Research Article
doi: 10.1590/01047760202632013689
vol(32), 2026

CERNE

Dependence between lianas and tree health in minas gerais: ridge-penalized logistic and multinomial copula models

Isadora Azevedo Perpétuo[1]*, Marcelo Vitor Gualberto Santos Chaves[1], Elias Manensa Sabe[2], Rafael Genaro[2], Marcelo Ângelo Cirillo[2], Samuel José Silva Soares da Rocha[1], Luis Marcelo Tavares de Carvalho[1].

[1]Federal University of Lavras, Department of Forest Sciences, Lavras, MG, Brazil.
[2] Federal University of Lavras, Department of Statistics, Lavras, MG, Brazil.

FOREST ECOLOGY

ABSTRACT

Background: This study presents an integrated statistical framework for the joint modeling of mixed ecological variables, simultaneously addressing issues of multicollinearity and dependence between response variables. Data from the National Forest Inventory of Minas Gerais were used to model liana presence (⟦*LP*⟧*_bin*), a binary variable, and tree health status (TH), an ordinal categorical variable with four levels.

Results: The marginal distribution of ⟦*LP*⟧*_bin* was fitted using logistic regression with the penalized AURPLCE estimator, which showed satisfactory predictive performance (AUC = 0.771). For the TH variable, multinomial logistic regression was applied, and an additional ordinal regression analysis including ⟦*LP*⟧*_bin* as a predictor of TH was also performed. Dependence between the responses was modeled using copulas, with the Clayton copula selected based on the Akaike Information Criterion (AIC = −72.0142). Estimates of Kendall's coefficient (τ = 0.1109) and Spearman's rank correlation (ρ_s = 0.1707) indicated a positive but weak dependence. The joint structure, represented by a probability table and heat map, revealed co-occurrence patterns, highlighting a higher probability of healthy trees in the absence of lianas (estimated probability of 0.3655).

Conclusions: Overall, the results underscore the potential of combining penalized estimators and copula-based models for the robust analysis of forest data, providing relevant quantitative insights into the relationship between liana presence and tree phytosanitary condition.

Keywords: *Joint modeling; multicollinearity; ecological analyses; forest health.*

HIGHLIGHTS

The article integrates a joint model of mixed ecological variables with high robustness.
Lianatre presence and tree health were modeled using penalized logistic and multinomial logistic regressions.
Dependence between response variables was modeled with Clayton copulas, indicating a weak relationship.
Healthy trees are more likely to occur in the absence of lianas.

PERPÉTUO, I. A.; CHAVES, M. V. G. S.; SABE, E. M.; GENARO, R.; CIRILLO, M. A.; ROCHA, S. J. S. S.; CARVALHO, L. M. T. Dependence between lianas and tree health in minas gerais: ridge-penalized logistic and multinomial copula models. CERNE, v. 32, e103689, 2026. DOI: 10.1590/01047760202632013689

✤Corresponding author: Isadora.perpetuo2@estudante.ufla.br
Scientific Editor: Camila Lais Farrapo

Received: January 14, 2026
Accepted: June 01, 2026

Open Access
CERNE
ISSN 0104-7760
UFLA UNIVERSIDADE FEDERAL DE LAVRAS

INTRODUCTION

Understanding the interactions between tree health and the presence of lianas is essential for the management and conservation of forest ecosystems, particularly in regions of high biodiversity. Trees in good phytosanitary condition constitute the structural and functional foundation of forests, ensuring the maintenance of essential ecological processes such as nutrient cycling, carbon sequestration, successional dynamics, and microclimatic regulation (Malhi *et al*., 1999; Perry *et al*., 2008; Pan *et al*., 2011; Li *et al*., 2023).

In this context, although lianas are natural components inherent to forest structure, complexity, diversity, and dynamics, they can act as important agents of competitive pressure (Estrada-Villegas *et al*., 2022; Vargas *et al*., 2020; Vargas *et al*., 2022). Their presence directly affects light availability, radial and height growth, mechanical stability, and, in more severe cases, tree survival (Guedes and Krupek, 2017; Gonzalez de tanago *et al*., 2017; Han and Sánchez-Azofeifa, 2022). Recent pantropical assessments further suggest that lianas play a significant role in regulating forest carbon dynamics, potentially reducing aboveground biomass accumulation and long-term carbon sequestration. These findings highlight the ecological importance of monitoring liana abundance as an indicator of forest structural changes and ecosystem resilience under global environmental change (Rueda-Trujillo *et al*., 2025; Gora *et al*., 2023).

The National Forest Inventory (NFI) provides a comprehensive database containing detailed information on tree and environmental attributes, collected between 2011 and 2024 across different Brazilian biomes (Brazilian Forest Service, 2024). In this study, NFI data from the state of Minas Gerais were analyzed to investigate the association between liana presence (LP) and tree health status (TH), the latter categorized into four distinct levels of vitality. A study conducted by Estrada-Villegas and Schnitzer (2018) indicated that the presence of lianas may directly reflect the phytosanitary condition of trees in a given area, reducing the growth, survival, fecundity, and physiological performance of individuals, thus serving as an indicator of forest conservation status (Guedes & Krupek, 2017).

One of the main statistical challenges in this type of analysis lies in the presence of multicollinearity among the explanatory variables. This condition can inflate standard errors and compromise the precision of estimates obtained via maximum likelihood estimation (MLE). To mitigate this issue in modeling the binary variable ⟦PL⟧_bin, the Almost Unbiased Ridge Principal Component Logistic Estimator (AURPLCE), proposed by Wu and Asar (2018), was adopted. This estimator combines ridge regularization with dimensionality reduction through principal component analysis.

For the ordinal response variable TH, with four levels, a multinomial logistic regression model was fitted using maximum likelihood estimation to estimate the probabilities associated with each health status category as a function of structural and environmental covariates. Since TH presents ordered categories, a complementary cumulative logit model (ordinal regression) was also fitted to directly account for the health gradient and provide more parsimonious and interpretable estimates of predictor effects.

Additionally, to model the joint dependence between ⟦PL⟧_bin and TH, bivariate copula models specifically designed for discrete marginals were employed, within the theoretical framework proposed by Panagiotelis, Czado, and Joe (2012). This approach enables the integration of marginals of different natures into a flexible joint model, overcoming limitations of classical multivariate models, which often require restrictive parametric assumptions for specifying the dependence structure.

