of the year	°C
C	tmed_cold_quarter_py	Average temperature in the coldest quarter of the previous year	°C
C	tmed_warm_quarter	Average temperature in the hottest quarter of the year	°C
C	tmed_warm_quarter_py	Average temperature in the hottest quarter of the previous year	°C
C	tmed_dri_quarter	Average temperature in the driest quarter of the year	°C
C	tmed_dri_quarter_py	Average temperature in the driest quarter of the previous year	°C
C	tmed_wett_quarter	Average temperature in the wettest quarter of the year	°C
C	tmed_wett_quarter_py	Average temperature in the wettest quarter of the previous year	°C
C	range_tmin	Range of Tmin in the year	°C
C	range_tmin_py	Range of Tmin in the previous year	°C
C	high_tmin	Highest Tmin of the year	°C
C	high_tmin_py	Highest Tmin of the previous year	°C
C	tmin_mean	Average Tmin in the year	°C
C	tmin_mean_py	Average Tmin in the previous year	°C
C	low_tmin	Lowest Tmin of the year	°C
C	low_tmin_py	Lowest Tmin of previous the year	°C
C	range_rh	Range of RH in the year	%
C	range_rh_py	Range of RH in the previous year	%
C	rh_max	Maximum RH in the year	%
C	rh_max_py	Maximum RH in the previous year	%
C	rh_mean	Average RH in the year	%
C	rh_mean_py	Average RH in the previous year	%
C	rh_min	Minimum RH in the year	%
C	rh_min_py	Minimum RH in the previous year	%
L	agriculture	Percentage of land occupied by agriculture	%
L	agriculture_pasture	Percentage of land occupied by agriculture and pasture	%
L	perennial_crop	Percentage of land occupied by perennial crops	%
L	wetlands	Percentage of land occupied by wetlands	%
L	grassland	Percentage of land occupied by grassland formation	%
L	forest	Percentage of land occupied by forest formation	%
L	others_occupations	Percentage of land occupied by other natural formations	%
L	pasture	Percentage of land occupied by pasture	%
L	forest_plantation	Percentage of land occupied by forest plantation	%
S	sand	Percentage of sand content in the soil	%
S	clay	Percentage of clay content in the soil	%
S	awc	Available water capacity of the soil	mm
T	altitude_range	Stand altitude range	m
T	altitude_mean	Stand average altitude	m
T	slope_range	Stand slope range	%
T	slope_mean	Stand average slope	%

Where the letters C, T, S, and L in the group column means Climate, Topography, Soil, and Land Use. ETP = Potential Evapotranspiration, PPT = Precipitation, RAD = Radiation, Tmed = Average Temperature, Tmax = Maximum Temperature, Tmin = Minimum temperature, RH = Relative Humidity

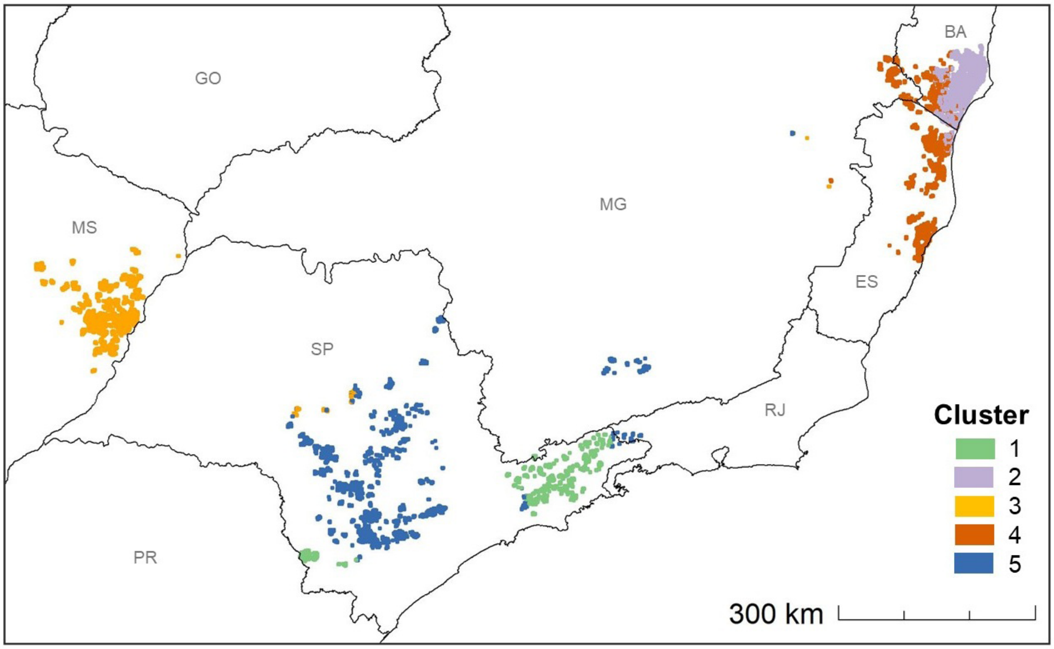

Figure 3: Spatial distribution of the cluster analysis results, considering edaphoclimatic variables as predictors

general model (Table 3). The general model for site-specific characterization performed well, with 85% accuracy and 77% sensitivity (Table 4).

Similarly, models evaluating the presence of large nests exhibited high accuracy without subsampling (80% General, 91% C1, 84% C2, 79% C3, 76% C4, 87% C5). However, their sensitivity was low, leading to false negatives. When using subsampling, accuracy decreased across all regions, but model sensitivity improved: 78% General 81% C1, 75% C2, 69% C3, 76% C4, 75% C5. This adjustment enhanced the identification of areas with large nest occurrences (Table 5). The general model for site-specific characterization demonstrated a balanced performance, with 72% accuracy and 77% sensitivity (Table 6).