MATERIAL AND METHODS

Data description

The data analyzed in this study were obtained from the National Forest Inventory (NFI), conducted by the Brazilian Forest Service, and refer to the state of Minas Gerais, covering the period from 2011 to 2024. The database includes detailed dendrometric, phytosanitary, and spatial information from trees sampled in 1,042 plots distributed across the Caatinga and Cerrado biomes, systematically allocated throughout the state, totaling a sampled area of 104.2 ha. After filtering the data corresponding to these biomes and selecting the variables of interest, the initial sample comprised 3,805 valid observations. During the data preparation stage, 345 observations identified as outliers were removed, resulting in a final dataset of 3,460 trees used in the analyses conducted in this study.

The response variable is tree health status (TH), recorded on an ordinal scale with four categories: 1 (healthy tree), 2 (mild signs of impairment), 3 (moderate impairment), and 4 (dead tree). The multinomial structure was preserved in the modeling in order to capture relevant nuances in the degradation of tree health. The observed frequencies in the sample were 1,993 trees with TH = 1, 1,279 with TH = 2, 215 with TH = 3, and 318 with TH = 4.

Liana presence (LP) was originally recorded as a categorical variable with the levels Yes (Y) and No (N), and later recoded as a binary variable for statistical analyses, where LP = 1 indicated presence and LP = 0 absence. Among the sampled trees, 1,304 exhibited lianas and 2,501 did not.

Additional tree attributes and spatial variables were included in the analyses, including total height (THE), bole height (BH), diameter at breast height (DBH), bole quality (BQ), latitude, and longitude. Continuous variables were standardized prior to modeling and selected based on their ecological relevance and association with the main study variables. The inclusion of tree-level attributes was supported by evidence that liana infestation may reduce resource availability and light interception in host trees, negatively affecting growth, structural stability, and overall performance (Álvarez-Cansino *et al*., 2015; Estrada-Villegas *et al*., 2022).

Therefore, dendrometric traits such as height, diameter, and bole quality may reflect the effects of liana occurrence and infestation intensity. Latitude and longitude

were included as spatial proxies to represent geographic variation and large-scale environmental gradients across the study area. Previous studies have shown that liana abundance is associated with spatially structured climatic conditions, particularly in regions with low mean annual precipitation and high precipitation seasonality (Dewalt *et al*., 2015; Parolari *et al*., 2020; Estrada-Villegas *et al*., 2022). Thus, geographic coordinates were incorporated to capture part of the spatial heterogeneity potentially related to liana occurrence and tree performance throughout Minas Gerais.

Identification and Removal of Outliers

Three statistical criteria were used to identify influential observations: Pearson residuals, leverage (hat values), and Cook's distance (Barreto, 2011). Pearson residuals were calculated as:

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i\left(1-\hat{\pi}_i\right)}}$$

with absolute values > 2 indicating potential outliers. Leverage was computed as $x_i^T\left(X^T W X^{-1}\right)x_i$, with observations exceeding $2p/n$ classified as influential. Cook's distance was given by $D_i = \frac{r_i^2 h_{ii}}{p(1-h_{ii})^2}$, with D_A>1 indicating significant influence. Observations exceeding any of these thresholds were removed to avoid bias and ensure robustness.

AURPLCE Estimator

To estimate coefficients under high multicollinearity, the Almost Unbiased Ridge-Type Principal Component Estimator (AURPLCE) proposed by Wu and Asar (2018) was employed. This estimator combines ridge regularization with dimensionality reduction via principal components. The methodology begins with the spectral decomposition of the logistic information matrix $S = X^T W X - T\Lambda T^T$, where T is the eigenvector matrix and the diagonal matrix of eigenvalues. The first r eigenvectors T_r explain the largest proportion of the variance (commonly > 70%). The AURPLCE estimator is given by equation 1:

$$\hat{\beta}_{AURPLSLE} = T_r T_r^T\left[I - k^2(S+kI)^{-1}(S+kI)^{-1}\right]\hat{\beta}_{MLE} \quad (1)$$

where k is the ridge penalization parameter and MLE is the maximum likelihood estimate (Wu and Asar, 2018).

ROC Curve and Area Under the Curve (AUC)

The Receiver Operating Characteristic (ROC) curve evaluates binary classifier performance by plotting the true positive rate against the false positive rate across decision thresholds. The Area Under the ROC Curve (AUC) quantifies discriminative ability and corresponds to the probability that the model assigns a higher score to a positive observation than to a negative one. An unbiased estimate of the AUC can be expressed as a Wilcoxon–Mann–Whitney statistic (Hanley and McNeil, 1982; Bamber, 1975), equation 2:

$$AUC\left(f\right) = \frac{1}{|D^0||D^1|}\sum_{t_0\in D^0}\sum_{t_1\in D^1} 1\left[f\left(t_0\right) < f\left(t_1\right)\right] \quad (2)$$

where D0 and D1 are the sets of negative and positive examples, ft is the predictive score, and 1 is the indicator function. Maximization of the AUC was adopted as the criterion for selecting the optimal r,k pairs of the AURPLCE estimator. The best performance was achieved with r=7 principal components and k∈0.1,1.5, yielding an AUC of 0.7713.

Marginal Modeling of Tree Health Status (TH)

Tree health status (TH) is an ordinal categorical variable with four levels (1: healthy, 2: mild signs, 3: moderate impairment, 4: dead). Two complementary approaches were adopted. First, a *multinomial logistic regression* was fitted to allow category-specific effects of covariates. The probability that observation *i* belongs to category *k* (with *k* = 1 as reference) is equation 3 and 4:

$$P\left(Y_i = k \middle| x_i\right) = \frac{\exp\left(x_i^T\beta_k\right)}{1+\sum_{j=2}^{4}\exp\left(x_i^T\beta_j\right)},\ k = 2,3,4 \quad (3)$$

$$P\left(Y_i = 1 \middle| x_i\right) = \frac{1}{1+\sum_{j=2}^{4}\exp\left(x_i^T\beta_j\right)} \quad (4)$$

where xi includes total height (THE), bole height (BH), diameter at breast height (DBH), latitude, longitude, bole quality (BQ), and liana presence (LP_bin). Estimation was performed via maximum likelihood using the multinom function in R (Goeman, 2016; Hosmer Jr et al., 2013). Second, a cumulative logit model (ordinal regression) was fitted to directly account for the ordered nature of TH and provide interpretable directional effects. The model assumes equation 5:

$$logit\left(P(x)\right) = a_k - x^T\beta, k = 1,2,3 \quad (5)$$

Where k are threshold parameters and β is a vector of coefficients. A positive coefficient indicates association with worse tree health. The proportional odds assumption was assessed and found reasonable. Estimation was performed using the clm function in R (Christensen, 2022). The model specification for both approaches was $TH \sim PL_{bin} + BH + THE + DBH + longitude + latitude + BQ$. The predicted probabilities from both models were used to construct pseudo-observations for joint copula modeling following Panagiotelis et al. (2012).