The importance of the 15 most influential edaphoclimatic and landscape variables in models with subsampling is illustrated in figures 5 and 6. The importance index quantifies each variable's contribution to the model outcomes. For initial nests models, climate-related variables were present in all clusters. However, in the General model, C1, and C3, altitude-related variables were also significant. In C3, agricultural and pasture percentages in stand buffers were particularly relevant. For large nest models, terrain, soil, and landscape variables were more frequently observed. Average altitude was a key predictor variable in all models except for C3.

DISCUSSION

Machine learning models applied to leaf-cutter ant monitoring in managed eucalypt forests provide evidence that edaphoclimatic and landscape variables influence the occurrence of both initial and large nests.

Segmenting the forest plots into five clusters facilitated the differentiation of distinct climates and soils, primarily driven by variations in precipitation and average temperature, and highlighted clear seasonal differences across regions. This regional segmentation also reduced the risk of overgeneralizing *Atta* species distribution.

Table 2: Edaphoclimatic characteristics of the clusters

Cluster	PPTa	PPT CV	Tmed	Tmed CV	Tmin	Tmax	Clay	Sand	Nº Stands	Area
	mm	(%)	(ºC)	(ºC)	(ºC)	(ºC)	(%)	(%)		(ha)
1	1,457	57	19.5	13.6	14.0	25.0	39	43	2,074	43,826
2	1,295	34	24.6	7.2	20.0	29.2	19	63	10,386	191,382
3	1,292	63	24.1	9.3	18.4	29.7	20	64	4,833	207,014
4	1,217	54	24.7	7.5	19.9	29.5	30	52	9,679	127,230
5	1,417	56	21.2	11.3	15.9	26.6	36	45	6,210	159,257

Where: PPTa = Average annual precipitation, PPT CV = Average monthly precipitation coefficient of variation, Tmed = Average annual temperature, Tmed CV = Average monthly temperature coefficient of variation, Tmin = Average annual minimum temperature, Tmax = Average annual maximum temperature, Clay = Average percentage of clay, Sand = Average percentage of sand

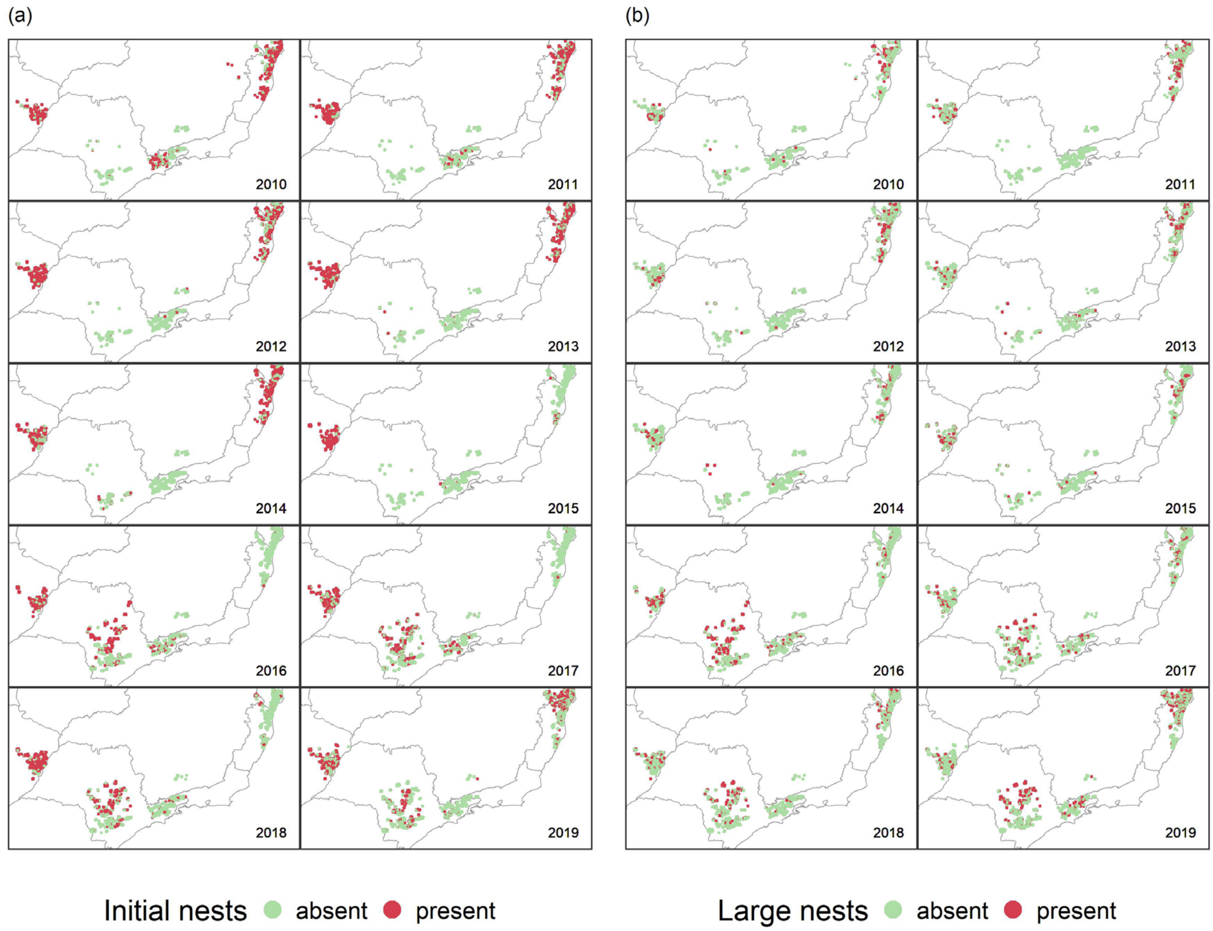

Figure 4: Spatial distribution of initial (a) and large nests (b) from 2010 to 2019

Although we did not conduct a species-level survey and all monitored nests were identified only at the genus level, it is well established that some *Atta* species are restricted to specific geographic areas (Delabie *et al.*, 2011; Schaefer *et al.*, 2021).

Models predicting initial nest showed high accuracy across both general and site-specific applications. Sensitivity improved with subsampling, reinforcing that environmental variables influence leaf-cutter ant distribution. The general model showed high accuracy (85% in average) for initial

Table 3: Performance metrics of models for initial nest

Model	Model Statistics									
	Accuracy		Precision		Sensitivity		Specificity		N training	
	RF	RF_{sub}	RF	RF_{sub}	RF	RF_{sub}	RF	RF_{sub}	RF	RF_{sub}
General	0.84	0.83	0.76	0.74	0.86	0.89	0.82	0.79	1=48,068 0=72,267	1=48,068 0=48,068
C1	0.89	0.75	0.72	0.34	0.33	0.85	0.98	0.73	1=1,093 0=7,090	1=1,093 0=1,093
C2	0.86	0.85	0.79	0.76	0.88	0.92	0.84	0.80	1=15,657 0=22,958	1=15,657 0=15,657
C3	0.81	0.75	0.82	0.90	0.97	0.75	0.30	0.74	1=13,964 0=4,446	1=4,446 0=4,446
C4	0.83	0.82	0.74	0.71	0.86	0.90	0.81	0.76	1=13,709 0=21,549	1=13,709 0=13,709
C5	0.88	0.81	0.76	0.49	0.53	0.82	0.96	0.81	1=3,674 0=16,194	1=3,674 0=3,674

Where RF = Random Forest model without subsampling, RFsub = Random Forest model with subsampling, N training = Number of observations in the training base, where 1 = presence of initial nest and 0 = absence

Table 4: Validation of the use of the general initial nest model with sub-sampling for site-specific regions

Model	Model Statistics				
	Accuracy	Precision	Sensitivity	Specificity	N test
C1	0.90	0.76	0.38	0.98	1=286, 0=1,760
C2	0.84	0.74	0.92	0.78	1=3,859, 0=5,795
C3	0.79	0.79	0.99	0.13	1=3,526, 0=1,077
C4	0.81	0.70	0.91	0.75	1=3,442, 0=5,373
C5	0.90	0.77	0.66	0.95	1=931, 0=4,036
Mean	0.85	0.75	0.77	0.72	

Where N test = number of observations in the test base, where 1 = presence of initial nest and 0 = absence.

nests classification across clusters. However, sensitivity was lower for C1 and C5, likely due to imbalanced test datasets, where initial nests absence was disproportionately high (86% in C1 and 81% in C5). Thus, the general model with subsampling is effective for C2, C3, and C4, while site-specific models are recommended for C1 and C5 to improve sensitivity. Our findings are consistent with different studies that also observed that edaphoclimatic and landscape variables influence the occurrence of different pests in agriculture and forest (De França *et al.*, 2024; Hentschel *et al.*, 2018; De Oliveira Aparecido *et al.*, 2020), and that subsampling techniques can enhance model sensitivity, especially when dealing with class imbalance as describe in Benkendorf *et al.* (2023).

The influence of edaphoclimatic and landscape variables on initial nests models varied across clusters. Key predictors included cumulative precipitation from the previous year, minimum radiation, minimum precipitation, maximum and minimum temperatures, minimum day length, and standard deviation of precipitation in the previous year. This is consistent with expectations, as initial nests (<1 m^2) originate from recent nuptial-flight (13-14 months old) influenced by preceding climatic conditions (Grandeza *et al.*, 1999). These variables highlight the role of climate seasonality in the nuptial-flights events and colony establishment, with nuptial flights typically occurring at the onset of rains following dry periods (Marinho *et al.*, 2011; Staab and Kleineidam, 2014) and optimal temperatures favoring colony survival (Staab and Kleineidam, 2014). Solar radiation affects environmental and soil temperatures, influencing chamber depth and colony viability (Sousa *et al.*, 2022). Photoperiod impacts social insects foraging (Lei *et al.*, 2019), with latitude-driven day length variations affecting C3 and C5.

Climatic variables were the primary influence across all regions. Average altitude was also significant in the general model and in regions C1 and C3. Notably, C3 was the only region where soil and landscape variables ranked among the 15 most influential factors.