Assessment of Coefficient Significance: Wald Chi-Square Test

The Wald test (Barreto, 2011) was applied to assess the statistical significance of coefficients in the marginal models. The Wald statistic for parameter jis given by equation 6:

$$Z_j = \frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)},\quad SE\left(\hat{\beta}_j\right) = \sqrt{[S^{-1}]j_j} \quad (6)$$

where $\hat{\beta}_j$ is the coefficient estimate and S is the information matrix. Under H_0:β_j=0, Z_j follows a standard normal distribution, and the p-value is calculated as $p_j = 2(1 - \Phi(|z_j|))$, with $\Phi(.)$ the standard normal CDF. The test was applied to each coefficient of the marginal models for LP_bin and TH_bin, using the optimal r and k values selected by AUC. Coefficients with p<0.05 were considered statistically significant.

Dependence Modeling Using Copulas

According to Sklar's Theorem (Sklar, 1959), any joint distribution can be expressed as $F(y_1,y_2)=C(F_1(y_1),F_2(y_2);\alpha)$, where $C(;.;\alpha)$ is a copula function. Table 1 presents the four bivariate copula families considered: Gaussian, Clayton, Gumbel, and Frank.

Table 1: Bivariate Copula Functions Used.

Cópula	Função C(u_i,u_2;α)
Gaussian	$\Phi_2\left(\Phi^{-1}\left(u_1\right),\ \Phi^{-1}\left(u_2\right);\rho\right)$
Clayton	$ma\left\{\left(u_1{}^{-\alpha}+u_2{}^{-\alpha}-1\right)^{-\frac{1}{\alpha}},0\right\}$
Gumbel	$ex\left\{-[(-log\ \ u_1)^{\alpha}+(-lou_2)^{\alpha}]^{\frac{1}{\alpha}}\right\}$
Frank	$-\frac{1}{\alpha}lo\left(1+\frac{(e^{-\alpha u_1}-1)(e^{-\alpha u_2}-1)}{e^{-\alpha}-1}\right)$

To model dependence between LP_bin (binary) and TH (ordinal), we followed the framework for discrete data proposed by Panagiotelis *et al*. (2012). Pseudo-observations on the uniform scale 0,1 were obtained from the predicted probabilities of the marginal models (Lin and Chaganty, 2021). For LP_bin, pseudo-observations were derived directly from Px; for TH, they were obtained from the predicted probabilities for each category using the Monte Carlo method (Panagiotelis *et al*., 2012). After fitting the copula to the pseudo-observations, the joint probability for discrete variables is given by the finite difference of the cumulative copula (Panagiotelis *et al*., 2012), equation 7:

$$P(Y_1 = y_1, Y_2 = y_2) = C(F_1(y_1), F_2(y_2)) - C(F_1(y_1 - 1), F_2(y_2)) \\ -C(F_1(y_1), F_2(y_2 - 1)) + C(F_1(y_1 - 1), F_2(y_2 - 1)) \quad (7)$$

Copula selection was based on the Akaike Information Criterion (AIC; Akaike, 1974), defined as AIC= -2log L +2p. The Clayton copula yielded the lowest AIC value and was selected as the most appropriate for representing the dependence between LP_bin and TH.

Measures of Concordance: Kendall's and Spearman's coefficients

Nonparametric concordance measures, invariant to monotonic transformations of the marginals, are particularly appropriate in the copula context (Nelsen, 2006). Kendall's τ and Spearman's ρ_s were used to characterize the dependence between LP_bin and TH. Kendall's τ quantifies the degree of concordance between two continuous variables. For two i.i.d. pairs (X_1, Y_1) and (X_2,Y_2) equation 8:

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (8)$$

In terms of a copula C(u,v), Kendall's is given by equation 9:

$$\tau = 4\int_{[0,1]^2} C\left(u,v\right)\mathrm{d}C\left(u,v\right) - 1 \quad (9)$$

Spearman's s measures overall monotonic association and is defined as the Pearson correlation of the ranks. It can be expressed as equation 10:

$$\rho_s = 12E\left[F_X\left(X\right)F_Y\left(Y\right)\right] - 3 = 12\int_{[0,1]^2} C\left(u,v\right)\mathrm{d}u\mathrm{d}v - 3 \quad (10)$$

Both coefficients range from -1 to 1, where positive values indicate direct association, negative values inverse association, and values near zero indicate no monotonic relationship (Nelsen, 2006).

RESULTS

Results of the Marginal Modeling of PLbin

The presence of lianas (LP_{bin}) was modeled using the Almost Unbiased Ridge-Type Principal Component Estimator (AURPLCE), with optimal parameters of r=7 principal components and ridge penalization k=0.8.

Prior to the final model fitting, a diagnostic step was conducted to identify and remove influential observations and outliers. Pearson residuals, leverage (hat) values, and Cook's distance were evaluated for this purpose. Figure 1 visually illustrates the distribution of these metrics, allowing the identification of discrepant observations that were subsequently excluded from the analysis to ensure the robustness of the estimates. After outlier identification based on Pearson residuals, leverage (hat values), and Cook's distance, 345 observations (9.1%) were excluded from the initial sample of 3,805 trees, resulting in a final sample of 3,460 trees used in all subsequent analyses.

After outlier removal, the estimated AURPLCE coefficients for the PL_{bin} model and Wald tests are presented in Table 2. BH, THE, latitude, longitude, and BQ were statistically significant (p < 0.05), while DBH was not. The AUC was 0.771, indicating satisfactory discriminative ability for liana presence (Hanley and McNeil, 1982). The overall classification accuracy of the AURPLCE model for PL_{bin} was 70.6% using a cut-off of 0.5. Figure 2 presents the ROC curve of the fitted PL_{bin} model.