Models predicting the presence of large leaf-cutter ant nests demonstrated high accuracy for both general and site-specific applications. Sensitivity improved with subsampling, highlighting the influence of environmental variables on nest growth. The general model achieved an

Table 5: Performance metric of models for large nests

Model	Model Statistics									
	Accuracy		Precision		Sensitivity		Specificity		N training	
	RF	RF$_{sub}$	RF	RF$_{sub}$	RF	RF$_{sub}$	RF	RF$_{sub}$	RF	RF$_{sub}$
General	0.80	0.67	0.75	0.37	0.10	0.78	0.99	0.64	1=25,946 0=94,389	1=25,946 0=25,946
C1	0.91	0.72	0.67	0.23	0.15	0.81	0.99	0.71	1=713 0=7,470	1=713 0=713
C2	0.84	0.72	0.63	0.35	0.15	0.75	0.98	0.72	1=6,844 0=31,771	1=6,844 0=6,844
C3	0.79	0.74	0.74	0.51	0.32	0.69	0.96	0.76	1=4,834 0=13,576	1=4,834 0=4,834
C4	0.76	0.70	0.69	0.48	0.27	0.76	0.95	0.67	1=1,005 0=25,208	1=1,005 0=1,005
C5	0.87	0.78	0.67	0.43	0.42	0.75	0.96	0.79	1=3,501 0=16,367	1=3,501 0=3,501

Where RF = Random Forest model without subsampling, RFsub = Random Forest model with subsampling, N training = number of observations in the training base, where 1 = presence of large nest and 0 = absence.

Table 6: Validation of the use of the general large nest model with subsampling for site-specific regions

Model	Model Statistics				
	Accuracy	Precision	Sensitivity	Specificity	N test
C1	0.88	0.42	0.64	0.91	1=192, 0=1,854
C2	0.76	0.39	0.70	0.78	1=1,663, 0=7,991
C3	0.60	0.39	0.86	0.51	1=1,236, 0=3,367
C4	0.56	0.38	0.92	0.41	1=2,506, 0=6,309
C5	0.82	0.48	0.74	0.83	1=853, 0=4,114
Mean	0.72	0.41	0.77	0.69	

Where N test = number of observations in the test base, where 1 = presence of large nest and 0 = absence.

accuracy of 72% across all clusters. However, its sensitivity was higher than that of the site-specific models only in clusters C3 and C4. In C5, both models performed similarly (general: 74%, site-specific: 75%). As observed in the initial nests model, the lower sensitivity in C1 and C2 was likely due to an imbalance in the test dataset, where absence of large nests accounted for 90% (C1) and 82% (C2) of the data.

Given that large nests pose an important potential risk for forest productivity, due to their high leaf consumption capacity (Viana *et al.*, 2004), this study prioritizes sensitivity over accuracy, accepting a higher Type I error rate (false positives). Under this approach, the general model with subsampling is suitable for classifying large nests in C3, C4, and C5, eliminating the need for site-specific models in

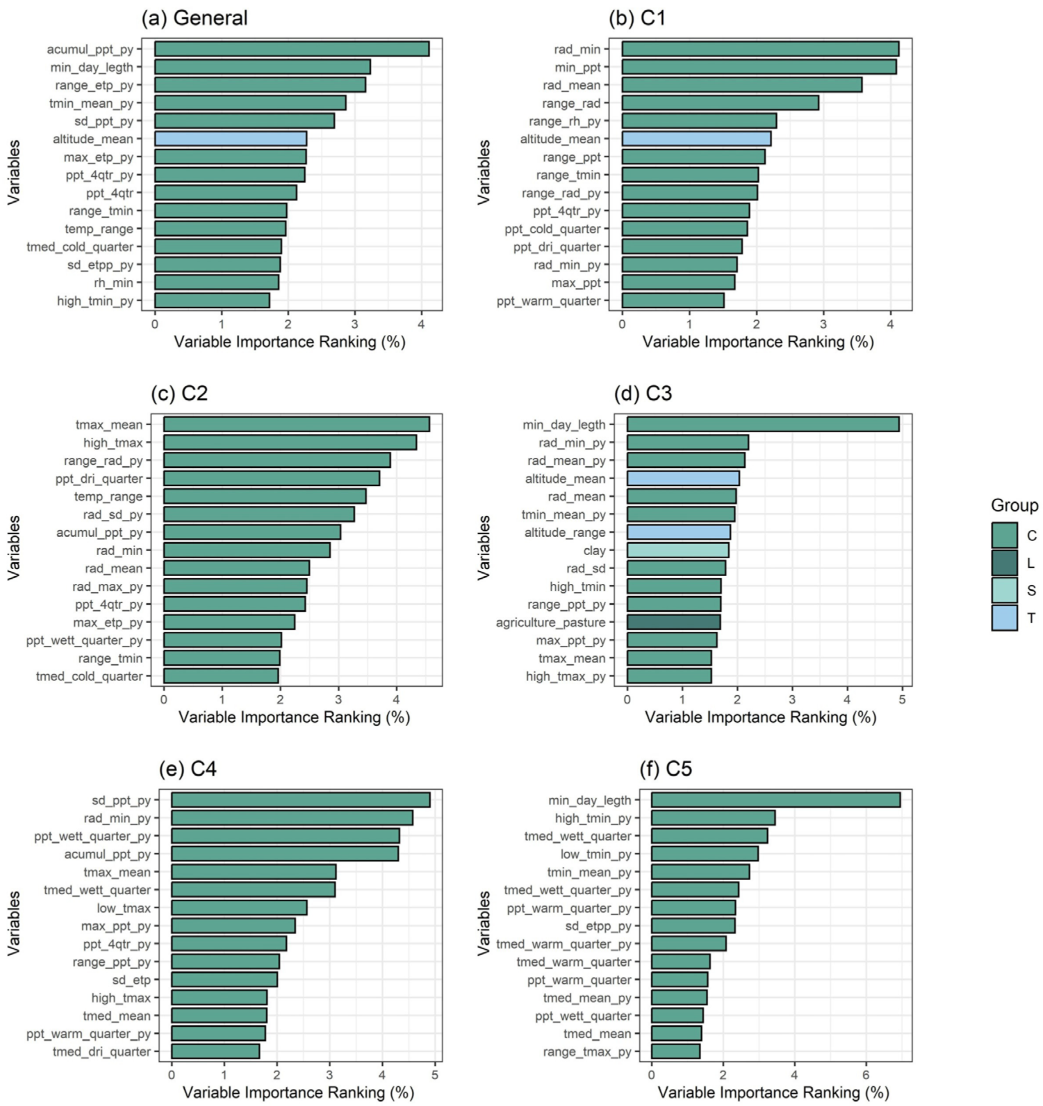

Figure 5: Variable importance (%) for initial nest models with subsampling across different clusters. Where C = climate, S = soil, T = terrain, and L = land use.