Results of the Marginal Modeling of Tree Health Status (TH)

The modeling of tree health (TH), an ordinal categorical variable with four levels, was performed using multinomial logistic regression, following the methodology detailed in Section 2.5. The results of the model fitting,

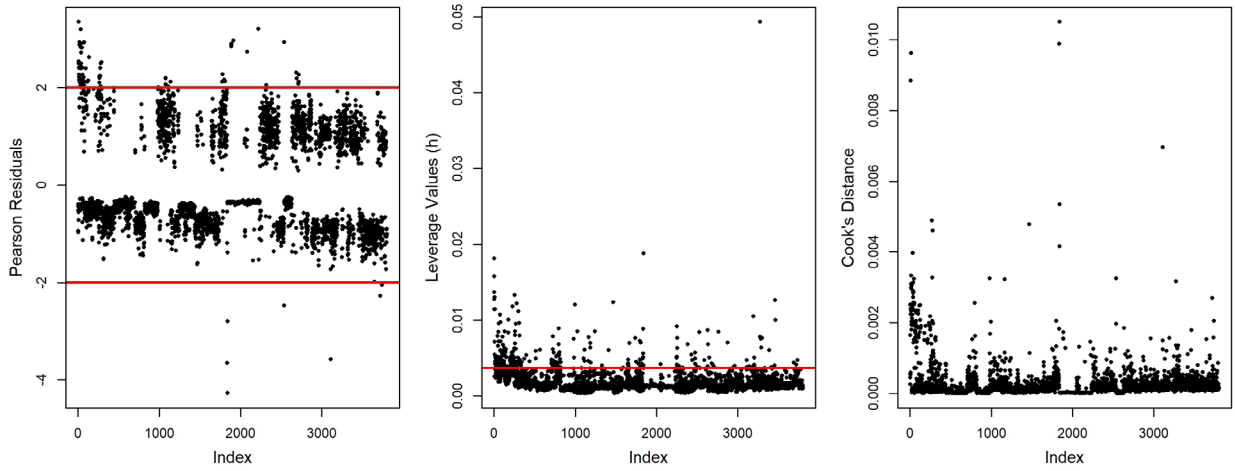

Figure 1: Outlier diagnostics in the PLbin modeling, showing Pearson residuals, leverage (hat values), and Cook's distance.

including the estimated coefficients and Wald test p-values for each covariate relative to the response variable categories (with TH = 1 as the reference category), are presented in Table 3. To complement the multinomial approach and directly account for the ordinal nature of TH, a cumulative logit model (ordinal regression) was also fitted, including liana presence (LP_bin) as a predictor. The ordinal model confirmed that LP_bin is significantly associated with worse tree health (coefficient = 0.2882, OR = 1.334, p = 0.0005). The overall accuracy of the ordinal model was 63.27%, with class-specific accuracies of 77.2% (TH=1), 54.3% (TH=2), 0% (TH=3), and 53.5% (TH=4); the poor performance for TH=3 reflects its low sample frequency (5.5%). The AIC values for the multinomial (5540.5) and ordinal (5632.1) models showed a difference of 91.6, indicating a better fit for the multinomial model. Nevertheless, the ordinal model remains useful as a complementary approach, providing a more parsimonious interpretation of directional effects along the tree health gradient.

Table 2: Estimated AURPLCE Coefficients and Wald Test Results for PL_{bin}

Variable	Coef.	Standard Error	Z	p-value
Intercept	-0.9502	0.0451	-21.06	< 0.0001
BH	-2.3019	0.1479	-15.56	< 0.0001
THE	1.6476	0.1551	10.63	< 0.0001
DBH	0.0498	0.0458	1.09	0.2761
Longitude	-0.3221	0.0485	-6.64	< 0.0001
Latitude	-0.6943	0.0569	-12.20	< 0.0001
TQH	0.1820	0.0533	3.41	0.0006

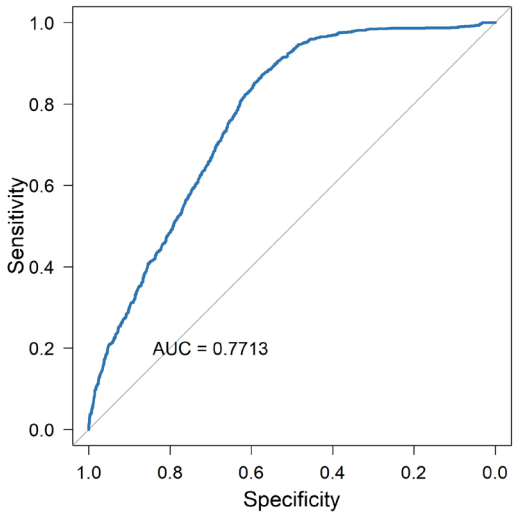

Figure 2: ROC Curve for the AURPLCE Model of the PLbin Variable

Table 3: Estimated Coefficients and Wald Test Results for the Multinomial Logistic Regression Model of TH

TH Category	Variable	Coef.	Standard Error	Z-Value	p-value
TH=2 vs TH=1	Intercept	-0.370	0.043	-8.665	< 0.0001
	BH	0.155	0.124	1.249	0.2115
	THE	-0.861	0.126	-6.844	< 0.0001
	DBH	0.192	0.047	4.080	< 0.0001
	Longitude	-0.340	0.048	-7.073	< 0.0001
	Latitude	0.204	0.045	4.553	< 0.0001
	TQH	0.753	0.060	12.494	< 0.0001
TH=3 vs TH=1	Intercept	-2.586	0.116	-22.246	< 0.0001
	BH	0.731	0.275	2.665	0.0077
	THE	-1.665	0.263	-6.329	< 0.0001
	DBH	0.419	0.076	5.513	< 0.0001
	Longitude	-0.378	0.083	-4.579	< 0.0001
	Latitude	0.475	0.081	5.840	< 0.0001
	TQH	1.362	0.108	12.585	< 0.0001
TH=4 vs TH=1	Intercept	-3.272	0.140	-23.298	< 0.0001
	BH	2.754	0.305	9.024	< 0.0001
	THE	-2.568	0.300	-8.560	< 0.0001
	DBH	0.530	0.080	6.662	< 0.0001
	Longitude	-0.313	0.087	-3.620	0.0003
	Latitude	0.822	0.086	9.547	< 0.0001
	TQH	2.602	0.120	21.627	< 0.0001

Results of the Copula-Based Modeling

After fitting the marginal models, pseudo-observations were calculated from the predicted probabilities and used to fit four bivariate copulas (Gaussian, Clayton, Gumbel, Frank) via maximum likelihood (Panagiotelis *et al*., 2012). Model comparison based on AIC (Akaike, 1974) is presented in Table 4. The Clayton copula yielded the lowest AIC (-72.0142), with estimated parameter θ = 0.2493 and Kendall's τ = 0.1109.