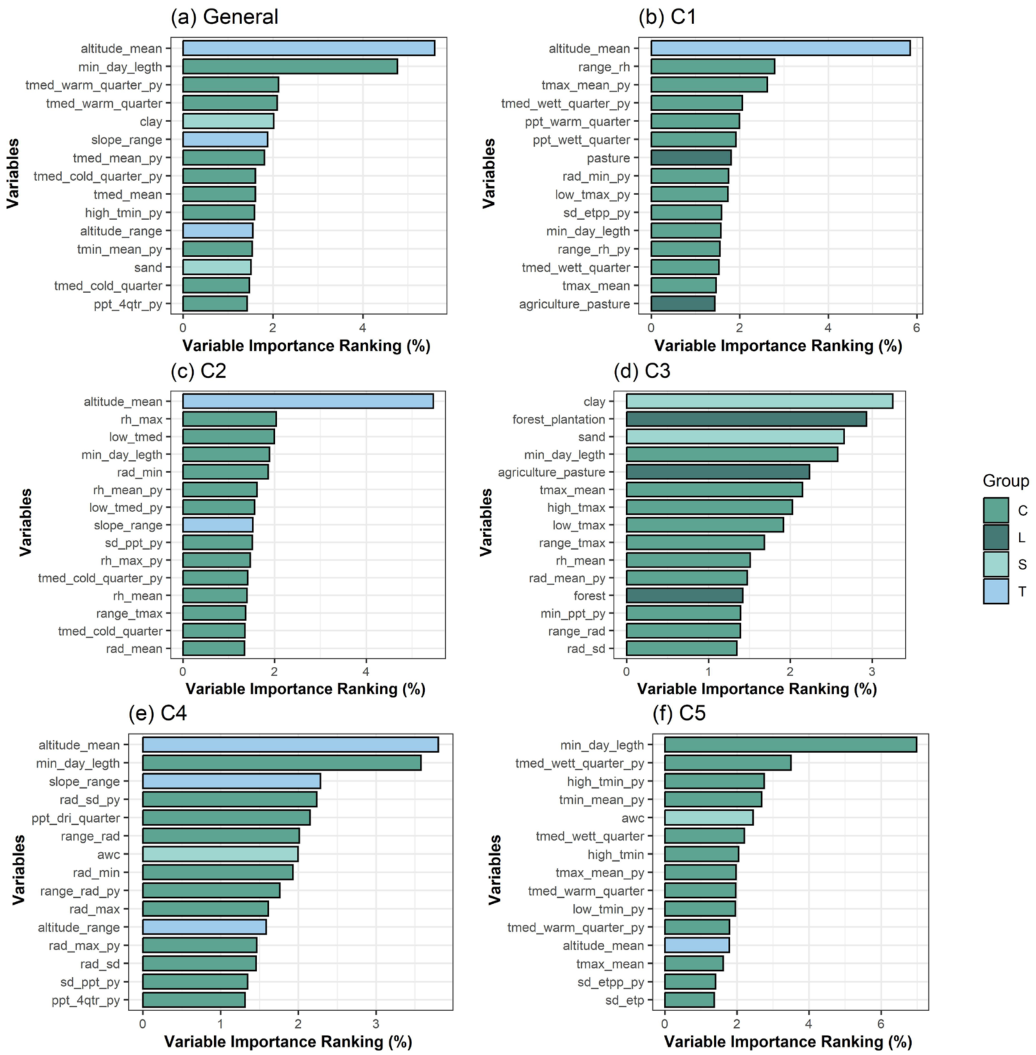

Figure 6: Variable importance (%) for large nest models with subsampling, across different clusters. Where C = climate, S = soil, T = terrain, and L = land use.

these regions. However, for C1 and C2, site-specific models are recommended to enhance sensitivity in detecting large nests. This approach is supported in ecological studies, where different data balancing techniques are used to produce more sensitivity and even more accurate models (Abdulwahab *et al.*, 2022;Benkendorf *et al.*, 2023; De Cubber *et al.*, 2023; Marchetto *et al.*, 2023).

Despite the strong influence of climate variables, large nest models also revealed significant terrain and soil effects. Altitude played a crucial role in all models except for C3, while soil variables, such as clay and sand percentages (for the general model and C3) and available water capacity (AWC) (for C4 and C5), also contributed to nest distribution. The influence of soil texture on the general model may be linked to the sandier nature of clusters C2, C3, and C4, which have historically shown a higher presence of large nests. Schoereder and Da Silva (2008) suggest that cohesive, clay-rich soils tend to limit nest size due to increased excavation difficulty.

Landscape factors such as forestry, agriculture, pasture, and forest formations were particularly influential in the C3 and C1 models. The presence of native vegetation adjacent to *Eucalyptus* plots has been negatively correlated with infestation (total loose soil per hectare) (Chiles *et al.*, 2022) and with the density of initial nests (<1 m^2) when in proximity to native vegetation strips (Zanetti *et al.*, 2000). However, Zanetti *et al.* (2000) also reported that plots near native forest fragments exhibited a higher density of initial nests, although no evidence was found indicating that these fragments influence the occurrence of large ant nests. Adjacent unmanaged *Eucalyptus* plots can serve as infestation sources, allowing unchecked nest expansion.

Regarding climatic influences, air temperature and humidity, modulated by altitude (a variable present in most models), directly impact leaf-cutting ant foraging (Viana *et al.*, 2004; Giesel *et al.*, 2013; Tizón *et al.*, 2014; Bustamante *et al.*, 2020). In the Cerrado, *Atta laevigata* exhibits a bimodal foraging pattern, ceasing activity on rainy days, which underscores the role of seasonal climate variation (Viana *et al.*, 2004). Changes in foraging behavior directly affect resource availability for colony survival and expansion.

The continuous use of leaf-cutting ant monitoring data is essential for improving predictive modeling and enhancing pest management strategies in commercial *Eucalyptus* plantations. By continuously integrating monitoring data into predictive frameworks, forest managers can enhance the efficiency of pest control operations, mitigate economic losses, and improve the long-term sustainability of *Eucalyptus* plantations.

It is important to note that this study was conducted in managed *Eucalyptus* forests, where the presence of large nests is also shaped by management interventions. Factors such as the frequency and effectiveness of ant control measures, economic constraints that may deprioritize pest management, and the expansion of plantations into previously unmanaged areas all contribute to colony dynamics. These operational factors underscore the need for continuous monitoring and adaptive management strategies to ensure the long-term effectiveness of leaf-cutting ant control.

In addition, physiographic conditions represent a key descriptor of the terrain that was not explicitly explored in this study. Stape and Alvares (2025) provided a comprehensive approach to the construction and use of water table depth (WTD) to relate its effects on the productivity of tropical planted *Eucalyptus* forests. For instance, WTD strongly influences soil type and moisture, which could in turn affect the ecological dynamics of leaf-cutting ants at the landscape scale within plantations. Future research should therefore investigate how within-stand physiographic variability interacts with establishment and persistence of ant colonies, providing an additional step forward in the modeling framework presented here.

CONCLUSIONS

This study highlights the environmental drivers of leaf-cutting ant establishments in *Eucalyptus* plantations, providing a quantitative basis for integrated pest management. Infestation rates varied from 39% to 90% across edaphoclimatic clusters, with large nests in up to 60% of some areas. Predictive models achieved 81%–89% accuracy but faced sensitivity limitations due to dataset imbalances, improved by subsampling techniques. The general model performed well (79% accuracy, 77% sensitivity), supporting large-scale pest monitoring. Soil texture, precipitation, and temperature fluctuations strongly influenced nest distribution, emphasizing the need for region-specific control strategies. Persistent large nests highlight the importance of early intervention. Machine learning models enhance infestation risk prediction, reducing chemical reliance and promoting sustainable management. Future research should integrate real-time climatic data and remote sensing for improved pest control efficiency.

ACKNOWLEDGMENTS

The authors would like to thank Suzano SA Company for providing monitoring and environmental datasets. We would like to thank FCA/UNESP for providing the necessary resources and environment during the Master's program of the first author, under the supervision of the second author. Acknowledgments are also due to CNPq for the financial support through the Inova Talentos project.

AUTHORSHIP CONTRIBUTION

Project Idea: CAA; EAVZ; RGM; VFS
Funding: CAA; EAVZ
Database: VSF; IRC; EPS; MMD; CAA
Processing: VSF; IRC; CAA
Analysis: VSF; IRC; MMD; CAA
Writing: VSF; CAA; MMD; EPS; IRC; EAVZ; RGM
Review: VSF; CAA

ACKNOWLEDGMENTS

The authors would like to thank Suzano S.A. company for providing the environmental and monitoring datasets. We also express our gratitude to FCA/UNESP for providing the necessary resources and infrastructure during the first author's Master's program, conducted under the supervision of the second author. Finally, we acknowledge CNPq for the financial support provided through the Inova Talentos project.

Data availability

The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

ABDULWAHAB, U. A.; HAMMILL, E.; HAWKINS, C. P. Choice of climate data affects the performance and interpretation of species distribution models. Ecological Modelling, v. 471, p. 110042, 2022. DOI: 10.1016/j.ecolmodel.2022.110042.

AIDOO, O. F.; SOUZA, P. G. C.; DA SILVA, R. S.; *et al.* A machine learning algorithm-based approach (MaxEnt) for predicting invasive potential of Trioza erytreae on a global scale. Ecological Informatics, v. 71, p. 101792, 2022. DOI: 10.1016/j.ecoinf.2022.101792.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; *et al.* Köppen's climate classification map for Brazil. Meteorologische Zeitschrift, v. 22, n. 6, p. 711–728, 2013.

ALVARES, C. A.; CEGATTA, Í. R.; SCOLFORO, H. F.; *et al.* Decision-tree application to predict and spatialize the wood productivity probabilities of Eucalyptus plantations. Forests, v. 14, n. 7, p. 1334, 2023. DOI: 10.3390/f14071334.

ARAÚJO, M. S.; DELLA LUCIA, T. M. C.; MAYHÉ-NUNES, A. J. Levantamento de Attini (Hymenoptera, Formicidae) em povoamento de eucalipto na região de Paraopeba, Minas Gerais, Brasil. Revista Brasileira de Zoologia, v. 14, n. 2, p. 323–328, 1997.