Table 4: AIC Values for the Fitted Copula Models

Copula	AIC
Gaussian	-38.3501
Clayton	-72.0142
Gumbel	-7.0485
Frank	-33.4095

The final model consisted of the fitted marginal distributions for the variables PLbin and TH , combined through a Clayton copula. The joint distribution function for a Clayton copula with parameter is given by equation 11:

$$C\left(u,v\right)=\left(u^{-\theta}+v^{-\theta}-1\right)^{\frac{1}{\theta}} \quad (11)$$

where u and v are the pseudo-observations of the marginal variables. Applying the estimated parameter = 0.2493: (Equation 12)

$$F_{Y_1Y_2}\left(y_1,y_2\right)=\left[F_{Y_1}(y_1)^{-0.2943}+F_{Y_2}(y_2)^{-0.2943}-1\right]^{-\frac{1}{0.2943}} \quad (12)$$

The marginal cumulative distribution functions FY1 and FY2 were estimated using the AURPLCE model for PLbin and multinomial regression for TH, respectively, as described in the previous sections.

Table 5 presents the estimated joint probabilities for all combinations of PLbin and TH categories, calculated from the fitted Clayton copula.

Table 5: Estimated Joint Probability Table (Clayton Copula)

TH	LP=0	LP=1
TH=1	0.3655	0.1583
TH=2	0.2081	0.1280
TH=3	0.0339	0.0226
TH=4	0.0497	0.0338

To quantify the strength of the dependence between liana presence (LP) and tree health status (TH), Kendall's tau and Spearman's rho were estimated from the pseudo-observations derived from the fitted marginals. The Kendall concordance coefficient, calculated from the estimated parameter of the Clayton copula, was (Equation 13 and 14):

τ=0.1109 (13)

Additionally, Spearman's rho was estimated directly from the rank-normalized pseudo-observations, yielding:

ρ_S=0.1707 (14)

Both estimates indicate the presence of a positive dependence between the variables analyzed.

DISCUSSION

The joint modeling of liana presence (PLbin) and tree health status (TH) using a marginal-copula approach provided important insights into the interrelationship between these factors in forest ecosystems. This section delves into the interpretation of the results obtained from the marginal models, the copula selection, and the estimated joint probabilities.

Discussion of the Estimated Parameters from the Marginal Models

Marginal Model for Liana Presence (PL_{bin})

The results of the logistic regression model for PL_{bin}, fitted using the AURPLCE estimator (Table 2), indicate that several environmental and dendrometric covariates significantly influence the probability of liana occurrence. The negative and significant coefficient for bole height (BH) and the positive coefficient for total height (THE) suggest a complex relationship with liana presence. Trees with longer boles may be associated with a lower probability of hosting lianas, whereas taller trees overall may have a higher probability. This pattern may indicate that lianas establish more easily on trees with shorter boles or that their growth is favored in trees that have already reached greater total height, since they use trees with self-supporting wood and apical dominance as structural support to reach and remain in the canopy, where they begin to compete intensely for aboveground and belowground resources such as light, water, and nutrients (Schnitzer; Bongers, 2011; Vargas *et al*., 2022).

Diameter at breast height (DBH) was not statistically significant (p=0.2761) in the marginal model for PL_{bin}, suggesting that, independent of the other covariates, tree diameter is not a strong predictor of liana presence. The geographic covariates longitude and latitude showed negative and significant coefficients, indicating a strong influence of geographic location on liana distribution. This pattern likely reflects climatic, edaphic, and land-use variation across the study area, as liana occurrence is generally favored in warm, humid environments with fertile soils. In addition, topographic and edaphic heterogeneity play an important role in species turnover and ecological interactions (Torres *et al*., 2023), creating favorable conditions for the establishment and persistence of lianas (Tuomisto *et al*., 2017; Gerolamo *et al*., 2022).

It is important to note that the analyzed data present a hierarchical structure, with multiple trees sampled within each plot. The latitude and longitude coordinates used correspond to the centroids of the plots rather than to the individual locations of the trees. Although the approach adopted in this study does not incorporate random effects to model the dependence among trees within the same plot, this simplification is recognized as a methodological limitation. Therefore, future studies may employ mixed models or copula models with random effects in order to more adequately represent the hierarchical structure and spatial dependence present in the data.

According to Engel, Fonseca, and Oliveira (1998), higher liana density across all diameter classes was observed in alluvial soils rich in organic matter, with high cation exchange capacity (CEC) and slightly acidic pH, compared to more acidic soils with low CEC. This finding highlights the high ecological plasticity of lianas and their ability to exploit fertile environments, suggesting that soil quality and nutrient availability directly influence liana abundance. Bole quality (BQ) also showed a positive and significant coefficient, indicating that trees with better-quality boles may have a greater probability of liana presence. This relationship may indirectly reflect more productive environments that favor both tree development and liana growth. As noted by Proctor *et al*. (1983), although tree density and basal area may remain relatively stable among soil types, lianas can benefit from favorable edaphic conditions to intensify competition for resources. Consequently, high liana abundance in fertile soils may increase pressure on trees, affecting forest dynamics, successional processes, and ecosystem resilience.

The model's high predictive capacity for PLbin, evidenced by an AUC of 0.771 (Figure 2), reinforces the adequacy of the selected covariates and the effectiveness of the AURPLCE estimator in handling multicollinearity, providing robust estimates and a model with strong discriminative performance.