BENKENDORF, D. J.; SCHWARTZ, S. D.; CUTLER, D. R.; *et al*. Correcting for the effects of class imbalance improves the performance of machine-learning based species distribution models. Ecological Modelling, v. 483, p. 110414, 2023. DOI: 10.1016/j.ecolmodel.2023.110414.

BUSTAMANTE, S.; AMARILLO-SUÁREZ, A.; WIRTH, R. Effects of pasture and forest microclimatic conditions on the foraging activity of leaf-cutting ants. Biotropica, v. 52, n. 4, p. 697–708, 2020.

CAMARGO, R. S.; AMORIM FONSECA, J.; LOPES, J. F. S.; *et al*. Influência do ambiente no desenvolvimento de colônias iniciais de formigas cortadeiras (Atta sexdens rubropilosa). Ciência Rural, v. 43, n. 8, p. 1375–1380, 2013.

CANUTO, D. M. F.; DOS REIS, S. F.; GODOY, W. A. C. Spatial Statistics and Age Structure of Leaf Cutting Ant Nests. Open Journal of Statistics, v. 9, n. 2, p. 196–217, 2019. DOI: 10.4236/ojs.2019.92015.

CARDOSO, S. R. S.; FORTI, L. C.; NAGAMOTO, N. S.; *et al*. First-year nest growth in the leaf-cutting ants Atta bisphaerica and Atta sexdens rubropilosa. Sociobiology, v. 61, n. 3, p. 243–249, 2014.

CHEN, S.; DING, F.; HAO, M.; *et al*. Mapping the Potential Global Distribution of Red Imported Fire Ant (Solenopsis invicta Buren) Based on a Machine Learning Method. Sustainability, v. 12, n. 23, p. 10182, 2020.

CHILES, C. R.; MELO, R. S.; OTTO, M. S. G.; *et al*. How can structure and composition of Eucalyptus plantation landscape reduce leaf-cutting ants? Forest Ecology and Management, v. 518, p. 120250, 2022. DOI: 10.1016/j.foreco.2022.120250.

DE CUBBER, L.; TRENKEL, V. M.; DIEZ, G.; *et al*. Robust identification of potential habitats of a rare demersal species (blackspot seabream) in the Northeast Atlantic. Ecological Modelling, v. 477, p. 110255, 2023. DOI: 10.1016/j.ecolmodel.2022.110255.

DELABIE, J. H. C.; ALVES, H. S. R.; REUSS-STRENZEL, G. M.; *et al*. Distribuição das formigas-cortadeiras Acromyrmex e Atta no Novo Mundo. In: DELLA LUCIA, T. M. C. (Org.). Formigas-cortadeiras: da bioecologia ao manejo. Viçosa, MG: UFV, 2011. p. 80–101.

DELLA LUCIA, T. M. C.; SOUZA, D. J. Importância e história de vida das formigas cortadeiras. In: DELLA LUCIA, T. M. C. (Org.). Formigas-cortadeiras: da bioecologia ao manejo. Viçosa, MG: UFV, 2011. p. 1–421.

DE FRANÇA, L. C.; PEREIRA, P. S.; SARMENTO, R. A.; *et al*. Artificial neural networks as a tool for seasonal forecast of attack intensity of Spodoptera spp. in Bt soybean. International Journal of Biometeorology, v. 68, n. 11, p. 2387–2398, 2024. DOI: 10.1007/s00484-024-02747-w.

DE OLIVEIRA APARECIDO, L. E.; DE SOUZA ROLIM, G.; DA SILVA CABRAL DE MORAES, J. R.; *et al*. Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. International Journal of Biometeorology, v. 64, n. 4, p. 671–688, 2020.

DIDAY, E. Optimisation en classification automatique et reconnaissance des formes. Revue Française d'Automatique, Informatique, Recherche Opérationnelle, v. 6, p. 61–96, 1972.

FARR, T. G.; KOBRICK, M. Shuttle radar topography mission produces a wealth of data. Eos, v. 81, n. 48, p. 583–585, 2000.

GIESEL, A.; BOFF, M. I. C.; BOFF, P. Seasonal Activity and Foraging Preferences of the Leaf-Cutting Ant Atta sexdens piriventris (Santschi) (Hymenoptera: Formicidae). Neotropical Entomology, v. 42, n. 6, p. 552–557, 2013. DOI: 10.1007/s13744-013-0160-2.

GRANDEZA, L. A. O.; MORAES, J. C.; ZANETTI, R. Estimativa do crescimento externo de ninhos de Atta sexdens rubropilosa Forel e Atta laevigata (F. Smith) (Hymenoptera: Formicidae) em áreas de reflorestamento com eucalipto. Anais da Sociedade Entomológica do Brasil, v. 28, n. 1, p. 59–64, 1999.

HENTSCHEL, R.; MÖLLER, K.; WENNING, A.; *et al*. Importance of ecological variables in explaining population dynamics of three important pine pest insects. Frontiers in Plant Science, v. 9, p. 1667, 2018. DOI: 10.3389/fpls.2018.01667.

INDÚSTRIA BRASILEIRA DE ÁRVORES. Relatório Anual 2023. São Paulo: IBÁ, 2023.

INSTITUTO DE PESQUISAS E ESTUDOS FLORESTAIS. Derogation for the use of Deltamethrin, Fenitrothion, Fipronil and Sulfluramid in FSC® certified forests in Brazil. Piracicaba: IPEF, 2016.