Marginal Model for Tree Health Status (TH)

The results of the multinomial logistic regression model for TH (Table 3), with TH = 1 (healthy) as the reference category, reveal how the covariates influence the transition to more compromised health categories. Total height (THE) and longitude consistently exhibited negative and significant coefficients for the comparisons TH = 2 vs. TH = 1, TH = 3 vs. TH = 1, and TH = 4 vs. TH = 1. This indicates that taller trees and those located at higher longitudes (depending on the geographic gradient direction) tend to have a lower probability of belonging to more compromised health categories compared to healthy trees. In contrast, diameter at breast height (DBH), latitude, and bole quality (BQ) showed positive and significant coefficients across all comparisons. This suggests that trees with larger DBH, situated at higher latitudes (depending on the gradient), and with higher bole quality are more likely to fall into more compromised health categories (TH = 2, TH = 3, TH = 4) relative to TH = 1. The positive relationship between DBH and BQ with health compromise may seem counterintuitive at first but could indicate that larger, high-quality trees are older, more exposed to stressors over time, or simply more prone to host conditions that lead to health decline, such as liana infestation. Bole height (BH) exhibited a variable pattern: its coefficient was positive and significant for TH = 3 vs. TH = 1 and TH = 4 vs. TH = 1, but not significant for TH = 2 vs. TH = 1. This indicates that for more severe health impairment levels (TH = 3 and TH = 4), trees with taller boles tend to have a higher probability of being in these categories—a pattern that warrants further investigation.

These results are consistent with previous studies on tree–liana dynamics and successional processes in tropical forests. Several authors have reported that liana diameter tends to increase as forests mature, reflecting successional advancement and greater structural complexity (Letcher & Chazdon, 2009; Castello *et al*., 2017). In mature forests, environmental stability, canopy closure, and the presence of larger trees favor the establishment of large lianas, which use trees as structural support to reach the canopy and maximize light acquisition (Estrada-Villegas *et al*., 2022).

In contrast, early successional stages are typically characterized by a high abundance of small-diameter lianas (Dewalt *et al*., 2000; Medeiros & Torezan, 2013; Castello *et al*., 2017), commonly associated with open and disturbed environments. In these conditions, high light availability favors fast-growing and highly colonizing species, including pioneer lianas. As succession advances, forests undergo structural reorganization, promoting the establishment of larger and more persistent lianas capable of stronger competitive interactions with host trees (Castello *et al*., 2017).

Discussion of Copulas and Joint Dependence

The selection of the copula that best represents the dependence structure between PL_{bin} and TH was based on a comparison of AIC values (Table 4). The Clayton copula exhibited the best performance and was therefore selected as the joint dependence model. This copula is particularly suitable in situations where there is a higher concentration of probability in the lower tail, which, in an ecological context, may suggest that the absence of lianas (LP = 0) and high levels of tree health (TH = 1) tend to occur together more frequently.

The estimated dependence parameter for the Clayton copula was θ=0.2493, corresponding to a Kendall's tau coefficient of =0.1109and a Spearman's rho of _S=0.1707. Although both measures indicate a positive association between liana presence and tree health compromise, the estimated values are of low magnitude. This characterizes a weak dependence between the variables, indicating that, while a tendency for co-occurrence exists, it is not sufficiently strong to constitute a pronounced joint association between PL_{bin} and TH.

Studies have shown that liana presence negatively affects tree growth, biomass accumulation, water status, and reproductive performance, regardless of forest type or management purpose (Estrada-Villegas *et al*., 2022). These impacts directly compromise tree physiological and structural conditions, increasing susceptibility to stress, functional decline, and phytosanitary impairment. In this context, the results of the present study reinforce the hypothesis that liana presence is associated with higher levels of tree health deterioration, acting not only as a structural forest component but also as an important source of ecological pressure on host trees.

Lianas can also reduce forest productivity through intense competition for essential resources such as water and light (Meunier *et al*., 2021). Furthermore, their presence may limit carbohydrate allocation to flower and fruit production compared to trees free of crown infestation (Staudhammer *et al*., 2013; Kainer *et al*., 2014; Estrada-Villegas *et al*., 2022). This persistent competition compromises tree physiological

balance by reducing photosynthetic capacity, limiting water availability, and increasing metabolic maintenance costs.

Consequently, increasing liana loads may promote gradual phytosanitary weakening, reflected in reduced vigor, greater susceptibility to mechanical damage and environmental stress, higher mortality risk, and lower regeneration and growth capacity (Estrada-Villegas *et al*., 2022). Therefore, tree–liana interactions play a central role in forest dynamics, directly influencing forest structure, regeneration, carbon storage, and ecosystem resilience.

Discussion of the Joint Probability Tables

Table 5 presents the estimated joint probabilities, providing a detailed view of the co-occurrence between the different levels of tree health status and the presence or absence of lianas. Analyzing Table 5:

The highest joint probability is P(TH=1,"LP"=0)=0.3655, indicating a high likelihood of encountering healthy trees in the absence of lianas. This result is expected and ecologically desirable.

In contrast, the probability P(TH=1,"LP"=1)=0.1583 is considerably lower. This indicates that, although healthy trees can still occur in the presence of lianas, their likelihood is substantially reduced compared to environments without lianas.

As tree health status deteriorates (TH = 2, TH = 3, TH = 4), the joint probabilities associated with both LP = 0 and LP = 1 tend to decrease. This trend reflects the progressively lower frequency of increasingly compromised health states. However, it is essential to interpret these values in relative terms.

For the most severely compromised health categories (TH = 3 and TH = 4), joint probabilities are relatively low regardless of liana presence. This pattern reflects not only the reduced prevalence of these categories in the sample but also the complexity and multifactorial nature of the processes leading to severe declines in tree health.

The joint analysis enabled by the copula framework provides a richer perspective than isolated marginal analyses, revealing subtle nuances in the dependence structure between the variables. For instance, the probability of a tree being classified in the most severe health category (TH = 4, dead) in the presence of lianas is P("TH"=4,"LP"=1)=0.0338. Although this value is low in absolute terms, its ecological relevance emerges when compared with P(TH=4,"LP"=0)=0.0497. This comparison suggests that, in this specific case, the presence of lianas may not be the dominant factor driving tree mortality, or that other stressors may lead to tree death independently of liana occurrence. Such an interpretation must be made with caution and should be considered in conjunction with domain knowledge.

This interpretation is consistent with Estrada-Villegas *et al*. (2022), who reported that liana removal did not significantly reduce tree mortality, although it marginally improved sap flow and tree growth in silvicultural systems. The authors also highlighted the significant contribution of lianas to the forest leaf area index (LAI), reinforcing their structural role within the canopy. These findings emphasize the complexity of tree–liana interactions, as lianas function not only as competitors but also as important components of forest structure, influencing canopy dynamics and ecological processes related to productivity and nutrient cycling.

In this context, the copula-based joint analysis proved particularly relevant because it enabled the simultaneous evaluation of dependencies between liana presence and tree health status, revealing patterns not captured by isolated marginal analyses. The results suggest that liana presence is more associated with gradual physiological and functional weakening of trees than with immediate mortality. Therefore, tree–liana interactions should be understood within a multifactorial ecological context, where biotic and abiotic factors jointly influence forest dynamics, resilience, and ecosystem stability.

CONCLUSION

This study employed a statistical framework combining the Almost Unbiased Ridge-Type Principal Component Logistic Estimator (AURPLCE) for marginal modeling of binary variables under multicollinearity with a copula-based approach to model dependence between discrete variables. The methodology proved effective for analyzing complex ecological data, particularly the relationship between liana presence and tree health status using data from the National Forest Inventory of Minas Gerais.

The marginal modeling of liana presence (PL_{bin}) using the AURPLCE estimator indicated significant effects of dendrometric characteristics (BH, THE), spatial variables (latitude, longitude) and the morphological variable BQ on liana occurrence, with satisfactory predictive performance (AUC = 0.771). For tree health status (TH), the multinomial logistic model revealed interactions between covariates and health impairment levels, highlighting THE, DBH, latitude, longitude, and BQ as important predictors.

The Clayton copula was selected as the most appropriate model to characterize the dependence between PL_{bin} and TH, reflecting an asymmetric dependence structure and yielding a Kendall's tau correlation of τ =0.1109. This choice highlights the importance of flexible methods capable of capturing nonlinear relationships and asymmetries in ecological data. The resulting joint probability table and heat map provide valuable tools for visualizing and quantifying the likelihood of different combinations of liana presence and tree health status. The results indicate that the absence of lianas is strongly associated with healthy trees (TH = 1), whereas liana presence shows a moderate positive association with more compromised health states, although other factors likely play a substantial role in cases of severe health decline.

Overall, the proposed methodology enabled a comprehensive analysis of factors influencing liana presence and tree health status individually and jointly. These findings provide relevant implications for forest management and ecosystem stewardship by supporting a quantitative understanding of ecological interactions and contributing to more effective conservation and monitoring strategies. Future studies may extend this framework using

conditional copulas or pair-copula constructions to model more complex dependence structures involving a larger set of interacting variables.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for funding this research and promoting scientific development in Brazil.

DATA AVAILABILITY

The datasets supporting the conclusions are included in the article.

AUTHORSHIP CONTRIBUTION

Project Idea: IAP; MVGS
Database: MVGS
Processing: EMS; RG
Analysis: SEM; RG; MAC
Writing: IAP; MVGS
Validation: MAC; SJSSR; LMTC
Review: IAP; MAC; SJSSR; LMTC

REFERENCES

AKAIKE, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, v.19, p. 716–723, 1974.

BAMBER, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology , v.12, p. 387–415, 1975. DOI: 10.1016/0022-2496(75)90001-2.

BARRETO, A. Modelos de regressão: teoria e aplicações com o programa estatístico R. Alexandre Serra Barreto, 2011.

CARVALHO, P.G.; MELLIS, J.V.; ASCENSÃO, J.C.; *et al.* Abundância e biomassa de lianas em um fragmento de floresta atlântica, Hoehnea , v.2, p.307–314, 2011. DOI: 10.1590/S2236-89062011000200010.

CASTELLO, A.C.D.; COELHO, S.; CARDOSO-LEITE, E. Lianas, tree ferns and understory species: indicators of conservation status in the brazilian atlantic rainforest remnants, southeastern brazil. Brazilian Journal of Biology , v.77, n. 2, 2017. DOI: 10.1590/1519-6984.07715.

DEWALT, S. J.; SCHNITZER, S. A.; ALVES, L. F.; *et al.* "Biogeographical patterns of liana abundance and diversity," in Ecology of Lianas, eds S. A. Schnitzer, F. Bongers, R. J. Burnham, and F. E. Putz (Oxford: John Wiley and Sons), p. 131–146, 2015. DOI: 10.1002/9781118392409.ch11.

DEWALT, S.J.; SCHNITZER, S.A.; DENSLOW, J.S. Density and diversity of lianas along a chronosequence in a central panamanian tropical forest. Journal of Tropical Ecology, v.16, n. 1, p.1–19, 2000. DOI: 10.1017/S0266467400001231.

ENGEL, V.L.; FONSECA, R.C.B.; OLIVEIRA, R.E. Ecologia de lianas e o manejo de fragmentos florestais. Série Técnica IPEF, v.12, n. 32, p.43–64, 1998.

ESTRADA-VILLEGAS, S.; NARVAEZ, S.S.P.; SANCHEZ, A.; *et al.* Lianas significantly reduce tree performance and biomass accumulation across tropical forests: a global meta-analysis. Front. For. Glob. Change, v. 4, 2022. DOI: 10.3389/ffgc.2021.812066.

FAWCETT, T. An introduction to roc analysis. Pattern Recognition Letters, v.27, p.861–874, 2006. DOI: 10.1016/j.patrec.2005.10.010

GEROLAMO, C.S.; COSTA, F.R.C.; ZUNTINI, A.R.; *et al.* Hydro-edaphic gradient and phylogenetic history explain the landscape distribution of a highly diverse clade of lianas in the Brazilian Amazon. Front. For. Glob., v. 5, 2022. DOI: 10.3389/ffgc.2022.809904.

GOEMAN, J. J. Foundations of Linear and Generalized Linear Models. A.Agresti (2015). HOBOKEN, N.J: John Wiley & Sons, Inc. 444 pages, Biometrical Journal, v. 58, n. 5, p. 1258–1259, 2016. DOI: 10.1002/bimj.201600036.

GONZALEZ DE TANAGO, J.; LAU, A.; BARTHOLOMEUS, H.; *et al.* Estimation of Above-Ground Biomass of Large Tropical Trees with Terrestrial LiDAR. Methods Ecol. Evol.,v. 9, p. 223–234, 2017. DOI: 10.1111/2041-210X.12904.

GORA, E. M.; SCHNITZER, S. A.; BITZER, P. M.; *et al.* Lianas increase lightning-caused disturbance severity in a tropical forest. New Phytologist, v. 238, n. 5, p. 1865-1875, 2023. DOI: https://doi.org/10.1111/nph.18856.