LEE, W. H.; SONG, J. W.; YOON, S. H.; *et al*. Spatial Evaluation of Machine Learning-Based Species Distribution Models for Prediction of Invasive Ant Species Distribution. Applied Sciences, v. 12, n. 20, p. 10260, 2022.

LEI, Y.; ZHOU, Y.; LÜ, L.; *et al*. Rhythms in Foraging Behavior and Expression Patterns of the Foraging Gene in Solenopsis invicta (Hymenoptera: Formicidae) in relation to Photoperiod. Journal of Economic Entomology, v. 112, n. 6, p. 2923–2930, 2019.

MARCHETTO, E.; DA RE, D.; TORDONI, E.; *et al*. Testing the effect of sample prevalence and sampling methods on probability- and favourability-based SDMs. Ecological Modelling, v. 477, p. 110248, 2023. DOI: 10.1016/j.ecolmodel.2022.110248.

MARINHO, C. G. S.; OLIVEIRA, M. A.; ARAÚJO, M. S.; *et al*. Voo nupcial ou revoada de formigas-cortadeiras. In: DELLA LUCIA, T. M. C. (Org.). Formigas-cortadeiras: da bioecologia ao manejo. Viçosa, MG: UFV, 2011. p. 165–172.

MOREIRA, A. A.; FORTI, L. C.; ANDRADE, A. P. P.; *et al*. Nest Architecture of Atta laevigata (F. Smith, 1858) (Hymenoptera: Formicidae). Studies on Neotropical Fauna and Environment, v. 39, n. 2, p. 109–116, 2004. DOI: 10.1080/01650520412331333756.

PEDREGOSA, F.; MICHEL, V.; GRISEL, O.; *et al*. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011.

PYTHON SOFTWARE FOUNDATION. Python (version 3.11). Wilmington: Python Software Foundation, 2022.

SCHAEFER, C. E. G. R.; HENRIQUES, R. J.; GOMES, L. P.; *et al*. Interplays between Atta ants (Formicidae: Attini), soils and environmental properties in the Brazilian Neotropics: a preliminary assessment. Revista Brasileira de Ciência do Solo, v. 45, p. e0210073, 2021.

SCHOEREDER, J. H.; DA SILVA, W. L. Leaf-Cutting Ants (Hymenoptera: Formicidae) and Soil Classes: Preference, Survival and Nest Density. Sociobiology, v. 52, n. 2, p. 403–415, 2008.

SOUSA, K. K. A.; CAMARGO, R. S.; CALDATO, N.; *et al*. The ideal habitat for leaf-cutting ant queens to build their nests. Scientific Reports, v. 12, n. 1, p. 4830, 2022. DOI: 10.1038/s41598-022-08918-2.

SOUZA, A.; ZANETTI, R.; CALEGARIO, N. Nível de Dano Econômico para Formigas-Cortadeiras em Função do Índice de Produtividade Florestal de Eucaliptais em uma Região de Mata Atlântica. Neotropical Entomology, v. 40, n. 4, p. 483–488, 2011.

SOUZA, C. M.; SHIMBO, J. Z.; ROSA, M. R.; *et al*. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. Remote Sensing, v. 12, n. 17, p. 2735, 2020.

SPARKS, A. H.; MIGUEZ, F.; SALMON, M.; *et al*. Nasapower: NASA-POWER Data from R. R package version 4.0.7.

STAAB, M.; KLEINEIDAM, C. J. Initiation of swarming behavior and synchronization of mating flights in the leaf-cutting ant Atta vollenweideri Forel, 1893 (Hymenoptera: Formicidae). Myrmecological News, v. 19, p. 93–102, 2014.

STACKHOUSE, P. W.; ZHANG, T.; WESTBERG, D.; *et al*. POWER Release 8.0.1 (with GIS applications) methodology: data parameters, sources, & validation. Hampton: NASA, 2018. 99 p.

STAPE, J. L.; ALVARES, C. A. Water table depth contributes to tropical Eucalyptus plantation yields in sandstone-derived landscapes. Forest Ecology and Management, v. 589, p. 122771, 2025.

TIZÓN, R.; WULFF, J. P.; PELÁEZ, D. V. The effect of increase in the temperature on the foraging of Acromyrmex lobicornis (Hymenoptera: Formicidae). Zoological Studies, v. 53, p. 1–13, 2014. DOI: 10.1186/s40555-014-0071-6.

VASQUES, G. M.; COELHO, M. R.; DART, R. O.; *et al*. Soil Clay, Silt and Sand Content Maps for Brazil at 0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm Depth Intervals with 90 m Spatial Resolution. Rio de Janeiro: Embrapa Solos, 2021.

VIANA, L. R.; SANTOS, J. C.; ARRUDA, L. J.; *et al*. Foraging Patterns of the Leaf-Cutter Ant Atta laevigata (Smith) (Myrmicinae: Attini) in an Area of Cerrado Vegetation. Neotropical Entomology, v. 33, n. 3, p. 391–393, 2004. DOI: 10.1590/S1519-566X2004000300018.

XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in Brazil (1980–2013). International Journal of Climatology, v. 36, n. 6, p. 2644–2659, 2016.

ZANETTI, R.; VILELA, E. F.; ZANUNCIO, J. C.; *et al*. Influência da espécie cultivada e da vegetação nativa circundante na densidade de sauveiros em eucaliptais. Pesquisa Agropecuária Brasileira, v. 35, n. 10, p. 1911–1918, 2000. DOI: 10.1590/S0100-204X2000001000001