GUEDES, J.S.; KRUPEK, R.A. Características ecológicas e fitossanidade de espécies arbóreas em um fragmento de floresta ombrófila densa da região sudeste do estado de São Paulo, Revista Ambiência, v.13, n. 2, p. 311–324, 2017. DOI: 10.5935/ambiencia.2017.02.04.

HAN, T.; SÁNCHEZ-AZOFEIFA, G.A. Extraction of liana stems using geometric features from terrestrial laser scanning point clouds. Remote Sens, v. 14, n. 4039, 2022. DOI: 10.3390/rs14164039.

HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology, v.143, p. 29–36, 1982. DOI: 10.1148/radiology.143.1.7063747.

HOSMER J.R. D.W.; LEMESHOW, S.; STURDIVANT, R.X. Applied Logistic Regression, 3 ed., John Wiley & Sons, Hoboken, NJ, 2013. DOI: 10.1002/9781118548387.

KAINER, K.A.; WADT, L.H.O.; STAUDHAMMER, C.L. Testing a silvicultural recommendation: Brazil nut responses 10 years after liana cutting. J. Appl. Ecol., v. 51, p. 655–663, 2014. DOI: 10.1111/1365-2664.12231.

LETCHER, S.G.; CHAZDON, R.L. Lianas and self-supporting plants during tropical forest succession. Forest Ecology and Management, v.10, p.2150–2156, 2009. DOI: 10.1016/j.foreco.2009.02.028.

LI, F.; ZI, H.; SONNE, C.; *et al.* Microbiome sustains forest ecosystem functions across hierarchical scales. Eco-Environment & Health, v. 2, p. 24-31, 2023. DOI: 10.1016/j.eehl.2023.03.001.

LIN, H.; CHAGANTY N.R. Multivariate distributions of correlated binary variables generated by pair-copulas. Journal of Statistical Distributions and Applications , v. 8, p. 1–14, 2021. DOI: 10.1186/s40488-021-00118-z.

MALHI, Y.; BALDOCCHI, D.D.; JARVIS, P.G. The carbon balance of tropical, temperate and boreal forests. Plant Cell Environ., v. 22, p. 715–740, 1999. DOI: 10.1046/j.1365-3040.1999.00453.x.

MEDEIROS, H.R.; TOREZAN, J.M. Evaluating the ecological integrity of atlantic forest remnants by using rapid ecological assessment. Environmental Monitoring and Assessment, v. 185, p. 4373–4382, 2013. DOI: 10.1007/s10661-012-2875-7.

METADADOS DO INVENTÁRIO FLORESTAL NACIONAL. Serviço Florestal Brasileiro, 2024. Disponível em: https://snif.florestal.gov.br/pt-br/inventario-florestal-nacional-ifn.

MEUNIER, F.; VERBEECK, H.; COWDERY, B. *et al.* Unraveling the relative role of light and water competition between lianas and trees in tropical forests: a vegetation model analysis. J. Ecol., v. 109, p. 519–540, 2021. DOI: 10.1111/1365-2745.13540.

NELSEN, R.B. An Introduction to Copulas. Springer Series in Statistics, 2006. 272p. DOI: 10.1007/0-387-28678-0.

PAN, Y.; BIRDSEY, R.A.; FANG, J.; *et al.*, A large and persistent carbon sink in the world's forests. Science, v. 333, p. 988–993, 2011. DOI: 10.1126/science.1201609.

PANAGIOTELIS, A.; CZADO, C.; JOE, H. Pair copula constructions for multivariate discrete data. Journal of the American Statistical Association, v.107, p.1063–1072, 2012. DOI: 10.1080/01621459.2012.682850.

PAROLARI, A. J.; PAUL, K.; GRIFFING, A.; *et al.* Liana abundance and diversity increase with rainfall seasonality along a precipitation gradient in Panama. Ecography, v. 43, p. 25–33, 2020. DOI: 10.1111/ecog.04678.

PERRY, D. A.; OREN, R.; HART, S. C. Forest Ecosystems. 2. ed. Baltimore, Maryland: Johns Hopkins University Press, 2008. DOI: 10.56021/9780801888403.

PROCTOR, J.; ANDERSON, J.M.; CHAI, P.; *et al.* Ecological studies in four contrasting lowland rain forests in gunung mulu national park, sarawak. Journal of Ecology, v.71, n. 1, p.237–260, 1983. DOI: 10.2307/2259975.

RUEDA-TRUJILLO, M. A.; VELDHUIS, M. P.; VAN BODEGOM, P. M.; *et al*. Global increase of lianas in tropical forests. Global Change Biology, 2025. DOI: https://doi.org/10.1111/gcb.17485.

SKLAR, A. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris, v. 8, p.229–231, 1959.

STAUDHAMMER, C.; WADT, L. H. O.; KAINER, K. Tradeoffs in basal area growth and reproduction shift over the lifetime of a long-lived tropical species. Oecologia, v. 173, p. 45–57, 2013. DOI: 10.1007/s00442-013-2603-1.

TORRES, C. M. M. E.; MEDINA-VEGA, J. A.; ROCHA, S. J. S. S.; *et al*. M. Drivers of tree demographic processes in forest fragments of the Brazilian Atlantic forest. Forest Ecology and Management, v. 549, p. 120893, 2023. DOI: 10.1016/j.foreco.2023.120893.

TUOMISTO, H.; RUOKOLAINEN, K.; VORMISTO, J.; *et al*. Effect of sampling grain on patterns of species richness and turnover in Amazonian forests. Ecography, v. 40, p. 840–852, 2017. DOI :10.1111/ecog.02453.

VARGAS, B.da.C; GROMBONE-GUARATINI, M.T.; MORELLATO, L.P.C. A dataset of neotropical liana research focusing on the strategies of control for forest restoration and management practices. Annals of forest Science, v. 79, n. 32, 2022. DOI: 10.5281/zenodo.6678112.

VARGAS, B.da.C; GROMBONE-GUARATINI, M.T.; MORELLATO, L.P.C. Lianas research in the Neotropics: data from 1950-2018. Zenodo, 2020. DOI: 10.5281/zenodo.4050477.

WU, J.; ASAR, Y. Performance of the almost unbiased ridge-type principal component estimator in logistic regression model. Communications in Statistics - Simulation and Computation, v.47, p. 2925–2937, 2018. DOI: 10.1080/03610918.2017.1364384.